IBERDROLA
INSTITUTO
TECNOLOGICO

JORNADAS DE INGENIERIA TELEMATICA

COMUNICACIONES

JITEL 97

INGENIERITZA TELEMATIKOARI BURUZKO JARDUNALDIAK

KOMUNIKAZIOAK

Bilbao
Septiembre 15-17, 1997
ETSII y de IT

Bilbo
1997ko Irailak 15-17
TI eta IIGME

1897
CENTENARIO
INGENIEROS BILBAO
MENDEMUGA

eman ta zabal zazu

universidad del país vasco
euskal herriko unibertsitatea

COMUNICACIONES DE LAS

Jornadas
de
Ingeniería
Telemática

JITEL 97

Ingenieritza
Telematikoari
Buruzko
Jardunaldiak

KOMUNIKAZIOAK

Bilbao
Septiembre 15-17, 1997
ETSII y de IT

Bilbo
1997ko Irailak 15-17
TI eta IIGME

Edita:
IBERDROLA INSTITUTO TECNOLOGICO
I.S.B.N.: 84-89654-04-2
Depósito Legal: BI-1653-97

Jornadas de Ingeniería Telemática Jitel 97

Comité de programa

Roberto Beitia	SARENET
Pedro Chas	TELEFÓNICA I+D
Mikel Emaldi	LABEIN
Iñaki Goirizelaia	ETSII y de IT, Bilbao
José Antonio López Egaña	ROBOTIKER
Jorge Mataix	ETSIT, Valencia
Juan Luis Núñez	EUSKALNET
Juan Quemada	ETSIT, Madrid
Emilio San Vicente	ETSIT, Barcelona
José M. Santos	ETSIT, Vigo

Comité de organización

Koldo Espinosa	ETSII y de IT, Bilbao
Armando Ferro	ETSII y de IT, Bilbao
Iñaki Goirizelaia	ETSII y de IT, Bilbao
Eduardo Jacob	ETSII y de IT, Bilbao
Mikel Olabe	ETSII y de IT, Bilbao
Juan José Uncilla	ETSII y de IT, Bilbao

Entidades Patrocinadoras

ETSII y de IT, Bilbao
Dep. Electrónica y Telecomunicaciones
IBERDROLA INSTITUTO TECNOLOGICO
ROBOTIKER

PRESENTACIÓN

Durante el presente año 1997, la Escuela Técnica Superior de Ingenieros Industriales y de Ingenieros de Telecomunicación de Bilbao cumple 100 años desde su fundación. Con este motivo nuestra escuela está celebrando diversos actos entre los que destaca la organización de congresos y jornadas de marcado carácter científico y tecnológico.

La celebración de las *Jornadas sobre Ingeniería Telemática, JITEL'97*, se enmarcan dentro de los actos conmemorativos del centenario siendo para nosotros un reto importante, puesto que a pesar de ser una escuela centenaria, la implantación del área de conocimiento de Ingeniería Telemática es relativamente reciente.

La razón de ser de nuestra área es lograr la comunicación de todo tipo datos y como decimos a nuestros alumnos/as, una buena comunicación depende fundamentalmente de la calidad de la señal y del medio de transmisión empleados. Quisiéramos que estas jornadas tengan un objetivo de comunicación que permita dar a conocer los trabajos y actividades de los diversos grupos de investigación, y que también sirvan para establecer un medio de transmisión adecuado para el intercambio del conocimiento generado.

La calidad de la señal a transmitir se refleja en las ponencias de las que vamos a disfrutar a lo largo de las jornadas y que tienen entre sus manos en este libro. Quisiéramos aprovechar para agradecer todas las aportaciones recibidas, sin olvidarnos de aquellas que por requisitos de espacio y tiempo nos hemos visto obligados a rechazar muy a nuestro pesar.

El medio que hemos elegido para la transmisión de este conocimiento, es la realización de estas jornadas. Esperamos y deseamos que este medio de transmisión que ahora iniciamos, tenga una continuidad a lo largo del tiempo y sea realmente un instrumento eficiente para dar a conocer nuestras actividades. Sabemos, sin embargo, que durante la puesta en marcha de esta jornadas habrá sin duda algunos fallos, por lo que de antemano pedimos disculpas. Pero también sabemos que estas jornadas van a suponer para todos un importante foro de discusión, del que creemos que nuestra área de conocimiento saldrá beneficiada sin ninguna duda.

Hemos organizado las jornadas teniendo en cuenta el contenido siguiente :

- Comunicaciones Avanzadas. Banda Ancha
- Impacto de las Tecnologías de Información en la Sociedad
- Aplicaciones en Educación

Creemos que son temas de máximo interés para nuestro área, como lo demuestra el elevado número de comunicaciones recibidas, lo cual nos ha obligado a organizar sesiones en paralelo. Estamos seguros que todos los congresistas van a disfrutar de las comunicaciones que serán defendidas a lo largo de estos tres días.

Para terminar la presentación de estas jornadas, sólo nos queda dar la bienvenida a Bilbao y a nuestra escuela a todos los participantes. Esperamos que vuestra estancia entre nosotros sea lo más agradable posible. Estamos a vuestra disposición.

Iñaki Goirizelaia

AURKEZPENA

1997. urte honetan, Bilboko Industri eta Telekomunikazio Ingeniarien Goi Mailako Eskolak haren sorreratik 100 urte betetzen ditu. Hori dela eta, gure eskola zenbait ekitaldi antolatzen ari da, haien artean batzarre eta jardunaldi tekniko eta zientfikoak nabarmentzen direlarik.

Ingenieritza Telematikoari Buruzko Jardunaldiak, JITEL'97, ekitaldi gomutagarri hauetan kokatzen dira. Guretzat erronka haundi bat suposatzen du, zeren eta nahiz eta ehun urteko eskola izan , Ingenieritza Telmatikoaren jakintza arlo honen ezarpena nahiko berria baita.

Gure arloaren helburua eta zeregina edozein datu-motaren komunikazioa lortzea da eta gure ikaslei esaten diegun bezala, komunikazio on bat seinalearen kalitatearen eta transmizio baliabidearen arabera lortzen da. Jaurdunaldi hauen bidez, Ikerketa talde desberdinen lanak eta ekitaldiak ezagutarazteko balio duen gure arteko komunikazioa lortzea gustatuko litzaiguke, eta era berean sortutako ezaguera elkar komunikatzeko transmizio baliabide egokia ezartzea.

Transmitidu behar dugun seinalearen kalitatea jaurdunaldi hauetan zuen eskuetan dauden eta gozatuko ditugun ponentzietan isladatzen da. Eskuratutako ekarpen guztiei eskerrak eman nahi dizkiegu, espazio eta denbora arazoengatik baztertu behar izan ditugunak ahaztu gabe.

Ezaguera honen transmiziorako aukeratutako baliabidea, jaurdunaldi hauek burutzea da. Orain hasten dugun transmizio baliabide honek denboran zehar jarraipena izan dezala eta benetan gure ekitaldiak ezagutaratzeko tresna baliagarria izan dadila nahi eta itxaroten dugu. Jakin badakigu halaere, jaurdunaldi hauek martxan ipintzerakoan huts batzu egon daitezkela, eta horregatik aldez aurretik barkamena eskatzen dugu. Baina beste aldetik ere, gure arloari on egingo dion eztabaidarako toki egokia izango dela uste dugu zalantzarik gabe.

Jaurdunaldiak ondoko edukin hau kontutan hartuta antolatu ditugu :

- Komunikazio Aurreratuak. Banda Zabala.
- Informazio Teknologien Eragina Gizartean
- Hezkuntzarako Aplikazioak

Gure arlorako interes handiena duten gaiak direla uste dugu, eta horren lekuko hor dago eskuratutako komunikazio kopuru altu eta adierazgarria. Hori dela eta jaurdunaldiak sesio paraleloetan antolatu behar izan ditugu. Hiru egun hauetan zehar defendiduko diren komunikazioekin batzarkideok une atseginak izango ditugula seguru gaude.

Jaurdunaldi hauen aurkezpena amaitzeko parte hartuko duten guztioi Bilbora eta eskolara ongi etorria ematea baino ez zaigu falta. Zuen egonaldia gure artean ahalik eta atseginena izan dadila espero dugu. Zuei laguntzeko prest gaude.

Iñaki Goirizelaia

CONTENIDO

CONTENIDO

CONTENIDO

CONTENIDO

CONTENIDO

Grupo I: Comunicaciones Avanzadas

Diseño e Interconexión de Redes

NETrix: Planificación estratégica de la estructura lógica de Redes de Banda Ancha

Alberto E. GARCIA, Klaus D. HACKBARTH †
GRUPO DE INGENIERÍA TELEMÁTICA, DEPARTAMENTO DE INGENIERÍA DE COMUNICACIONES ††
ETSII Y IT, UNIVERSIDAD DE CANTABRIA
Avda. de Los Castros S/N, 39005 SANTANDER
Correo electrónico: agarcia@tlmat.unican.es, klaus@tlmat.unican.es

Abstract:

NETrix is a strategic tool composes by specific user/client specification, and broadband service classification. For each service class implements the corresponding traffic model, calculating the two main stochastic moments of the source traffic (mean value and variance). Nodes can be classified considering nodes with or without predefined functionality following a heuristic algorithm based on the total source traffic and the geographical distance. For the source traffic distribution between each node pair NETrix uses a generical distribution function considering the influence of the predefined parameters for each broadband service in the origin-destination nodes and the traffic values and obtains the distribution of the user/client nodes into the two basic levels of the network: access layer and switching layer, establishing corresponding network hierarchy and the service area of each switch. As the network topology is obtained, NETrix realises the calculation of the corresponding traffic vectors of each switch with its concentrators. As the traffic vectors are known, NETrix calculate the final traffic matrix for the switching layer, and the network is dimensioned totally over the proposed scenario.

1. Introducción

El proyecto NETrix es una herramienta software desarrollada por el Grupo de Telemática del Departamento de Ingeniería de Comunicaciones de la Universidad de Cantabria que obtiene la distribución de los nodos de usuario en los dos niveles básicos de la red. Por un lado, la capa de acceso, constituida básicamente por concentradores, y por otro la capa de conmutación, con sus correspondientes conmutadores. A lo largo de los siguientes apartados se va a dar una idea básica de los fundamentos teóricos y prácticos en los cuales se basa este proyecto, esto es, la modelización del tráfico de fuente generado por los distintos servicios que componen un escenario.(apartados 2 y 3). Como complemento a dichas ideas, en el apartado 4 se desarrollan dos ejemplos prácticos de dimensionamiento, para finalizar con el apartado 5, en el que se analiza el comportamiento del programa bajo distintas plataformas.

2. Modelos de Trafico aplicables a la capa ATM

El concepto de tráfico, desde el punto de vista de la tecnología ATM, puede ser considerado, independientemente de los servicios, a tres niveles de resolución en el tiempo (Fig. 1):

- nivel de llamada, se corresponde con el comportamiento de las fuentes de tráfico sobre las correspondientes conexiones virtuales,
- nivel de ráfaga, se corresponde con el comportamiento estadístico del usuario, y
- nivel de célula, que describe la generación de células a nivel físico.

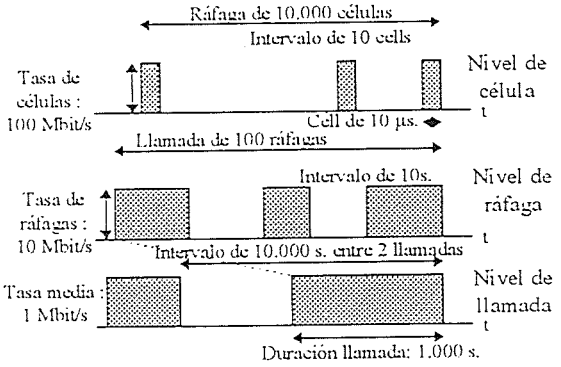


Fig.1.- Escalas temporales en el flujo ATM

Para realizar el correcto dimensionado de los elementos de la red, los parámetros a prever estarán dirigidos en dos vertientes: por un lado, determinar el tráfico ofrecido, considerando las características del flujo entrada/salida a nivel de ráfaga y célula respecto a ciertos servicios, las características del flujo de entrada a nivel de llamada, y las características del flujo resultante de la superposición de varias fuentes; por otro, especificar a nivel de llamada y de célula los parámetros de calidad de cada servicio.

Para la mayoría de las definiciones de señales estocásticas en redes ATM, no es necesario el conocimiento de la distribución de probabilidades para el proceso completo, sino solamente parámetros calculados o medidos en un periodo de tiempo determinado. En estos casos, los modelos de procesos modulados de Markov (MMP) con estados finitos pueden ser aproximaciones válidas para la descripción de fuentes, siendo los parámetros principales la velocidad media (S_1), la velocidad máxima (S_2), el momento de segundo orden, e incluso el momento de tercer orden. [1]

2.1 Servicios orientados a la conexión

Para los servicios orientados a la conexión tenemos que combinar los dos valores de salida obtenidos por la tasa de invocación (α_c) y la duración media del servicio (T_c) con los parámetros internos obtenidos en el cambio de estado mientras dure el servicio. Por otro lado, atendiendo a los tres parámetros de la señal de la fuente activa hemos de considerar un estado adicional "0", conectado al estado activo "1" y "2" por α_c, T_c, ó $a_c=\alpha_c \cdot T_c$. Aproximando la tasa de llamadas a una distribución poissoniana y la duración del servicio a una exponencial negativa, obtenemos una cadena de Markov de tres estados (con probabilidades p_0, p_1, p_2), que a su vez puede ser simplificada mediante un modelo de dos estados (p_0, p_i) siendo $p_i=p_1+p_2=\overline{a}_c$ y $p_0=1-\overline{a}_c$ (Ver Fig. 2).

La función de distribución común de las cadenas de bits de varias fuentes puede ser obtenida mediante la superposición de los modelos MMP de cada fuente. Esto solamente es cierto bajo el supuesto de que todas las fuentes y servicios son independientes, y puesto que la mayoría de los servicios pueden estar relacionados (caso de los servicios multimedia), esta suma es solamente una aproximación.[2],[3],[4]

2.2 Servicios no orientados a la conexión

Para los servicios no orientados a la conexión tenemos que tener en cuenta dos modelos on/off (Fig. 3). El primero obtiene los periodos en donde la fuente está en estado activo o en silencio (Ta, Ts), mientras que el segundo indica cuándo la fuente, estando en el estado activo, está transmitiendo o no (t_a, t_s) [5].

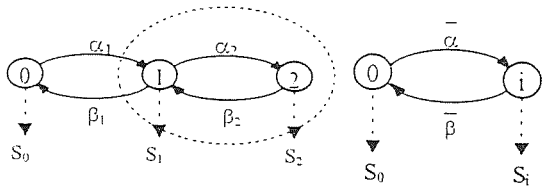


Fig.2.- Modelo MMP3 y su MMP2 correspondiente (servicios orientados a la conexión).

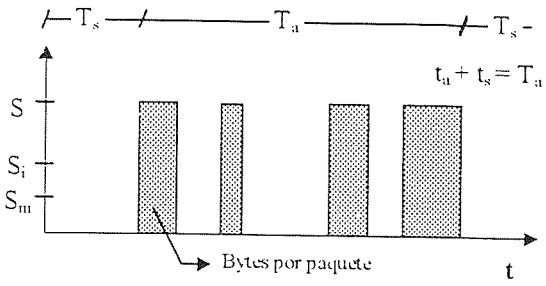


Fig.3.- Distribución temporal de las tramas no orientadas a la conexión.

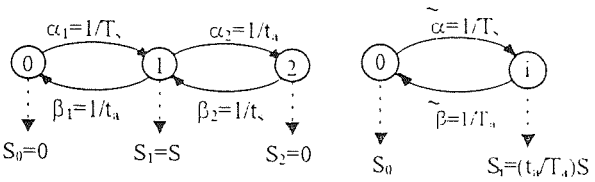


Fig.4.- Modelo MMP3 y su MMP2 correspondiente (servicios no orientados a la conexión).

Todo lo dicho para los servicios orientados a la conexión de dos y tres estados puede ser aplicado al caso de los servicios no orientados a la conexión (Fig.4).

3. Escenarios de Servicios

La planificación de la red debe realizarse teniendo en consideración todas las situaciones posibles en las cuales la red va a encontrarse desde su creación. Para poder realizar este estudio concienzudamente solo es posible el empleo de la emulación de dichas situaciones, para lo cual, a parte de necesitar un programa informático que lleve a cabo esa operación, es necesario alimentarlo con el correspondiente conjunto de datos de entrada, es decir, la información concreta acerca de la situación en la cual se encuentra la red

Un escenario de servicios debe aportar toda la información necesaria para concretar tanto el número de usuarios que va a soportar la red, como el conjunto de servicios que van a ser ofrecidos por la misma. En concreto, NETrix requiere de tres tipos fundamentales de información de entrada: los parámetros físicos de la red, los usuarios conectados a la misma, y los servicios provistos a los mismos.

3.1 Parámetros físicos de la red

Determinan la realidad física de la red emulada, e incluyen informaciones tales como la situación geográfica de los nodos de la red, el número de nodos de conmutación, las funcionalidades predefinidas para cada nodo, y los parámetros de corrección para el cálculo de la distancia o de la capacidad equivalente de tráfico generada en cada nodo.

3.2 Usuarios de la red

Es una información directamente relacionada con los datos individuales pertenecientes a cada nodo. Para el cálculo de la matriz de tráfico es necesario determinar áreas de usuario específicas, es decir, lugares donde todos los usuarios de ese área están conectados a un primer nodo de red común. Haciendo una simple analogía con las redes públicas clásicas, se corresponderían con áreas de la Red de Acceso interconectadas por un centro de conmutación de usuario (remoto, o a través de concentradores).

Para el caso ATM, podemos hacer un cálculo similar de las áreas de usuario y asociar a cada usuario su correspondiente demanda de servicios. Esta descripción resultaría excesiva si se llevara a la práctica, por lo que se debe recurrir a la clasificación de los usuarios en grupos con características de servicio comunes. Siguiendo la analogía con las redes clásicas, podemos distinguir entre usuarios privados y usuarios de negocios (divididos a su vez en usuarios de pequeña, mediana y gran empresa). Con esta descripción cabría suponer que el interés y la demanda de los servicios en la red serían iguales para todos los usuarios de un mismo grupo. Sin embargo, como esto no es del todo cierto, se puede establecer para cada grupo otra subdivisión según el nivel de demanda (pequeña, mediana y alta), con lo cual la clasificación total de los usuarios se realizaría de acuerdo a la Tabla I.

Esta clasificación permite asignar a cada grupo de usuario una demanda de servicio común, y calcular, a partir de los parámetros correspondientes el valor medio y la varianza del tráfico que será utilizado en el primer nodo común, y por consiguiente, continuar con el cálculo del tráfico de todo un área de acceso de usuario. De esta forma, una vez determinado el área a la que pertenece cada usuario, no necesitamos conocer cada usuario en particular, ni su localización. Tan solo será necesario conocer la cantidad de usuarios en cada grupo para calcular el tráfico total de ese área de servicio.

3.3 Servicios

Es el punto clave de la especificación del escenarios de servicios, abarcando todos los parámetros físicos necesarios para el cálculo del tráfico generado por cada usuario que haga uso del mismo, de acuerdo con el modelo al que se corresponde. NETrix realiza la clasificación de los servicios de Telecomunicación, siguiendo las definiciones de la UIT-T, en cuatro grandes grupos:

Tabla I.- Clasificación de usuarios utilizada por NETrix.

	TIPO DE USUARIO			
GRADO DE SERVICIO	PRIVADOS	DE NEGOCIOS		
		pequeña	mediana	grande
Bajo	uprb	unpb	unmb	ungb
Medio	uprm	unpm	unmm	ungm
Alto	upra	unpa	unma	unga

3.3.1 Servicios conversacionales

Un servicio conversacional es todo aquel servicio diseñado para la comunicación en tiempo real usuario-usuario o usuario-host. La información es generada por el/los usuario/s y es transmitida a uno o más usuarios finales. Para los servicios convencionales orientados a la conexión el conjunto mínimo de parámetros necesario es:

- *Velocidad* ó *tasa de bits* : Los parámetros de velocidad, tres para cada sentido de comunicación, cubren casi todos los atributos de simetría, esto es :servicios *unidireccionales* (solo valores en un sentido) ó *bidireccionales* (valores en ambos sentidos), *simétricos* (mismos valores en ambos sentidos) ó *asimétricos* (valores distintos en ambos sentidos). Teniendo en cuenta los valores concretos, podemos definir si el modelo de fuente es de *velocidad variable* (V_{min}, V_m y V_{max} distintas), o de *velocidad constante* (V_m y V_{max} coinciden).

- *Tasa de invocación* : definida para cada uno de los doce grupos de usuario, indica el valor medio del número de llamadas realizadas por cada usuario en el periodo de la hora cargada.

- *Duración media de la llamada* : también definida para cada grupo de usuario.

- *Número medio de destinos* : permite establecer los atributos para servicios punto a punto (valor 1) y punto multipunto (valor>1) para cada grupo de usuario.

- *Coeficiente interno, coeficiente de distancia* y *coeficiente de tráfico* : son parámetros adicionales definidos por NETrix para llevar a cabo diversas correcciones y ponderaciones en los cálculos de la matriz de tráfico, introduciendo las influencias tanto de la distancia, como la de la capacidad equivalente de tráfico de cada usuario.

La caracterización de cada uno de estos servicios es laboriosa teniendo en cuenta que parámetros tales como la duración de la llamada o la tasa de invocación dependen de múltiples factores, como por ejemplo el desarrollo económico y cultural del área de servicio. De esa forma, los valores concretos para cada parámetro deben ser obtenidos estadísticamente de acuerdo con las características tecnológicas, económicas y culturales del área de cobertura de la RDSI-BA (Ver Tabla II).

3.3.2 Servicios no orientados a la conexión

Un servicio no orientado a la conexión es todo aquel servicio que para comunicar dos puntos extremos no requiere del establecimiento y

mantenimiento de una conexión entre usuarios. Los parámetros fundamentales ahora serán :

- *Velocidad o Tasa de bits* : a diferencia con los servicios conversacionales, los servicios no orientados a la conexión se caracterizan por una velocidad media, pudiendo corresponderse con una fuente constante o variable, teniendo en cuenta que estos servicios responden a modelos a ráfagas, con períodos de actividad y silencio. Por ello se hace necesario la especificación del período interno de una ráfaga típica. Además, estos servicios pueden tener características especiales de simetría, por lo cual se deben especificar los valores de velocidad y de ráfaga en ambos sentidos de la comunicación.

- *Período activo* : medido en segundos, indica el tiempo total durante el cual la fuente no orientada a la conexión está transmitiendo dentro de un período.

- *Período de silencio* : de la misma forma, se mide en segundos, correspondiéndose con el tiempo total dentro de una ráfaga, en el que la fuente está en reposo (sin transmitir).

- *Nº medio de destinos* : estos servicios, aunque normalmente suelen dar comunicación usuario a usuario, presentan un modo de transmisión especial, denominado *multicast*, según el cual una fuente puede realizar una difusión de su mensaje a todo un grupo de usuarios.

- *Coeficiente interno, coeficiente de distancia y coeficiente de tráfico* : idénticos a los explicados para los servicios conversacionales.

Dependiendo de como se llevan a cabo estos servicios, algunos servicios convencionales podrían ser considerados también como servicios no orientados a la conexión. De entre ellos, cabría destacar todos los de transferencia de información, especialmente los servicios de comunicación LAN y los servicios de transferencia de datos (Ver Tabla II).

3.3.3 Servicios de extracción

En un servicio de extracción o de consulta, el usuario puede consultar información almacenada en centros específicos que en general son de uso público (aunque la tendencia es a privatizarlas). La información solo es facilitada cuando un usuario lo solicita, pudiendo controlar este el momento en el que desea comenzar a comprobar dicha información. Los parámetros a indicar ahora son :

- *Velocidad o tasa de bits* : De la misma forma que para los servicios conversacionales, y teniendo en cuenta que estos servicios suelen ser orientados a la conexión, se especifican seis valores de velocidad, tres para cada sentido de la comunicación (V_{min}, V_m y V_{max}), cubriendo todas las características de simetría.

- *Tasa de invocación* : definida para cada uno de los doce grupos de usuario, indica el valor medio del número de llamadas realizadas por cada usuario en el período de la hora cargada

Tabla II.- Clasificación y parámetros de servicios

Conversacional	origen-destino			destino-origen		
	min	med	max	min	med	max
Videotelefonía	384K	2M	10M	384K	2M	10M
Videoconf	384K	2M	10M	384K	2M	10M
Videovigilancia	0	2M	6M	0	64K	128K
vídeo/audio	0	50M	100M	0	0	0
Radiofonía	0	64K	128K	0	0.6K	2.4K
Transm. digital	0	2M	10M	0	2M	10M
Transf. ficheros	2M	2M	2M	2M	2M	2M
Teleacción	0	64K	128K	0	2M	10M
Telefax	2M	2M	2M	2M	2M	2M
Transf.imagen	0	128K	10M	0	128K	10M
Transf doc.	0	128K	10M	0	128K	10M

No orientados	origen-destino			destino-origen		
	Act	Silenc.	Tasa	Act.	Silenc	Tasa
LAN-Ethernet	0.001	0.0056	10M	0	64K	128K
Mensajería	0.001	0.0056	10M	0	64K	128K

Extracción	centro-usuario			usuario-centro		
	min	med	max	min	med	max
Videotex	0	10M	50M	0	64K	128K
VOD	0	10M	50M	0	64K	128K
BD gráficas	0	1M	10M	0	0.6K	2.4K
BD doc.	0	1M	2M	0	0.6K	2.4K
Consulta	0	1M	2M	0	0.6K	2.4K

Distribución	centro-usuario			usuario-centro		
	min	med	max	min	med	max
InterVideotex	0	10M	50M	0	64K	128K
Distrib. digital	2M	2M	2M	/	/	/
Distrib. vídeo	0	10M	50M	0	64K	128K
Distrib. doc	0	384K	2M	0	0.6K	2.4K
Distrib. TV	0	10M	50M	0	64K	128K

- *Duración media de la llamada*: también definida para cada grupo de usuario.

- *Coeficiente interno*: Con el mismo significado que en los servicios conversacionales. Los coeficientes de tráfico y distancia no tienen sentido en este tipo de servicio, ya que los centros de información siempre pueden tener situados centros secundarios en cada nodo ATM. De esta forma, el tráfico generado por una consulta, solo va a afectar al tramo de red correspondiente al área de servicio del usuario, sin introducir tráfico adicional a la red dorsal. Cuando la información requerida no se encuentre en la base de datos local, el sistema gestor del servicio se encargará de establecer una aplicación de transferencia de información entre el centro central de información y el centro local (servicio conversacional ó no orientado a la conexión anteriormente considerados).

3.3.4 Servicios de distribución

En los servicios de distribución, la información es distribuida por una central hasta un gran número de usuarios, pudiendo diferenciar entre servicios de difusión con control del usuario, y servicios de difusión sin control del usuario. En los servicios con control del usuario la información es proporcionada como una secuencia de entidades de información con repetición cíclica. De este modo, el usuario tiene la posibilidad de acceder individualmente a la información distribuida cíclicamente, pudiendo controlar el instante de comienzo y el orden de presentación. Por su parte, los servicios sin control por parte del usuario incluyen todos los servicios de difusión, en los que se produce un flujo continuo de información que es distribuido desde una central origen hasta un número ilimitado de usuarios autorizados. En ambos casos, los parámetros a especificar son idénticos a los utilizados para los servicios de extracción (Ver Aptdo..3.3.3 y Tabla II).

4. NETrix: Un par de ejemplos prácticos

4.1 Dimensionamiento de una nueva red ATM

Supongamos que nos encontramos en la necesidad de establecer el primer dimensionamiento de una red ATM, en un área a nivel nacional, problema que podría enunciarse como:

"Se pretende establecer una infraestructura de red ATM sobre el territorio peninsular español. Todas las provincias estarán servidas por una central (concentradora), situada en la capital (47 centrales), de las cuales 5 serán nodos conmutadores. Se desea obtener la jerarquía de red resultante indicando cual es la situación ideal de los conmutadores, y cual es el posible dimensionamiento de los enlaces entre cada par de nodos."

Nota.- Se supone un escenario de servicios típico, y estructura completamente mallada (a efectos lógicos).

4.1.1 Datos de entrada

Los datos a introducir por el operador son todos aquellos que se refieren a la situación geográfica de cada nodo, así como el número de usuarios a quienes da servicio, clasificados estos de acuerdo con los doce grupos de usuarios indicados en la Tabla I. El operador de la red dispone de todos estos datos estadísticos pudiendo ser extrapolados a partir de las infraestructuras de red actuales. En nuestro ejemplo se tuvo en cuenta la población (tomada del Censo de 1993) y el número de usuarios del servicio telefónico básico y del servicio postal.

Para cada nodo deben introducirse:

- Coordenadas geográficas: longitud y latitud.

- Función del nodo: podemos establecer que todos los cálculos se realicen considerando desde el principio al nodo como concentrador o conmutador.

- Usuarios de la red: agrupados por grupo de usuario.

Por su parte, la información referente a los servicios que van a ser ofrecidos por la red es la misma que se ha expuesto en el apartado 3.3. El conjunto de servicios utilizados para este ejemplo son los de la Tabla III, pudiendo utilizar los datos de la Tabla II.

4.1.2 Ejecución de NETrix

Una vez introducida toda la información de entrada, podemos proceder a la ejecución de los tres algoritmos que componen NETrix :

Traffic : Realiza el cálculo de la capacidad equivalente de tráfico generado por el conjunto de usuarios y servicios en cada uno de los nodos. Con sus resultados podemos establecer una clasificación de nodos por tráfico generado.

Tabla III.- Escenario de servicios típico

Conversacionales	POTS (Telefonía) Voz comprimida (Radiofonía) Videoconferencia Transferencia de ficheros Telefax
No orientados a la conexión	MMMS (correo electrónico)
Extracción	Videotex VOD (Vídeo bajo demanda)
Distribución	TV

Clasig : Introduciendo como datos de entrada el *número de nodos de conmutación* que deseamos tenga la red, la *distancia mínima* que debe existir entre dos nodos de conmutación y *parámetros de ponderación* del tráfico y la distancia, NETrix realiza la asignación de las funciones específicas de conmutación o concentración para cada nodo individual. Además, establece las correspondientes áreas de conmutación, es decir, determina el área total en la que un único nodo conmutador interconecta las distintas áreas de usuario, servidas a su vez cada una por un concentrador. De esta forma NETrix determina la jerarquía de la red ATM (plano de acceso y plano de conmutación), asignando a cada nodo conmutador un conjunto de nodos concentradores.

Matrix : Una vez establecida la jerarquía de la red, NETrix realiza el cálculo del tráfico al nivel origen-destino para cada servicio y proyecta dichos valores a la estructura lógica de la red. Como resultado se calculan los *vectores de acceso* entre cada nodo concentrador y su correspondiente nodo conmutador (un vector para cada sentido, ya que los enlaces no son totalmente simétrico). Por otro lado, se calcula la matriz de tráfico ATM, que no es otra cosa que los vectores de acceso entre cada par de nodos conmutadores, teniendo en cuenta que desde el plano de la conexión de cruce, todos los conmutadores están enlazados con todos.

4.1.3 Análisis de los resultados

La capacidad equivalente que requiere el tráfico generado en cada nodo para el conjunto de los servicios especificados anteriormente, se corresponde con uno de los ficheros de resultados del algoritmo Traffic.

Para el algoritmo de clasificación introducimos como parámetros el número de conmutadores (probamos para 5, 6 y 7), una distancia mínima de 100 Km entre conmutadores, y parámetros de ponderación de 1.00 para la distancia y 0.00 para el tráfico (la asignación se realizará en función de distancias sin tener en cuenta el nivel de tráfico del nodo destino). Se obtuvieron configuraciones de red como las de la Fig. 5.

Con estas configuraciones se puede realizar el cálculo de las correspondientes matrices de tráfico. Gráficamente se representarían de la misma forma que la Fig.6 para los tres casos, aunque para seis y siete conmutadores cambian los valores de la leyenda gráfica.

Analizando los distintos casos se puede observar como al aumentar el número de nodos conmutadores los enlaces individuales reducen su carga de tráfico. Aún así, los valores de tráfico son muy elevados (cientos de Gbits por segundo), lo que puede llevar a pensar en varias posibilidades :

- Aumentar el número de nodos con función de conmutación, lo que incrementa la dimensión de la matriz de tráfico reduciendo los valores en sus enlaces y en los vectores de acceso.
- Separar el tráfico de diferentes grupos de servicios, sobre todo asíncronos y síncronos, y encaminarlo sobre diferentes estructuras lógicas en forma de caminos virtuales. Esto permite introducir memoria adicional para los servicios asíncrono, y un mejor rendimiento en el dimensionamiento.
- Dividir los nodos superiores con alta carga de tráfico en dos o más nodos separados, introducir algunos nodos superiores adicionales y asignar cada nodo inferior a dos nodos superiores con división de carga de tráfico. De esta forma se consigue una estructura más fiable contra fallos y sobrecarga, y una matriz de tráfico más equilibrado.

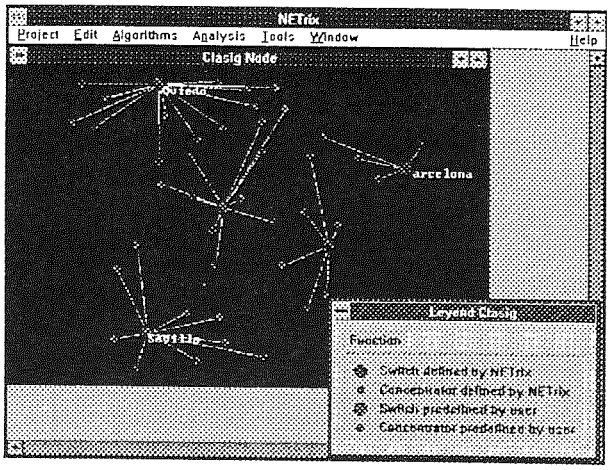


Fig. 5: Asignación de áreas de conmutación (cinco conmutadores)

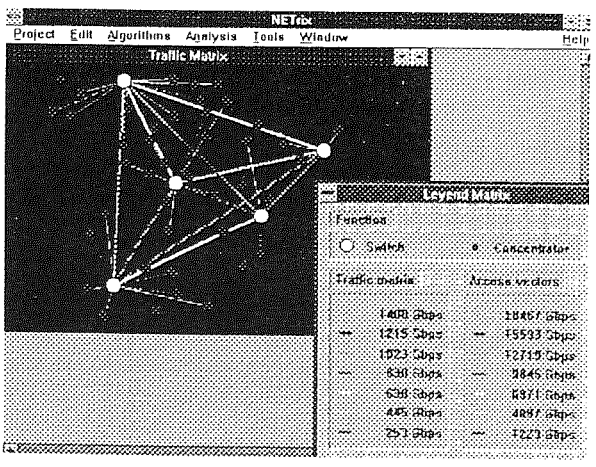


Fig. 6.- Representación gráfica de la matriz de tráfico y los vectores de acceso (5 conmutadores).

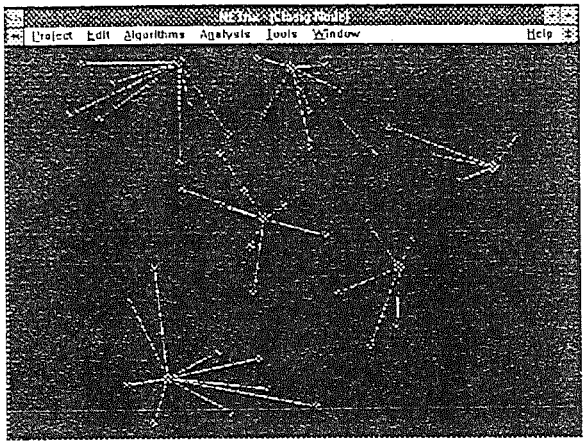


Fig. 7: Situación final introduciendo Bilbao como nodo conmutador

4.2 Redimensionamiento de la red

Ahora el supuesto podría ser: *"Las necesidades de la red ATM y criterios de posicionamiento estratégico para el Operador obligan a colocar una central de conmutación en Bilbao. En este supuesto, determinar qué centrales concentradoras dependerán de este nodo y calcular la matriz de tráfico de la nueva red."*

Partiendo de una red como la calculada para el primer supuesto, compuesta por cinco nodos conmutadores situados en Oviedo, Madrid, Barcelona, Valencia y Sevilla, los datos a introducir son completamente los mismos. La única variante es que deberemos preasignar la función de conmutación a los cinco nodos ya existentes, y además indicaremos el nuevo nodo de conmutación (Bilbao).

La capacidad equivalente de tráfico no varía puesto que no se han modificado ni el número de usuarios de cada nodo, ni el escenario de servicios, por lo que los valores coinciden con los del primer supuesto.

Sin embargo, ahora la asignación de nodos ha variado (ver Fig 7), y al igual que en el primer supuesto , al introducir un nuevo nodo conmutador el tráfico proyectado a cada enlace individual disminuye, no solo por la aparición de nuevas rutas (lo cual descongestiona a las existentes), sino también por la reducción del número de nodos concentradores conectados a los nodos ya existentes (el nuevo nodo recoge nodos concentradores de los nodos conmutadores circundantes).

5. Algunas cifras de NETrix

El GIT-DICOM ha desarrollado a lo largo de estos dos últimos años tres versiones de NETrix, sobre MSDOS, WINDOWS y UNIX/LINUX.

Todos estos productos han sido ampliamente testeados con el fin de comprobar la fiabilidad de los programas y de los escenarios introducidos bajo puntos de vista diferentes:

- La potencia de cálculo y la portabilidad del programa: se comprobó el correcto funcionamiento de los programas bajo UNIX/LINUX y WINDOWS hasta un total de 2700 nodos (últimos ejemplos constaban de 3000 nodos). Para ello se utilizaron plataformas diferentes, tanto PC's 486 y Pentium, como estaciones HP y SUN (*Nota.- Adicionalmente, la empresa alemana DeTeLine verificó el funcionamiento del programa bajo nuevas plataformas, con procesadores PENTIUM y estaciones SUN IPX.*).

- La velocidad de cálculo: al mismo tiempo que se verificaban los límites de NETrix, también se midieron los tiempos de ejecución para todos los algoritmos y ambas versiones (Tablas IV y V).

Observando los dos gráfico de la Fig.8 se puede comprobar la diferencia entre las distintas plataformas y sistemas. Podríamos considerar 750 como el límite para sistemas de baja potencia, mientras que para sistemas potentes no existen problemas para la manipulación de un número mucho más elevado de nodos (más de 3000). De hecho, podemos comprobar gráficamente la evolución del comportamiento de la versión LINUX de NETrix a partir de las emulaciones hechas en cada una de las plataformas utilizadas.

Tabla IV: Tiempos de ejecución para la versión final de NET-LINUX (NETrix UNIX/LINUX)

Nodes / *Processor*	*50* *Traffic/Clasig/Matri x*	*200* *Traffic/Clasig/Matri x*	*750* *Traffic/Clasig/Matri x*	*1250* *Traffic/Clasig/Matri x*	*2700* *Traffic/Clasig/Matri x*
486DX50/8M	1"/ 1"/ 3"	2"/ 3"/ 36"	5"/ 15"/ 9' 45"	10"/ 40"/ X'	
486DX100/16M	<1"/ 1"/ 3"	1"/ 3"/27"	4"/ 11"/ 6' 10"	5"/ 26"/ 11' 15"	13"/ 2' 25"/ X
P150/16M	<1"/<1"/<1"	<1"/1"/12"	2"/13"/3' 3"	4"/1' 1"/9' 15"	7"/11' 47"/61'22"
P100/32M	<1"/<1"/1"	1"/1"/19"	2"/19"/4'41"	4"/1'35"/13'15"	9"/16'51"/67'19"
HP9000/32M	<1"/ 1"/ 3"	1"/ 2"/ 15"	3"/ 7"/ 2' 45"	5"/ 17"/ 20'	9"/ 1' 20"/ 42'
SUN IPX/32M	1"/<1"/5"	2"/1"/37"	11"/59"/17'38"	18"/5'1"/48'34"	29"/58'5"/224'52"

Tabla V: Tiempos de ejecución para la versión "demo" NET-WIN (NETrix WINDOWS)

Nodes / *Processor*	*50* *Traffic/Clasig/Matri x*	*200* *Traffic/Clasig/Matri x*	*750* *Traffic/Clasig/Matri x*	*1250* *Traffic/Clasig/Matri x*	*2700* *Traffic/Clasig/Matri x*
486DX50/8M	3"/ 3"/ 10"	6"/ 9" / 1' 30"	10"/ 3'/ 40"	11"/ 5'/ X	X
486DX100/16M	1"/ 2"/ 3"	2"/ 14"/ 34"	4"/ 2' 20"/ X	X	X

X: Overflow o tiempos de ejecución elevados.

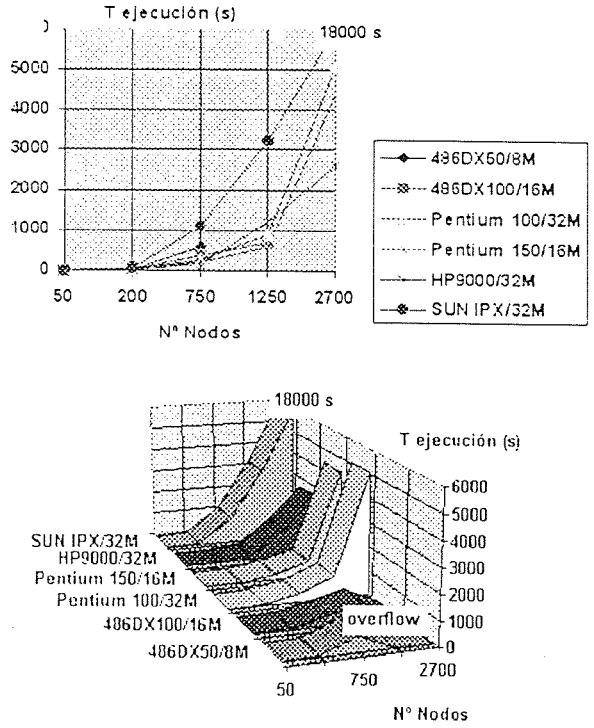


Fig. 8: Tiempos totales de ejecución en función del número de nodos y de las plataformas utilizadas (Versión NETrix UNIX/LINUX).

4. Conclusiones

Analizando los resultados obtenidos por NETrix pueden obtenerse las siguientes conclusiones :

• La evolución de los servicios de Telecomunicación suponen un alto grado de incertidumbre en el conocimiento de los requerimientos del ancho de banda. Sin embargo es posible la parametrización de las características de cada tipo de servicio mediante el uso de modelos estocásticos sencillos, con lo cual dicha evolución puede reflejarse en la simple modificación de los valores de dichos parámetros.

• La modelización del comportamiento de las fuentes de cada tipo de servicio permite agrupar a los mismos estableciendo diferentes escenarios de servicios que pueden hacer corresponderse con situaciones reales en la red ATM. De la misma forma que cada servicio puede modelarse mediante un conjunto de parámetros, un escenario puede modelarse mediante la agrupación de los conjuntos de parámetros correspondientes a los diferentes servicios considerados.

• Ciertas herramientas software, como es el caso de NETrix, pueden hacer uso de estos escenarios de servicios, pudiendo establecer situaciones ficticias dentro de una red ATM existente o no. El análisis de los resultados permite establecer, por ejemplo,

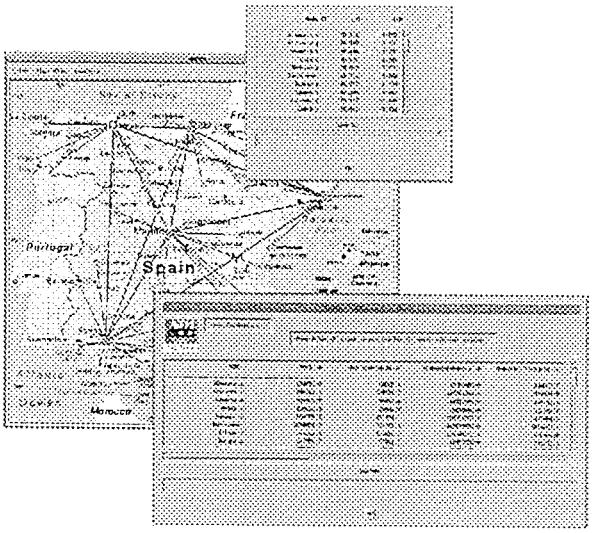


Fig. 9: Interfaz gráfica en la versión NETrix UNIX/LINUX.

los límites de una estructura lógica determinada (ya sea en cuanto a número de usuarios o a número y tipo de servicios), emular situaciones críticas en la red o incluso diseñar la estructura lógica de nuevas redes ó nuevos tramos de la red.

• El rango de aplicación de la generación de escenarios, los algoritmos heurísticos utilizados para la clasificación y jerarquización de la red, y los modelos estocásticos utilizados para el modelado de servicios va desde ejemplos sencillos de redes privadas de muy pequeño número de nodos (por ejemplo una red de vídeo privado a nivel provincial), hasta la implementación de grandes redes ATM de uso público con miles de nodos (un ejemplo desarrollado por la empresa DeTeLine considera 2800 nodos de la futura red ATM alemana), con resultados altamente satisfactorios, tanto a nivel de tiempos de ejecución, como de fiabilidad de los resultados.

Para finalizar, solo haremos mención a las líneas de trabajo en las cuales actualmente se está desarrollando el proyecto NETrixa:

• Creación de modelos de clasificación de usuarios, donde podemos considerar dos vertientes distintas: por un lado, la clasificación de usuarios en función de datos estadísticos en posesión del operador; y por otro lado, la estimación y predicción de la evolución del número de usuarios (y grupo) en función del tipo de servicio. De esta forma podría considerarse clasificaciones de usuarios para cada tipo de servicio, evitando hacer uso de una clasificación general que puede no ser del todo exacta para todos los servicios (por ejemplo dentro del grupo de usuarios de negocios, no utilizan los

mismos servicios empresas de automoción, o multinacionales de alimentación).

- Ampliación de la estructura jerárquica ATM, es decir, incorporar la clasificación y definición de la capa de acceso a la red ATM desarrollada por la actual versión NETrix. Ahora, los datos de entrada no son solo nodos concentradores o conmutadores, sino que la especificación de los nodos abarque hasta el nivel de usuario. Es decir, a partir de los puntos de conexión de usuario se establece el nivel de acceso (se determinan los nodos de los niveles inmediatamente superiores (concentradores y conmutadores), para posteriormente establecer la jerarquía del nivel superior, tal y como se realiza en la versión NETrix analizada en este trabajo (Fig 10 y Fig 11).

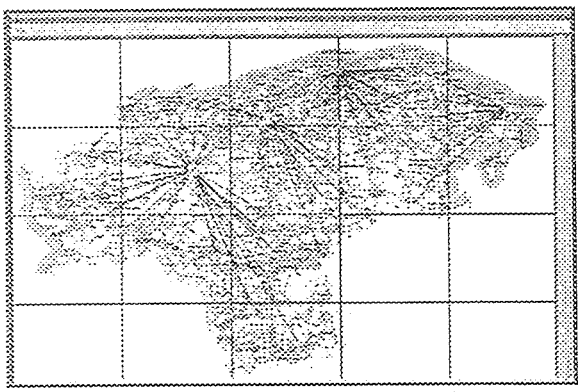


Fig. 10.- Acceso a nivel regional (versión LINUX).

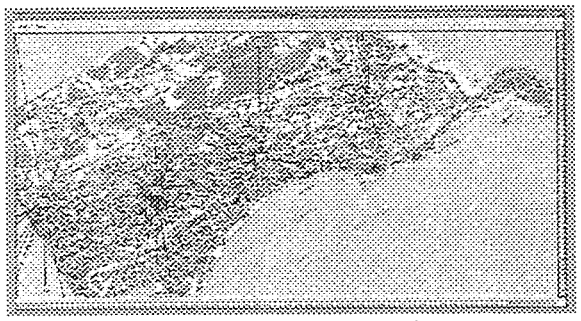


Fig. 11.- Acceso a nivel local (versión LINUX).

- Perfeccionamiento de los modelos estocásticos MMP3 y MMP2 y desarrollo de nuevos modelos de servicios (por ejemplo modelos brownianos para modelado de servicios multimedia) para su integración en nuevas versiones de NETrix u otros programas informáticos futuros.

- Verificación y modificación del algoritmo heurístico para la la clasificación y asignación de los nodos realizada por el algoritmo CLASIG de NETrix.

Referencias

[1] Andrade Parra. "Statistical Parameters to Describe Cell Traffic Generated by Broadband Services". *Comunicaciones de Telefónica I+D*, **4**, 2, 1990.

[2] Grüber. "A Comparison of Measured and Calculated Speech Temporal Parameters Relevant to Speech Activity Detection" *IEEE Trans. Of Com.***30**, 4 (1982).

[3] Heffes, Lucantoni. "A Markov Modulated Characterization of Packetized Voice and Data Traffic", *IEEE JSAC*, **4**, 6, (1986).

[4] Saito et all. "An Analysis of Statischical Multiplexing in an ATM Transport Network", *IEEE JSAC*, **9**, 3, (1991).

[5] Onvural Raif O. "Asynchronous Transfer Mode Networks: Performance Issues", *Artech House, INC* (1992).

Interconexión de redes FDDI a través de una red ATM

Maribel Narganes, Enrique Areizaga
ROBOTIKER, LÍNEA DE REDES DE DATOS
PARQUE TECNOLÓGICO DE ZAMUDIO, EDIFICIO 202
48170 ZAMUDIO, BIZKAIA
Correo electrónico: maribel@robotiker.es

Abstract:

The Broadband Integrated Services Digital Network (BISDN) will come with a great amount of different services such as multimedia videotext, interactive multimedia communications, video on demand, and so on. All this services will be offered to the users by means of an unique subscriber's access. This will be possible thanks to its enormous capacity in terms of the high speed transmissions supported and to the new schemes of switching and transmission, based on the Asynchronous Transfer Mode (ATM), which are used. In this article we will explore one of the most imminent application of the BSIDN: interconnection of present networks which cover a geographical restricted area (LANs and MANs) through a public BISDN.

1. Introducción

A lo largo del presente artículo se presenta una solución a la problemática que involucra el desarrollo de un equipo Adaptador FDDI-ATM capaz de interconectar redes FDDI (Fiber Distributed Data Interface) remotas a través una red de banda ancha basada en la tecnología ATM, tal y como muestra la Fig. 1. Este equipo Adaptador permite la extensión de las redes FDDI usando la red pública de Banda Ancha, de tal forma que el ámbito de cobertura de este tipo redes (limitado actualmente a los 100 Km de extensión del anillo) quede ampliado a toda la extensión cubierta por la red de banda ancha.

El Adaptador se encuentra, por una parte, conectado a una red FDDI, a la que ofrece el servicio de interconexión, a través del interfaz adecuado. Desde el punto de vista de los diversos terminales de la red, el Adaptador será considerado como un Encaminador (Router) que les permite comunicarse con terminales FDDI pertenecientes a redes FDDI remotas de una forma totalmente transparente.

Por otra parte, el Adaptador FDDI-ATM se conecta al punto de acceso a la RDSI BA a través del punto de acceso Tb (interfaz usuario/red) de la red utilizando circuitos semipermanentes. Cada circuito semipermanente activo facilitaría la interconexión con un equipo Adaptador, y por tanto con una red FDDI remota.

De esta forma la funcionalidad básica del Adaptador consistiría en analizar la información recibida a través de ambos interfaces y decidir a través de qué interfaz debe retransmitirse, realizando en cada caso la conversión de formato adecuada.

En la solución propuesta, el Adaptador realiza el encaminamiento de la información sobre el nivel 3 o nivel de red del modelo de referencia OSI utilizando el protocolo de encaminamiento del protocolo de nivel 3 de Internet, denominado comúnmente protocolo IP (Internet Protocol).

Adicionalmente, el Adaptador va a incorporar la capacidad de ser gestionado tanto de forma local, a través de un terminal conectado directamente al Adaptador, como de forma remota a través de un Agente SNMP.

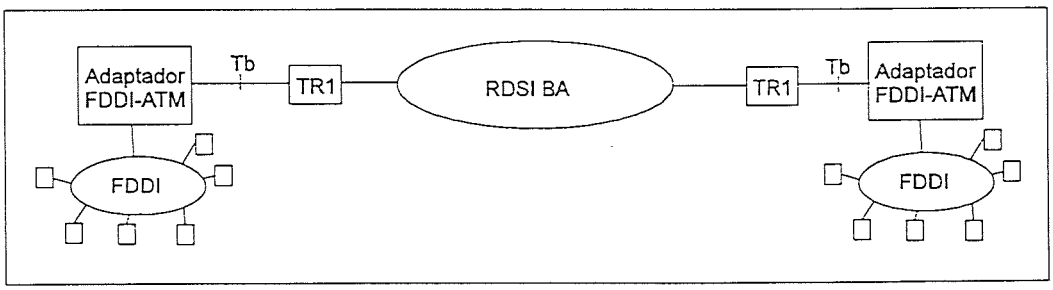


Figura 1. Diagrama de Interconexión del Adaptador FDDI/ATM

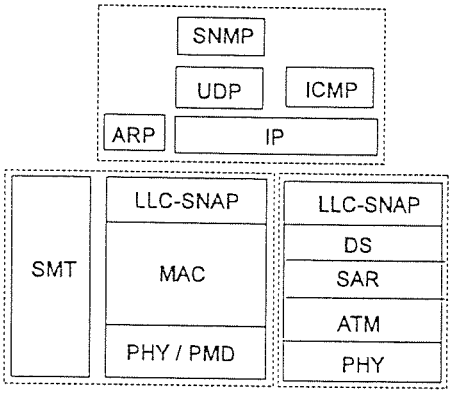


Figura 2. Arquitectura de red básica del Adaptador FDDI/ATM

2. Arquitectura de red del Adaptador FDDI/ATM

La Fig. 2 muestra la Arquitectura de Red básica del Adaptador FDDI/ATM, en la cual se reflejan las entidades de protocolo de los diferentes niveles de la arquitectura incluidos en el Adaptador FDDI. En esta Arquitectura es posible diferenciar tres módulos principales, cada uno de los cuales agrupa un conjunto de entidades de protocolo relacionadas entre sí.

2.1 Módulo Interfaz FDDI

Este módulo implementa el interfaz físico multimodo necesario para el acceso a un anillo doble FDDI. Este interfaz incorpora los dos niveles inferiores del modelo de referencia OSI: el nivel físico y el nivel de enlace de datos.

El nivel físico se subdivide a su vez en dos subniveles: el nivel físico dependiente del medio o PMD que proporciona la comunicación punto a punto en banda base digital entre dos nodos de la red, definido en el estándar ISO/IEC 9314-3 [3] (ANSI X3.166); y el protocolo de nivel físico PHY, definido en el estándar ISO/IEC 9314-1 [1] (ANSI X3.148), que proporciona la conexión entre el PMD y el nivel inmediatamente superior o nivel de enlace de datos. El interfaz físico incorpora dos entidades PMD y dos entidades PHY, de tal forma que cada par PHY-PMD corresponde a uno de los puertos, puerto A y puerto B, de un interfaz con doble conexión (DAS: Dual Attachment Interface).

El nivel de enlace de datos se encuentra a su vez subdividido en el subnivel de control de acceso al medio o subnivel MAC y el subnivel de control del enlace lógico o subnivel LLC.

El subnivel MAC, definido por el estándar ISO/IEC 9314-2 [2] (ANSI X3.139), proporciona el acceso al medio utilizando el protocolo de Rotación de Testigo Temporizado (TTR: Timed Token Rotation) basado el la utilización de un Testigo que otorga el derecho a transmitir, controlando y organizando las transmisiones dando soporte a dos tipos de servicios de transmisión, uno síncrono y otro asíncrono. Además, el subnivel MAC, se encarga tanto de la generación y envío, como de la recepción y procesamiento de tramas MAC haciendo uso del servicio de intercambio de tramas que le ofrece el subnivel PHY. El protocolo de nivel MAC que gobierna el intercambio de tramas entre entidades de nivel MAC no sólo proporciona un servicio de transporte de información al nivel superior, sino que además permite llevar a cabo una serie de procedimientos de control y gestión del anillo tales como: inicialización del anillo, procedimiento de petición del testigo, monitorización del anillo, inserción de nuevas estaciones en el anillo, etc.

El subnivel LLC realiza un servicio LLC de la Clase 1 (Servicio sin Conexión y sin Reconocimientos) definido en el estándar ISO 8802-2 [4] (IEEE 802.2).

Para llevar a cabo el encapsulamiento de los datos procedentes de los niveles superiores, se ha optado por el encapsulamiento SNAP definido en el RFC 1188 [10]. La Fig. 3 representa el formato de las tramas LLC así como su codificación en base al encapsulamiento SNAP.

Por último, las funciones de gestión de todo el interfaz FDDI son llevadas a cabo a través de la entidad SMT (Station Management) según lo especificado en el estándar ANSI X3T9/90 [5], soportando todos los objetos de carácter mandatorio y parte de los objetos de carácter opcional de la Base de Información de Gestión o MIB definido en dicho interfaz para estaciones de tipo DAS.

DSAP	SSAP	CONTROL	ORG.CODE	ETH. TYPE	Inform. (Trama MAC)

DSAP: (1 byte) SAP Destino [AA h]
SSAP: (1 byte) SAP Origen [AA h]
CONTROL: (1 byte) Campo de Control [UI, XID o TEST]
ORG. CODE: (1 byte) [00 h]
ETH. TYPE: (4 bytes) [0800 h para IP y 0806 h para ARP]

Figura 3. Formato de las tramas LLC y su codificación en base al encapsulamiento SNAP

2.2 Módulo Interfaz ATM

Este módulo consta de los siguientes niveles y subniveles: el nivel físico o PHY, el nivel ATM, el nivel de Adaptación ATM o AAL, subdividido en los subniveles SAR (Subnivel de Segmentación y Reensamblado) y CS (Subnivel de Convergencia), y el nivel LLC-SNAP.

El nivel PHY engloba las funciones de capa física que hacen posible el transporte de celdas ATM en el campo de carga útil de tramas SDH (Synchronous Digital Hierarchy) a una velocidad de 155 Mbps.

La funcionalidad del nivel ATM se basa en la formación de una celda ATM de 53 octetos por cada segmento de datos de 48 bytes recibido del subnivel SAR inmediatamente por encima de él, así como el procesamiento de cada celda ATM recibida del nivel PHY y su posterior entrega al nivel SAR de la unidad de datos resultante de la eliminación de los 5 bytes de la cabecera ATM. Este nivel opera en base a la especificación recogida en la Recomendación I.361 del CCITT [6].

Para cubrir la funcionalidad de la capa de Adaptación ATM o capa AAL se ha adoptado el estándar AAL Tipo 3/4 especificado en las Recomendaciones I.362 e I.363 [7][8] para el transporte de información a velocidad variable. El Subnivel de Segmentación y Reensamblado (SAR), dentro del nivel AAL, lleva a cabo la segmentación y reensamblado de las unidades de información procedentes o dirigidas hacia el Subnivel de Convergencia (CS) de acuerdo a como se especifica en las citadas recomendaciones.

El subnivel SAR proporciona, por una lado, el medio necesario para la transferencia simultánea de múltiples paquetes de longitud variable a través de un mismo interfaz ATM. De esta forma, por cada paquete de datos procedente del subnivel CS, durante el proceso de segmentación se generarán un conjunto de SAR-PDUs de 48 bytes de longitud (44 bytes de carga útil más 4 bytes de sobrecarga) cada una de las cuales constituirán el campo de carga útil de una celda ATM. Por otro lado, durante el proceso de reensamblado, el subnivel SAR deberá reconstruir paquetes destinados al subnivel CS a partir del conjunto de segmentos de 48 bytes recibidos del nivel ATM.

En lo que se refiere al Subnivel de Convergencia o CS, se ha adoptado el protocolo DS (Data Service) que soporta el servicio de datos sin conexión sobre la Red de Banda Ancha, de acuerdo a la especificación de Bellcore del servicio SMDS para el interfaz B-UNI (Broadband User-Network Interface) [9].

La Fig. 4 muestra el formato de cada una de las unidades de datos manipuladas por estos niveles y subniveles, así como la relación entre ellos.

De forma análoga a lo especificada para el interfaz FDDI, en el interfaz ATM el encapsulamiento de los datos procedentes de los niveles superiores se realiza en base al encapsulamiento LLC-SNAP tal y como lo recomienda el RFC 1209 [11], utilizando el servicio LLC sin conexión y sin reconocimientos de la Clase 1 cuyas tramas de información (tramas UI) incluyen una cabecera SNAP, de la misma forma que se refleja en la Fig. 3.

2.3 Módulo Stack IP

Dentro de este módulo se han agrupado todas aquellas entidades de protocolo de la familia Internet que, por un lado, hacen posible el encaminamiento de los datagramas IP entre uno y otro interfaz físico y, por otro lado, permiten la gestión remota del Adaptador. Estas entidades de protocolo son: IP, ARP, ICMP, UDP y SNMP.

La funcionalidad específica de cada una de estas entidades de protocolo se recoge en los RFCs 768, 791, 792, 826 y 1157 [12][13][14][15][16].

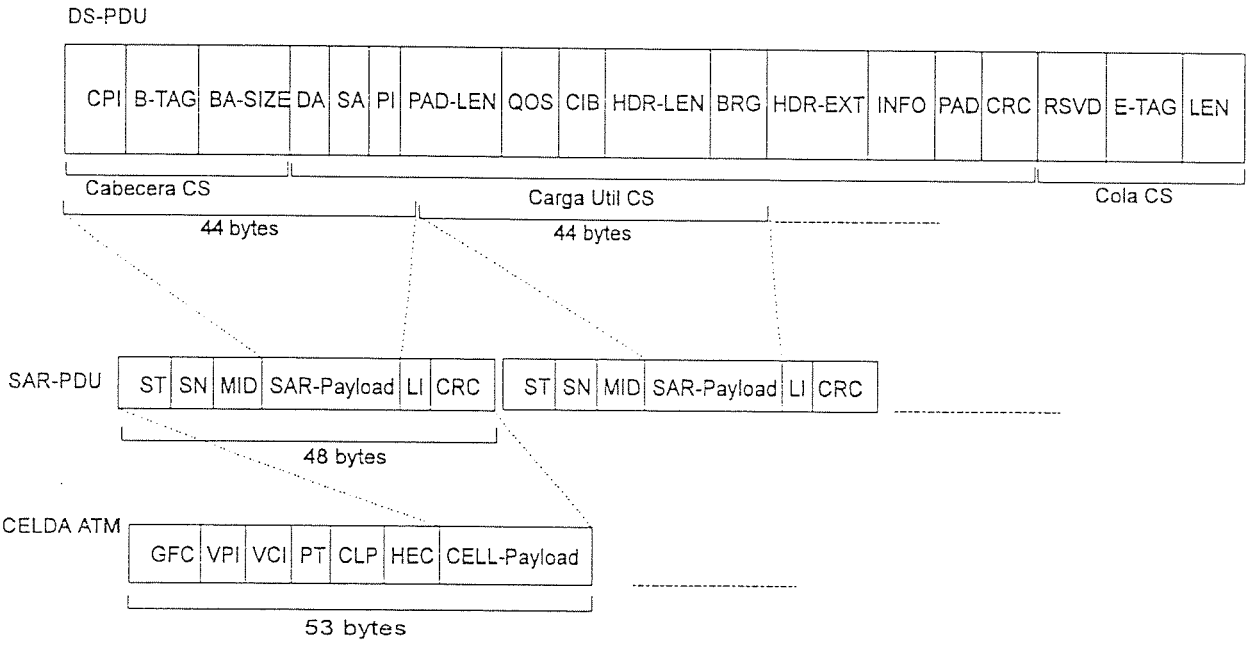


Figura 4. Formato de las tramas por niveles

3. Gestión local y remota del adaptador

La Fig. 5 muestra la arquitectura funcional de los módulos que permiten la gestión, control y mantenimiento del Adaptador tanto de forma local por parte de un operador como de forma remota respondiendo a las órdenes recibidas de un gestor de red.

En líneas generales, el Agente SNMP junto con la entidad de protocolo SNMP y todas las entidades de protocolo subyacentes necesarias para soportar dicho protocolo de nivel de aplicación (entidades UDP, IP, y las asociadas a cada interfaz físico) hacen posible la gestión remota a través del protocolo SNMP.

El Agente Local, por otro lado, permite realizar las mismas operaciones de gestión que las que es posible realizar desde una gestor de red remoto a través de SNMP, más algunas funciones de operación y mantenimiento adicionales.

Finalmente, para hacer posible la gestión y control, tanto de forma local como remota, del Adaptador, éste mantiene una Base de Información de Gestión o MIB adecuada al conjunto de entidades de protocolo y a la funcionalidad que proporciona. Esta MIB no es otra cosa que un almacén virtual de información de gestión en la que se almacenan los objetos gestionables del Adaptador, objetos que mediante los mecanismos de acceso adecuados son accesibles tanto por parte del Agente SNMP como por el Agente Local.

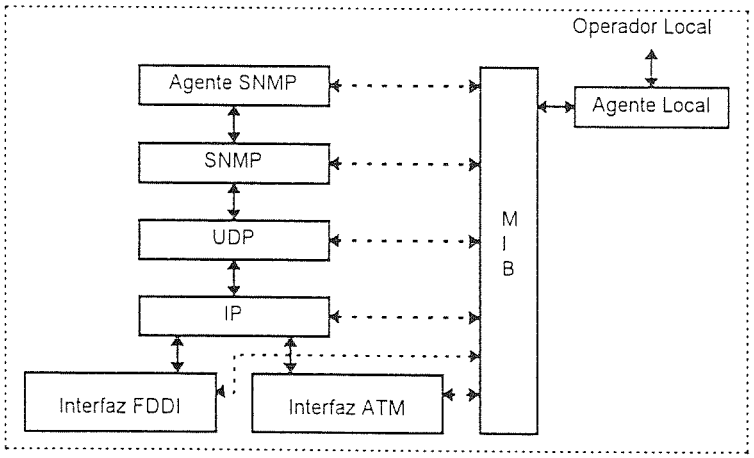


Figura 5. Arquitectura de las entidades de gestión, control y mantenimiento del Adaptador

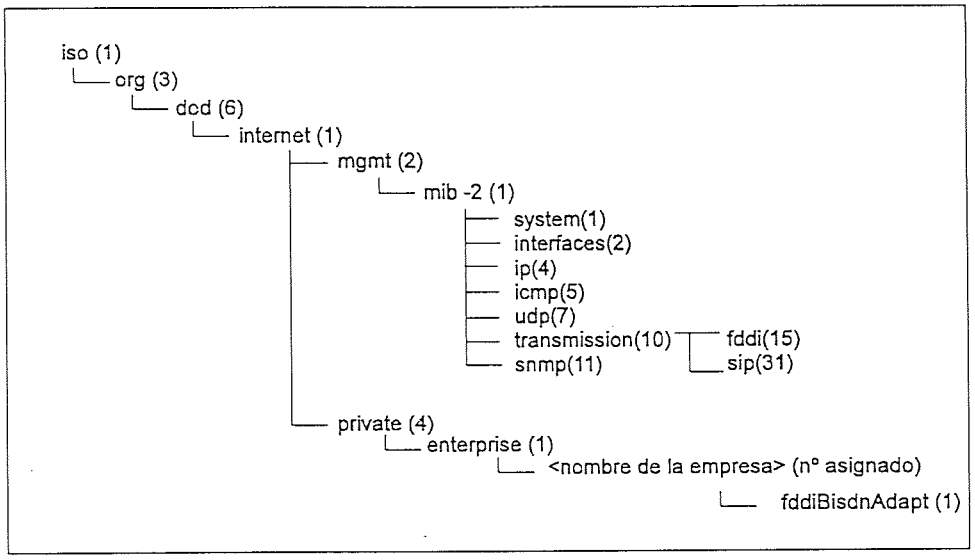


Figura 6. Estructura jerárquica de gestión de la MIB del Adaptador FDDI-ATM.

Teniendo en cuenta la arquitectura de red del Adaptador, éste deberá soportar una MIB que incluya todos aquellos grupos de objetos definidos en el RFC 1213 [17] como de obligada implementación, siempre y cuando tengan aplicabilidad en el Adaptador. Estos grupos son: *system, interfaces, ip, icmp, udp y snmp*. Además la MIB deberá incorporar grupos de objetos definidos en otros RFCs y que se consideran de aplicabilidad al Adaptador. Estos son: el grupo *fddi*, definido en el RFC 1285 [18] y el grupo *sip*, definido en el RFC 1304 [19]. Por último es posible incorporar grupos de objetos no estandarizados, dentro de lo que se denomina MIB privada. Esto permite la introducción de objetos considerados de utilidad en el caso particular del Adaptador FDDI-ATM. Para la creación de esta MIB privada es necesario solicitar al organismo responsable de la asignación de Identificadores de Objetos privados (IANA) la asignación de un identificativo propio a partir del cual crear una estructura de objetos de gestión particularizada.

La Fig. 6 recoge la estructura jerárquica de gestión de la MIB que soportaría el Adaptador FDDI-ATM.

Cada entidad funcional del Adaptador será la responsable de mantener aquella parte de la MIB que le corresponda. Tanto la entidad de aplicación Agente SNMP como la entidad de aplicación Agente Local accederán a la MIB para dar respuesta a las peticiones de gestión recibidas de un gestor de red o del operador local respectivamente. Las peticiones de gestión recibidas se reducen a operaciones de tipo *Set* (modificación) o *Get* (lectura) sobre objetos de la MIB. En determinados casos la modificación de un determinado objeto de la MIB puede suponer el desencadenamiento de una determinada acción sobre el Adaptador. Existe por último la posibilidad de que la aplicación de gestión genere mensajes de gestión de forma asíncrona ante la ocurrencia de algún evento singular en el Adaptador. A este tipo de mensajes se les conoce con el nombre de *Traps*.

4. Conclusiones

En definitiva, un equipo desarrollado en base a esta solución permitiría ampliar el ámbito de cobertura de este tipo de redes, hasta ahora limitado a los 100 Km., a todo el área geográfica cubierta por la red de banda ancha. Las únicas restricciones que impone esta solución son las derivadas de su propia arquitectura, como son:

- Los nodos de las redes FDDI que se intercomunican deberán incorporar una arquitectura de red acorde a la del Adaptador (protocolo IP como protocolo de nivel de red).

- Los Adaptadores FDDI-ATM en ambos extremos de la conexión semipermanente deberán implementar el mismo encapsulamiento IP. En nuestro caso ambos deberán incluir el encapsulamiento LLC-SNAP especificado y como Subcapa de Covergencia, la capa DS del servicio SMDS de Bellcore.

Apéndice I: Aplicación práctica

La solución expuesta a lo largo del presente artículo para el desarrollo de un Adaptador FDDI-ATM ha sido aplicada al equipo AFTER (Adaptador Flexible de Terminales) desarrollado dentro del Programa PLANBA (Plan de Acción Nacional para la I+D en Comunicaciones Integradas de Banda Ancha) promovido por la Dirección General de Telecomunicaciones durante el periodo 1992-1995.

El objetivo del Programa PLANBA ha sido incentivar el desarrollo de las tecnologías de telecomunicaciones de Banda Ancha en las industrias y centros de investigación españoles a través de 15 proyectos integrados en torno a un demostrador de Red Experimental de Comunicaciones de Banda Ancha (RECIBA) aportado por Telefónica. Cada uno de estos proyectos ha sido llevado a cabo por un consorcio de empresas y centros de investigación, tanto públicos como privados, responsable de desarrollar uno o varios elementos de la red.

Los 15 proyectos PLANBA se han agrupado en tres bloques: aplicaciones, terminales y adaptadores, e infraestructuras. Dentro del grupo de terminales e infraestructuras se incluye el proyecto AFTER. Este proyecto ha tenido como objetivo el desarrollo de un equipo capaz de interconectar redes FDDI y líneas de velocidad constante a 2 y 34 Mbit/s a la red pública de Banda Ancha.

La Fig. 7 muestra la arquitectura interna y el diagrama de interconexión del equipo AFTER. Consta de un bus interno basado en tecnología ATM y soportado por un bastidor al que se conectan las distintas tarjetas según las necesidades de la instalación a la que vaya destinado. Un bus estándar VME permite la comunicación entre tarjetas para las tareas de gestión y configuración. Las tarjetas que incorpora el equipo son:

- Tarjeta Interfaz de Red (TIR) que realiza el acceso a la red ATM.
- Tarjeta de Gestión (TCPU) que permite la supervisión y configuración, tanto local como remota de todo el equipo.
- Tarjeta G.703 (TVC) que ofrece al usuario el interfaz de línea G.703 a dos posibles velocidades: 2 y 34 Mbit/s.
- Tarjeta Interfaz FDDI (TIF) que permite la interconexión del equipo a un anillo de fibra óptica de una red FDDI.
- Tarjeta Adaptadora FDDI-ATM (TAD) que realiza las funciones de encaminamiento del tráfico entre la red FDDI y la red de Banda Ancha.

Cada equipo AFTER deberá incorporar una tarjeta TIR y una TCPU, y podrá incluir varias tarjeta TVC y varias parejas TFDDI-TADPT, todo ello en función del tamaño del bastidor.

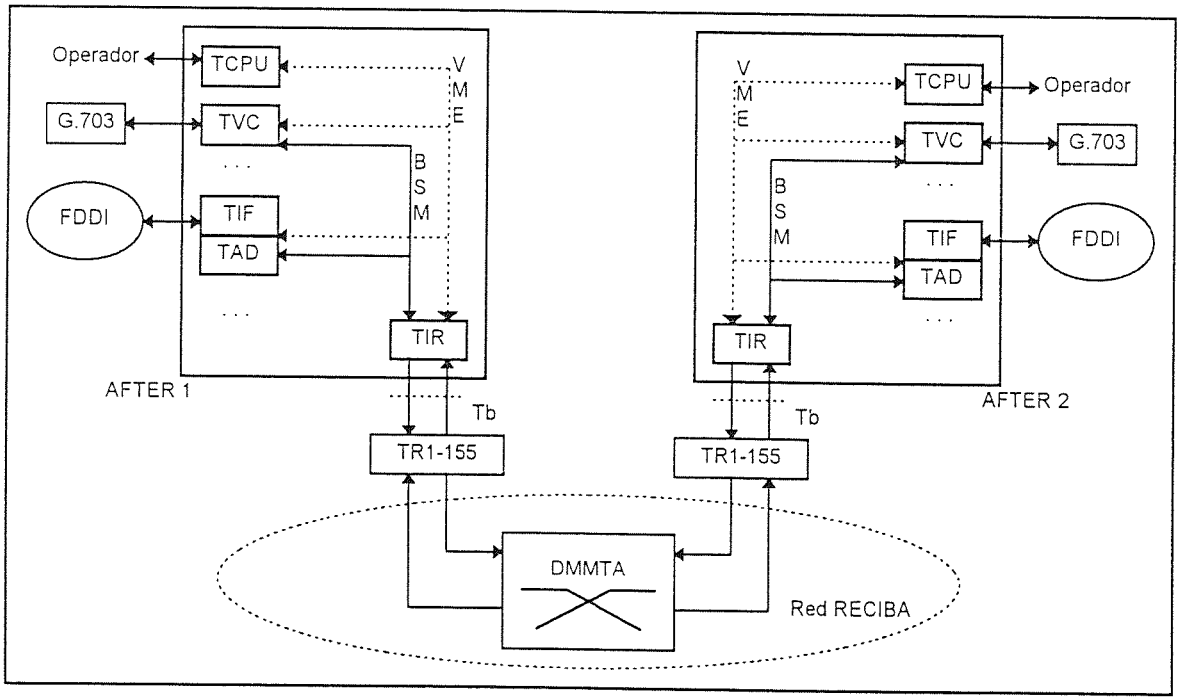


Figura 7. Arquitectura interna y diagrama de interconexión del equipo AFTER

Con la arquitectura interna definida para el equipo AFTER, el Adaptador FDDI-ATM está constituido, como puede deducirse, por una tarjeta TIF más una tarjeta TAD interconectadas directamente entre sí a través de un conector directo entre ambas. La arquitectura de red del adaptador queda de esta forma implantada entre ambas tarjetas tal y como muestra la Fig. 8.

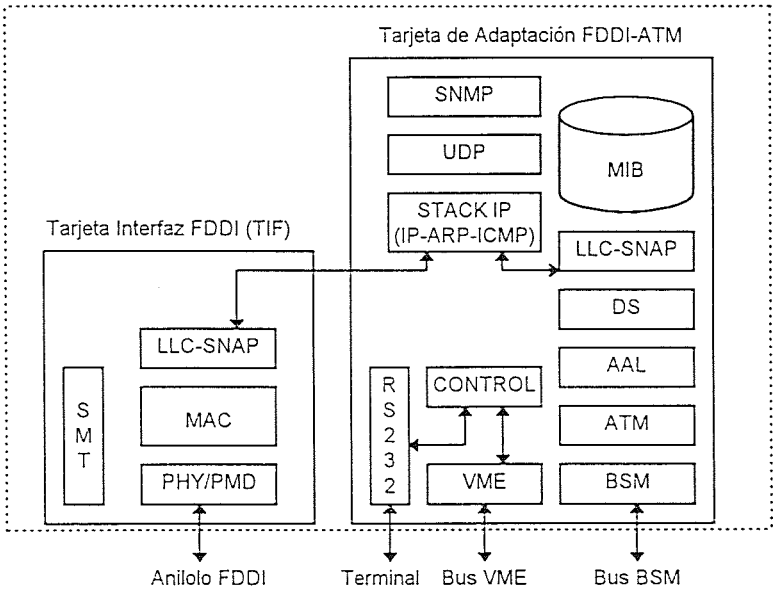


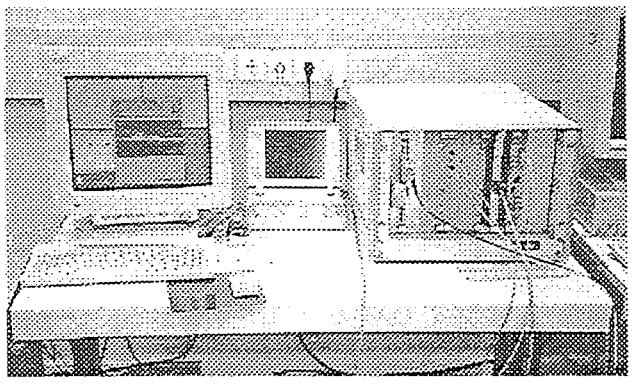
Figura 8. Arquitectura de red del Adaptador FDDI-ATM en el equipo AFTER

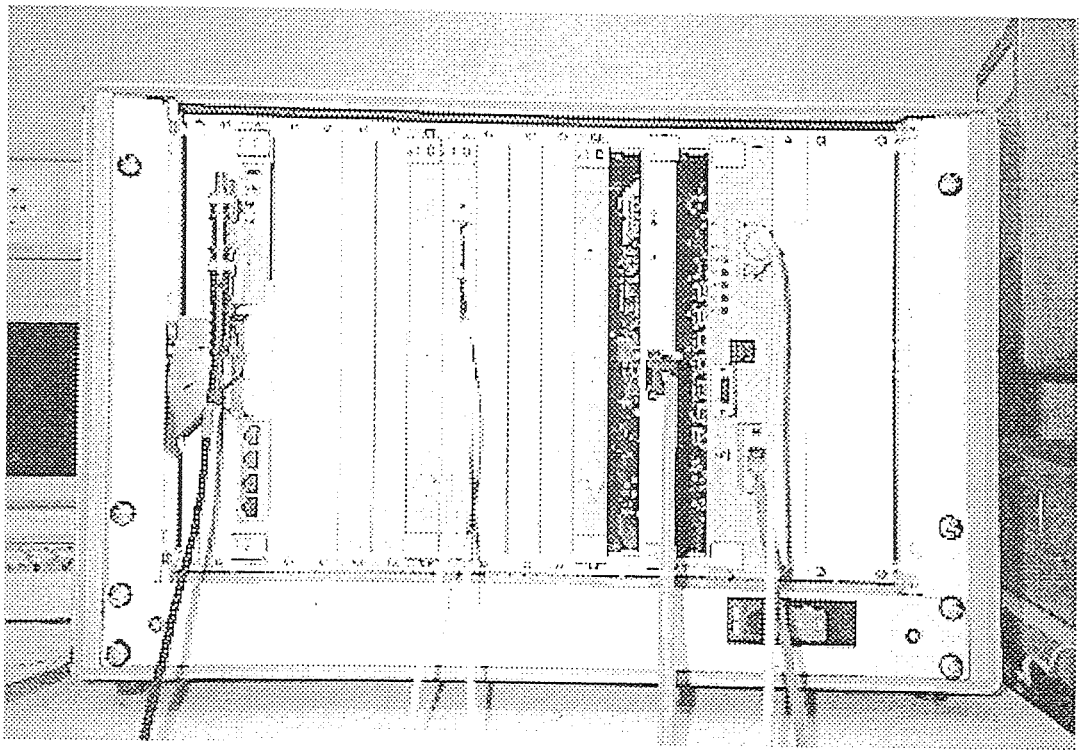
El consorcio formado para el desarrollo del equipo AFTER estuvo formado por las siguientes entidades:

- ALCATEL SESA, como coordinador del proyecto y desarrollador de la Tarjeta FDDI.
- TELEFONICA I+D, como entidad desarrolladora de la Tarjeta G.703 y de la Tarjeta Interfaz a la Red de Banda Ancha.
- LABEIN y ROBOTIKER, como entidades desarrolladores de la Tarjeta de Adaptación de tráfico FDDI-ATM y de la gestión remota y local de este Adaptador.
- TCP S.I., como entidad desarrolladora de la Tarjeta VME de Gestión y Control del equipo AFTER.
- UPC DAC, como entidad participante en la especificación del equipo.
- UPM DIT, como entidad participante en el desarrollo y ejecución de un sistema de simulación del equipo.

Entre Octubre de 1995 y Enero de 1996 se realizaron las pruebas de integración del equipo AFTER, así como las pruebas de campo correspondientes conectando el equipo a la red RECIBA de Telefónica.

Las siguientes fotografías muestran dos instantáneas del equipo AFTER en el laboratorio de TIDSA durante la realización de las pruebas de integración del equipo.





Los resultados obtenidos fueron totalmente satisfactorios y fueron demostrados durante los días 1 y 2 de Febrero en el Palacio de Telecomunicaciones de Madrid en las Jornadas PLANBA organizadas por la Dirección General de Telecomunicaciones y en las que se pudieron ver además de los resultados del Proyecto AFTER, los resultados del resto de los proyectos enmarcados dentro del Programa PLANBA.

En la actualidad existe un equipo AFTER instalado en el laboratorio del Centro de Comunicaciones de Banda Ancha (CCBA) en la Universidad Politécnica de Cataluña, dentro del departamento de Arquitectura de Computadores (DAC-UPC).

El laboratorio está enlazado a la red piloto ATM paneuropea JAMES a través del nodo ATM en Barcelona de la red ATM de Telefónica GIGACOM por medio de un acceso de fibra óptica monomodo, tal y como muestra de Fig. 9.

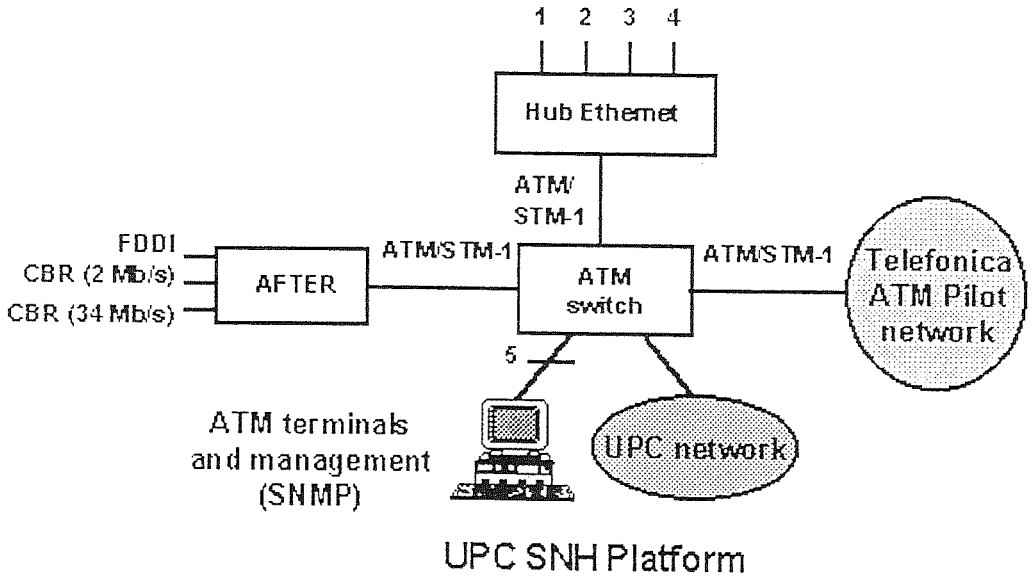


Figura 9. Estructura de interconexión del equipo AFTER en el DAC de la UPC.

Agradecimientos

Los autores del presente articulo quisieran agradecer a todas las entidades participantes del proyecto AFTER su contribución en el desarrollo del equipo AFTER, que ha hecho posible demostrar la viabilidad práctica de la solución expuesta.

Referencias

[1] "FDDI Physical Layer Protocol (PHY)". *ISO/IEC 9314-1* .
[2] "FDDI Media Access Control (MAC)". *ISO/IEC 9314-2.*
[3] "FDDI Physical Layer Medium Dependent (PMD)". *ISO/IEC 9314-3.*
[4] "Logical Link Control". *ISO 8802-2.*
[5] "FDDI Station Management (SMT) Rev. 6.2". *ANSI X3T9/90.*
[6] "B-ISDN ATM Layer Specification". *Recomendación I.361 del CCITT.*
[7] "B-ISDN ATM Adaptation Layer Functional Description". *Recomendación I.362 del CCITT.*
[8] "B-ISDN ATM Adaptation Layer Specification". *Recomendación I.363 del CCITT.*
[9] "BELLCORE SMDS". *TR-TSV-000772. Issue, 01/05/91.*
[10] "Proposed standard for the transmission of IP datagrams over FDDI networks". *RFC 1188.*
[11] "Transmission of IP datagrams over the SMDS service". *RFC 1209.*
[12] "User Datagram Protocol (UDP)". *RFC 768.*
[13] " Internet Protocol (IP)". *RFC 791.*
[14] "Internet Control Protocol (ICMP)". *RFC 792.*
[15] "An Ethernet Address Resolution Protocol (ARP)". *RFC 826.*
[16] "Simple Network Management Protocol (SNMP)". *RFC 1157.*
[17] "Management Information Base for Network Management of TCP/IP based internets: MIB-II". *RFC 1213.*
[18] "FDDI Management Information Base". *RFC 1285.*
[19] "Definitions of Managed Objects for the SIP Interface Type". *RFC 1304.*

Planificación de la red de señalización para servicios de Banda Ancha

Florentino Fernández Cuesta
Email: cuesta@tid.es
Tlfno: +34 1 337 47 48
Fax: +34 1 337 46 02

Federico Lozano Rozalén
Email: flr@tid.es
Tlfno: +34 1 337 98 74
Fax: +34 1 337 46 02

Telefónica I+D, Emilio Vargas 6, 28043 Madrid, SPAIN

Abstract

The fact that Broadband Networks signalling requires additional procedures and funcionalities, beyond the ones used in Narrowband Networks, as well as the wide spectrum of servicoes that B-ISDN can support, and the flexibility of structure that the ATM technology offers, leads to the development of new planning and dimensioning methods for resources devoted to carry the Broadband Networks signalling.

In order to aproach the problem, we propose a first strategy of study that consist of characterising and evaluating the signalling requirements and the functinal elements used by each one of the B-ISDN services, then define the architecture of the ATM network that supports the B-ISDN, and finally define the scenarios of demand in order to be able to dimension the resources allocated for the signalling data transport in each network element.

1 Introducción

A la hora de abordar la planificación de una red comercial RDSI-BA basada en tecnología MTA (Modo de Transferencia Asíncrono) se hace necesario analizar los mecanismos de señalización requeridos para soportar los nuevos servicios de Banda Ancha. Este análisis permitirá, en una primera etapa, desarrollar nuevos métodos de evaluación del tráfico de señalización en la red para luego, en una segunda etapa, determinar las estrategias de planificación que permitan optimizar tanto la arquitectura como la propia estructura de la red de señalización de Banda Ancha.

De acuerdo con el anterior planteamiento este artículo cubre los siguientes aspectos:

Selección de los servicios de Banda Ancha para los que se prevé una mayor demanda a corto plazo y agruparlos atendiendo a sus requisitos de señalización.

Análisis del estado actual de la definición de las posibles estructuras de red MTA así como las interfaces y protocolos de señalización de las mismas.

Descripción de los requisitos de señalización para los grupos de servicios seleccionados.

Estimación del tráfico de señalización generado por cada grupo de servicios, teniendo en cuenta la demanda y penetración de los mismos.

2 Servicios MTA

Una gran parte de los servicios MTA es la formada por los servicios de redes digitales de banda ancha (RDSI-BA), la recomendación UIT-T I.210 establece cuales son estos servicios definiendo dos amplias familias de acuerdo al nivel de prestación con el que se ofrecen por el proveedor o administrador de la red.

Servicios portadores: engloba aquellos servicios que, ofreciendo una capacidad de transferencia entre dos usuarios RDSI-BA, sólo proveen *funciones de capa inferior*. (niveles 1 al 3 de la arquitectura OSI). Cada servicio portador está caracterizado por un conjunto de atributos de capa inferior clasificados en tres categorías: atributos de transferencia de información, atributos de acceso y atributos generales. Las dos primeras categorías definen cada una de las capacidades portadoras de la Red.

Sólo dos tipos de servicio portador de banda ancha han sido definidos de una forma estable en la UIT-T: Servicio portador de banda ancha orientado a conexión (recomendación F.811) y Servicio portador de banda ancha no orientado a conexión (recomendación F.812); el resto de servicios portadores considerados se encuentran en fase de estudio: Servicio portador de banda ancha para servicios nx64 Kbits/s (F.N64), Servicio portador de banda ancha para transporte de circuitos no estructurado (F.UCTBS), etc.

Teleservicios: proporcionan plena capacidad de comunicación entre usuarios RDSI-BA por medio de funciones de terminal y de red y dependiendo de la naturaleza del servicio de funciones proporcionadas por centros especializados. Todas estas funciones corresponderían tanto a ***funciones de capa inferior como funciones de capa superior***. Un teleservicio hace uso de un servicio portador o un pequeño número de ellos.

Los teleservicios que permiten la transferencia de diversos tipos de información (***componentes del servicio***) establecidos en una red basada en MTA pueden ofrecer varios canales virtuales, uno por cada componente del servicio. Así un teleservicio viene definido por un conjunto de atributos que se aplican de forma global a todos los componentes del servicio junto con un subgrupo de atributos que caracterizaran cada una de las componentes.

En la recomendación UIT-T I.211 se definen las siguiente clases de teleservicios de banda ancha:

Servicios interactivos que se dividen a su vez en tres clases:

- **servicios conversacionales**: que proporcionan en general los medios para una comunicación bidireccional con transferencia de información en tiempo real de extremo a extremo. La información es producida por el usuario emisor y se dirige a uno (punto a punto) o más (punto a multipunto) copartícipes de la comunicación situados en el lado receptor. Ejemplo de este tipo de servicios son la videotelefonía, la videoconferencia y la transmisión de datos a alta velocidad y siendo muy diversas sus posibles aplicaciones (tele-enseñanza, tele-compra, etc.)

- **servicios de mensajería**: que ofrecen la comunicación entre usuarios individuales por medio de unidades de almacenamiento con funciones de buzón electrónico y/o tratamiento de mensajes. Son ejemplos de aplicación para este tipo de servicios aquéllas en las que existe un servidor de mensajes como medio para la comunicación entre usuarios (correo electrónico).

- **servicios de consulta**: El usuario de los servicios de consulta puede consultar la información almacenada en centros de información para uso público. Esta información se enviará al usuario solamente si la solicita controlando el instante en que debe comenzar una secuencia de información. Todo servicio de acceso a una información almacenada con control del usuario es un ejemplo de servicios de consulta, servicio de consulta de datos, de imágenes, de Vídeo (vídeo bajo demanda), etc.

Servicios de distribución que proporcionan un flujo continuo de información que es distribuido desde una fuente central a un número ilimitado de receptores. Dependiendo de cual es el control del usuario sobre la presentación de la información existen:

- **Servicios de distribución sin control de la presentación por el usuario**: El usuario puede acceder a este flujo de información, sin la posibilidad de determinar en qué instante debe comenzar la difusión de la cadena de información

- **Servicios de distribución con control de la presentación por el usuario**: El usuario puede tener acceso individual a la información distribuida cíclicamente, y controlar el instante de comienzo y el orden de presentación. El ejemplo clásico de este tipo de servicio es la distribución de Televisión con y sin control del usuario.

Para establecer la agrupación de servicios RDSI-BA atendiendo a su señalización se han analizado cada una de estas clases teniendo en cuenta los siguientes aspectos:

- características de las llamadas en lo que se refiera tipos de conexiones o componentes que emplean que utilizan

- etapas diferenciadas en la tramitación del acceso a un servicio completo

- entidades de red que intervienen en la prestación del servicio

- interacciones entre el equipo de usuario y las entidades de red, que se producen en la prestación del servicio

Como consecuencia de esta análisis se han determinado la siguientes clases de servicios RDSI-BA representativas en cuanto a los requisitos funcionales de señalización MTA.:

- Servicios conversacionales punto a punto
- Servicios conversacionales punto a multipunto
- Servicios de consulta

Con las dos primeras clases se agrupan los servicios RDSI-BA de tipo conversacional mientras la tercera clase corresponde a los servicios de consulta que se ofrecen bajo demanda.

Los servicios de mensajería no se consideran por un doble motivo: no parece que a corto plazo estos servicios sean ofrecidos directamente por la propia red MTA y por otra parte la señalización empleada, en todo caso, para soporta estos servicios sería similar a la contemplada para los servicios conversacionales punto a punto.

Los servicios de distribución no se han contemplado pues el establecimiento de sus componentes o conexiones se realizará normalmente de forma permanente y por tanto empleando el plano de gestión sin hacer uso de la señalización.

En cuanto a la señalización empleada para soportar servicios portadores establecidos bajo demanda sería similar a la contemplada para los servicios conversacionales punto a punto.

Por tanto, las tres clases de servicios antes mencionadas cubrirían de forma global los diferentes esquemas de señalización MTA empleados para soportar los servicios RDSI de banda ancha.

3 Estructura de red MTA

Para hacer una estimación cuantitativa del tráfico de señalización que se generará en el uso normal de una red MTA es preciso, dada la amplia gama de posibles configuraciones de red que permite la tecnología MTA, plantear escenarios de referencia que tengan en cuenta los siguientes aspectos:

- estructura de red formada por los equipos de conmutación MTA y aquellos servidores necesarios para la provisión de cada servicio en cada escenario.

- arquitectura de red estableciendo las interfaces entre los diferentes elementos de red de acuerdo a las funciones asignadas a cada uno de ellos.

Ambos aspectos son decisivos a la hora de identificar los trayectos de señalización y la carga de

tráfico correspondiente a una demanda dada de los diferentes servicios para un número determinado de usuarios de la red.

La estructura de red MTA contemplada está formada por dos niveles, un primer nivel (Nivel de tránsito) formado por nodos de tránsito conectados de forma totalmente mallada y un segundo nivel (Nivel de Acceso) compuesto por una serie de nodos de acceso conectados en estrella a un solo nodo de tránsito.

Además de estos nodos MTA, la red de Banda Ancha tendrá que contar con los diversos servidores necesarios para soportar los servicios considerados. Algunos de estos servicios se pueden ofrecer por medio de elementos especializados o bien por medio de funciones de inteligencia de red. En este último caso es conveniente considerar en los escenarios genéricos aquellos elementos de Red Inteligente involucrados en la provisión de estos servicios.

A continuación se analizan los elementos de red que pueden emplearse para proporcionar cada uno de los servicios de Banda Ancha seleccionados.

- Servicios conversacionales: No se prevé la necesidad de elementos de red especiales

- Servicios de consulta: se pueden considerar dos posibilidades, bien un elemento de acceso al servicio o bien los elementos de Red Inteligente: Modulo de Funciones Especiales y Centro de Inteligencia de Red, si el servicio se supone soportado por la Red Inteligente.

Mientras que en el caso de los nodos MTA (nodos de acceso y de tránsito), el número de elementos de red depende fundamentalmente del número de abonados y del tráfico ofrecido por abonado, el número de elementos de red requeridos para proveer la funcionalidad de los diferentes servicios dependerá de la estructura de red a nivel del servicio que se implante.

Así, por ejemplo, para un servicio dado, se puede usar un escenario de referencia con estructura centralizada en que exista un solo elemento de red con las funciones dedicadas a ese servicio que daría cobertura a la totalidad de usuarios, en tanto que para un escenario de referencia con estructura de servicio distribuida se puede considerar que existen varios elementos de red dedicados a ese servicio para cubrir diferentes zonas de la red MTA.

En la estructura de red MTA contemplada la señalización correspondientes al nivel de acceso se realiza mediante la propia red MTA a través de la interfaz UNI (User Network Interface) con la pila de protocolos mostrada en la figura 1.

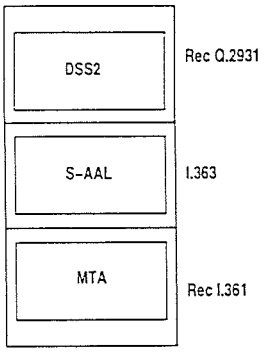


Figura 1: Pila de protocolos para la interfaz UNI

Es en el nivel de tránsito a través de la interfaz NNI (Node Node Interface) donde se ha de evaluar si la señalización (protocolo B-ISUP) se cursa mediante la propia red MTA o mediante la ya consolidada red de señalización por canal común N° 7 (RSCC#7). Para ello se han de considerar dos aspectos:

- Posibilidad técnica de que la RSCC#7, concretamente el nivel PTM-3 (Parte de Transferencia de mansajes) soporte el protocolo B-ISUP.

- Evaluarción de la capacidad de la actual RSCC#7 para soportar el tráfico de señalización ofrecido por los servicios de Banda Ancha.

En el siguiente apartado se analiza el soporte del protocolo B-ISUP por parte de la RSCC#7.

3.1 Soporte de la RSCC #7 del protocolo B-ISUP

La consideración de la RSCC #7 como posible alternativa para soportar las conexiones de señalización entre los elementos de la red MTA aparece contemplada en la recomendación de la UIT-T Q.2010 (aprobada el 7 de febrero de 1995) la cual dice textualmente *"En la NNI puede utilizarse la red MTA o, como una opción nacional, la red existente del sistema de señalización N. 7 para la transferencia de información de señalización"*. Esto permite que si un operador lo considera conveniente emplee su RSSC # 7 para transportar la señalización de Banda Ancha. La propia recomendación hace un primer planteamiento de cuales serían las pilas de protocolos consideradas en ambos casos para soportar el protocolo B-ISUP e incluye el siguiente gráfico:

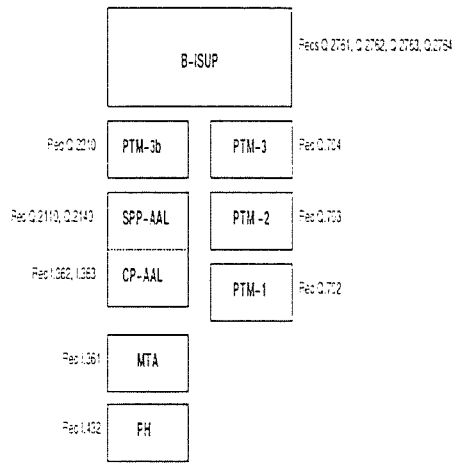


Figura 2: Pilas de protocolos para la interfaz NNI

En ambas pilas se propone que B-ISUP sea soportado por el protocolo PTM de capa 3 asumiendo que los servicios de esta capa son independientes de los mecanismos de transporte inferiores.

El protocolo PTM-3 fue inicialmente especificado por la UIT-T para soportar la parte de señalización del usuario de banda Estrecha (ISUP) [Rec. Q.704] sin que en posteriores versiones de esta recomendación se hayan establecido modificaciones para el caso en que PTM-3 este dedicado a soportar B-ISUP. Sin embargo recientemente tanto en la propia UIT-T [Q.2210] como en diferentes documentos de arquitectura elaborados por proyectos internacionales (EURESCOM[1], INSIGNIA [2]) se contempla un segundo protocolo (PTM-3b) para referirse al protocolo PTM de capa 3 soportado por S-AAL y empleado en redes de Banda Ancha para transportar mensajes B-ISUP.

Así, usando esta misma terminología, la red RSSC#7 podría transportar la señalización B-ISUP si el protocolo PTM-3 pudiese soportar directamente B-ISUP o bien si existiesen dispositivos de interfuncionamiento entre PTM-3 y PTM-3b. Este interfuncionamiento podría ser una opción para la migración de la red RSCC#7 hacia una red de señalización basada en protocolos de Banda Ancha, si bien es cierto que todavía tal dispositivo no ha sido considerado en la recomendación que describe las disposiciones generales para el interfuncionamiento entre la RDSI de Banda Ancha y la RDSI de Banda Estrecha [I.580].

En cuanto a la otra posible alternativa, PTM-3 podría, en teoría, transportar los mensajes B-ISUP (pues las primitivas con la capa superior son las mismas en PTM-3 y PTM-3b [Q.2210]) a menos que la longitud de estos excediesen la limitación del PTM-2 de 272 octetos por mensaje. Incluso para solventar esta limitación, las recomendaciones que especifican el protocolo B-ISUP para el grupo de capacidades CS 1 [Q.2764] y CS2.1 [Q.2721] establecen la opción nacional de segmentación simple que permitiría transportar mensajes B-ISUP de hasta 544 octetos.

Por otra parte, en el caso de relaciones de señalización entre elementos de Red Inteligente, se establece el empleo del protocolo PTM-3b para soportar los mensajes B-INAP por medio de los protocolos PACT (Parte de aplicación de las capacidades de Transacción) y PCCS (Parte de Control de la Conexión de Señalización) [2]. Estos dos protocolos podrían facilitar el empleo de la RSSC # 7 para transportar mensajes B-INAP pues permitirían que la capa PTM-3 fuese totalmente transparente para las entidades de inteligencia de red.

4 Señalización MTA de los servicios contemplados

En lo siguientes apartados se realizará la descripción formal de la señalización MTA empleada para soportar cada una de las clases de servicios RDSI-BA seleccionadas. Cada descripción se realizará siguiendo el planteamiento propuesto en la recomendación UIT-T I.375 para el servicio de banda ancha Vídeo bajo Demanda:

- Identificación de configuración de referencia del servicio y de sus componentes funcionales.
- Descripción de la funcionalidad de los elementos de la configuración de referencia.
- Identificación de puntos de referencia e interfaces entre elementos de red.
- Descripción de los flujos de comunicación, de acuerdo con el escenario de acciones que se desencadenan en el uso del servicio.

4.1 Servicios conversacionales punto a punto

Corresponden a esta clase de servicios RDSI-BA todos aquellos servicios basados en la transferencia de cualquier tipo de información (voz, vídeo, datos) entre dos emplazamientos. La variedad de teleservicios englobados en esta clase es muy amplia, aquí se relaciona algunos de los más destacados:

Telefonía
Videotelefonía
Servicio de transferencia de información
Telefax
Servicio de comunicación de documentos

De igual forma resultarían innumerables las aplicaciones de banda ancha que pueden ser ofrecidas a través de este clase de servicio. En general corresponde a esta clase de servicio cualquier aplicación en la que participen dos usuarios y el establecimiento

de la conexiones MTA entre ellos se realice de forma directa sin más participación que la de los nodos MTA por donde se cursarán la conexiones. Además de la aplicaciones clásicas de telecomunicación entre dos abonados de banda ancha (telefonía, videotelefonía) serían también aplicaciones de esta clase de servicios:

tele-enseñanza
tele-compra
tele-banco
tele-trabajo
tele-juegos

siempre que se realice una conexión directa entre el abonado y el proveedor de la información.

La llamada asociada a esta clase de servicio podrá tener asociada una o más conexiones MTA dependiendo de las necesidades de la aplicación. Así por ejemplo, la aplicación tele-trabajo podría requerir más de una conexión MTA, cada una de ellas dedicada a la transferencia de un tipo de información diferente (voz, datos, vídeo, etc). Cada una de estas conexiones MTA será del tipo punto a multipunto si bien el resto de características de las capas bajas (capacidad de transferencia MTA, Calidad de servicio, capa AAL, bidireccional/unidireccional, simétrica/asimétrica, etc.) variarían de acuerdo al tipo de información que se transfiera por ellas.

4.1.1 Componentes funcionales

Los componentes funcionales que forman parte de un servicio conversacional punto a punto son los siguientes:

- **TE-BA origen**: usuario de banda ancha origen de la conexión,
- **TE-BA destino**: usuario de banda ancha destino de la conexión
- **la red MTA**, en la cual se han considerado dos nodos de conmutación MTA lo que nos permitirá analizar las interfaces internas de la red.

Las interfaces entre usuarios y nodos MTA serán de tipo UNI empleando el protocolo de señalización DSS2 (Q.2971) mientras que las interfaces entre dos nodos MTA será de tipo NNI utilizando el protocolo de señalización para la parte de usuario de banda ancha B-ISUP.

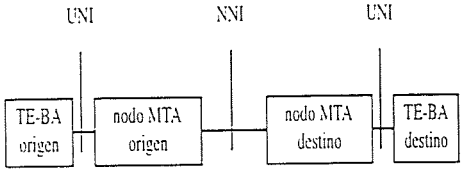


Figura 3: Componentes e interfaces asociados a un servicio conversacional punto a punto

4.1.2 Flujos de comunicación

Aunque estaba previsto que el grupo de capacidades CS 2.1 incluyese el control de llamada con múltiples conexiones, esta capacidad no está recogida, al menos de forma explícita, en la actual versión de las capacidades CS 2.1 para el protocolo B-ISUP (Rec. UIT-T Q.2721.1) y por otra parte la recomendación que establece esta capacidad para el protocolo DSS2 (Q.2968) está en una fase de desarrollo muy provisional.

Por consiguiente, se considera que sólo es posible establecer llamadas monoconexión con lo que cualquier aplicación que requiera múltiples conexiones estas se establecerán y se liberarán de forma aislada siguiendo los flujos de comunicación que a continuación se detallan.

La recomendación UIT-T Q.2650 establece el interfuncionamiento entre la parte de usuario RDSI-BA (B-ISUP) y el sistema de abonados digitales DSS2. El ***establecimiento de una conexión*** admite diferentes variantes en función de cómo el usuario origen informa a la red de la petición de establecimiento (en bloque cuando los dos abonados son de banda ancha o por etapas cuando existe interfuncionamiento con la banda estrecha) y si el usuario destino emite la respuesta de forma automática o enviando previamente un mensaje de aviso.

La siguiente figura muestra de un modo gráfico los flujos de comunicación para el establecimiento y liberación de una conexión asociada a un servicio conversacional punto a punto

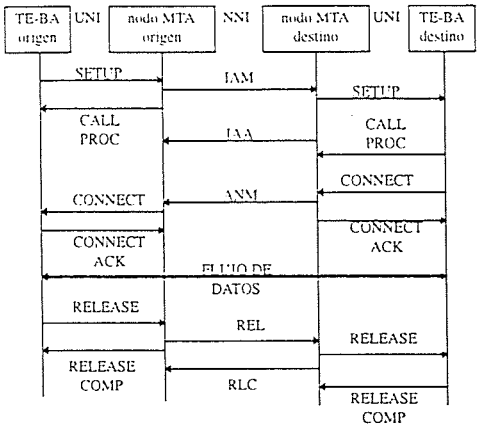


Figura 4: :Flujos de comunicación asociados a servicio conversacionales punto a punto

El flujo de comunicación aquí detallado debería ejecutarse para cada una de las conexiones MTA necesarias para ofrecer el servicio conversacional que se trate. Corresponde, por tanto, este flujo al prototipo de establecimiento y liberación de una conexión bajo demanda punto a punto el redes de banda ancha. Este prototipo sería aplicable no sólo a cada conexión asociada a un servicio conversacional punto a punto sino también a todas aquellas conexiones punto a punto establecidas bajo demanda correspondientes a cualquier servicio portador definido en la re-

comendación F.811 o teleservicios de banda ancha que empleen conexiones punto a punto (por ejemplo los de mensajería con conexión directa entre el usuario y el dispositivo que controle el buzón electrónico).

4.2 Servicios conversacionales punto a multipunto

La clase de servicio conversacional punto a multipunto estaría formada por los servicios RDSI-BA basados en la transferencia de cualquier tipo de información (voz, vídeo, datos) entre más de dos emplazamientos. El teleservicio más representativo de esta clase es la vídeo conferencia (con más de dos participantes) si bien cualquiera de los teleservicios mencionados en el apartado anterior (telefonía, videotelefonía, transferencia de información) pertenecería a esta clase de servicio si fuesen más de dos los usuarios involucrados. De igual modo cualquiera de la aplicaciones reseñadas en el apartado anterior (tele-enseñanza, tele-compra, etc.) emplearía esta clase de servicios si el número de usuarios a los que se le ofrece la información fuese mayor de dos. En general son aplicaciones de esta clase de servicios todas aquellas en las que un servidor distribuye información a varios abonados de forma conversacional (distribución de señales radiofónicas, de datos, de imágenes, etc.).

Cuando la aplicación requiera la transferencia entre usuarios en ambas direcciones sería necesario el establecimiento de conexiones bidireccional punto a multipunto, sin embargo la UIT-T sólo considera conexiones unidireccionales punto a multipunto en el grupo de capacidades CS2.1 [Rec. Q.2721.1]. Por lo tanto, una aplicación de esta clase de servicios en la que cada usuario desea enviar y transmitir datos requería establecer tantas conexiones unidireccionales punto a multipunto como usuarios se hallen envueltos en la comunicación.

Al igual que ocurría con los servicios conversacionales punto a punto no se prevé la capacidad de controlar llamadas con múltiples conexiones por los que si una aplicación requiere la transferencia de varios tipos de información (datos, vídeo, etc.) será necesario establecer de forma individual una conexión para cada tipo de información.

Teniendo en cuenta estos condicionantes, por ejemplo una videoconferencia entre N usuarios con transferencia bidireccional completa entre ellos requerirá N conexiones punto a multipunto unidireccionales para la transmisión de voz e imágenes (establecidas por cada usuario que desee enviar voz e imagen) y otras N conexiones punto a multipunto unidireccionales para la transmisión de datos

4.2.1 Componentes funcionales

ATM Forum [4] y la recomendación UIT-T Q.2722.1 consideran similares aunque no los siguientes componentes funcionales en una conexión punto a multipunto

TE-BA raíz: terminal que inicia la conexión

Nodo PMP : Nodo Punto a multipunto, es un nodo de red donde el flujo de información usa una conexión ya establecida para informar a una nueva rama de la conexión

Nodo LC (*Last Common Node*): Es un nodo de red donde el flujo de información emplea una conexión ya establecida para informar a una nueva rama de la conexión establecida en el lado entrante del nodo, en el lado saliente el flujo de información usa una interfaces no usada por el resto de ramas de la conexión

Nodo SP (*Single Party Node*): nodo de la red donde la conexión se usa en ambos lados por un único participante en la videoconferencia

TE-BA destino, representa a cada uno de los usuarios que reciben la información por cada una de las ramas de la conexión.

La siguiente figura describe cada uno de estos componentes funcionales en una conexión punto a multipunto así como las interfaces entre ellos establecidos (NNI con señalización B-ISUP entre cada dos nodos MTA y UNI empleando el protocolo DSS2 entre cada usuario y el nodo MTA al que está conectado)

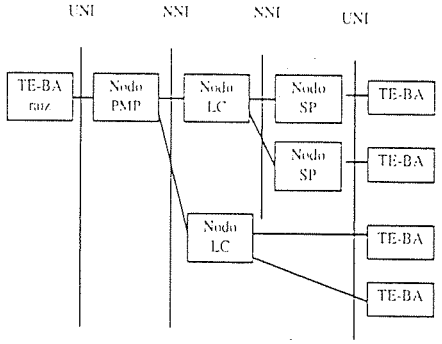


Figura 5: Componentes e interfaces asociados a un servicio conversacional punto a multipunto

4.2.2 Flujos de comunicación

Los nodos SP, son realmente nodos intermedios dentro de la conexión por lo que su papel se reduce a la retransmisión de mensajes, por lo tanto, no se consideran en la descripción de los flujos de señalización entre los componentes. A continuación se detalla el flujo de información asociado la establecimiento y liberación de una conexión unidireccional punto a multipunto entre un usuario (TE-BA raíz) y otros dos participantes en el servicio (TE-BA destino 1 y TE-BA destino 2).

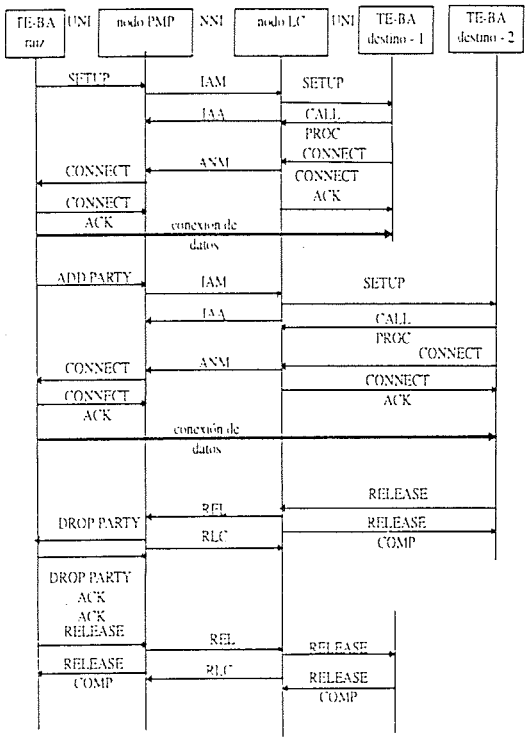


Figura 6:. Flujos de comunicación asociados a un servicio conversacional punto a multipunto

4.3 Servicios de consulta

Pertenecen a esta clase de servicios aquellos servicios que ofrecen al usuario la posibilidad de acceder a la información almacenada en centros de información de forma que es el propio usuario quien controla el instante en que debe comenzar cada secuencia de información. La recomendación I.211 establece como teleservicios pertenecientes a esta clase:
Videotext de banda ancha
Servicio de consulta de vídeo
Servicio de consulta de documentos
Servicio de consulta de datos

La aplicaciones de estos teleservicios serían aquellas en las que el usuario realiza una consulta a un Centro de Información controlando tanto el instante en que va a recibir la información y como que la información que va a recibir. Tele-compra y tele-capacitación serían ejemplo de este tipo de aplicaciones siempre que estén orientadas la realización de consultas con control del usuario. En cualquier caso la aplicación más representativa de esta clase de servicios es la de Vídeo bajo Demanda y en ella se ha basado la caracterización de la señalización que conlleva el empleo de esta clase de servicios.

4.3.1 Componentes funcionales

Aunque diversos organismos de estandarización (ETSI, UIT-T, DAVIC, ATM Forum) y proyectos europeos (Eurescom, INSIGNIA) están o han desarrollado arquitecturas de red para ofrecer el servicio Vídeo bajo Demanda, la mayoría de las arquitecturas consideran una misma configuración con parecidos componentes funcionales, lo que permitirá definir una arquitectura genérica que nos servirá de base para establecer los principales los componentes funcionales asociados a cualquier servicios de consulta.

A continuación se detallan los componentes funcionales involucrados señalización MTA necesaria para ofrecer los servicios de consulta:

Equipos de usuario (TE-BA)

Formado por todo el equipamiento bajo control del usuario del servicio, básicamente se compone de un equipo terminal y un dispositivo Set Top Box (STB) necesario para controlar la interfaz con el equipo terminal, mantener el control remoto de la información, controlar las interfaces humanas (ratón, etc.) e interconectar al usuario con la red.

Red central MTA (Core Network)

La red central MTA es la encargada de conectar todos los otros componentes funcionales entre si, para ello proporciona funciones de transmisión tanto de señalización como de transferencia de datos.
Acceso al Servicio (AS)

Este componente funcional es la interfaz entre los usuarios y los Centro de Información encargada de identificar y conectar a cada usuario con el correspondiente centro de información. Un Acceso al Servicio sirve a un grupo cerrado de usuarios, de forma que su número de se incrementará conforme aumente el número de usuarios. Las principales funcionalidades del Acceso al Servicio son la de ofrecer al usuario un primer nivel de selección entre todos los proveedores de servicio disponibles, así como negociar y definir la identidad de cada usuario.

Centro de información

El Centro de Información es el suministrador de los datos siendo también responsable de gestionar la información y permitir al usuario el control interactivo de la información que desea consultar.

Estos son los componentes funcionales básicos para ofrecer un servicio de consulta, adicionalmente, existen otros componentes orientados fundamentalmente al plano de gestión como son el Gestor de Red responsable de la gestión de la red MTA y el componente de operaciones de servicio que gestiona diversos aspectos del servicio (medidas, tarificación, etc), estos componentes no se han incluido en el esquema inicial pues no afectan a la señalización.

Tanto el Centro de Información como el Acceso al Servicio son considerados por la red MTA como usuarios de banda ancha por los que la interfaz

de señalización empleada entre cada uno de ellos y los nodos MTA será la misma que entre el usuario de banda ancha (TE-BA) y los nodos MTA, UNI con el protocolo DSS2. La interfaz entre los nodos que forman la red MTA será NNI empleando el protocolo B-ISUP.

La siguiente figura muestra los componentes e interfaces contemplados para los servicios de consulta. Se han considerado dos nodos en la red MTA, el usuario del servicio está conectado a un de ellos (nodo MTA origen) mientras que el Acceso al Servicio y el Centro de Información están conectados al otro (nodo MTA destino). De esta forma que será posible describir el flujo de mensajes B-ISUP entre nodos MTA.

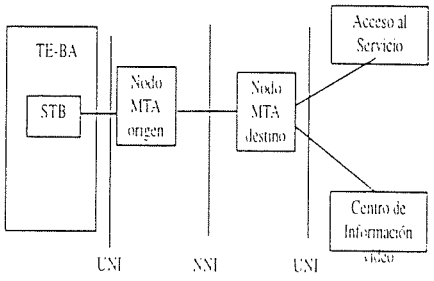


Figura 7: Componentes y puntos de referencia e interfaces asociados a un servicio de consulta

4.3.2 Flujos de comunicacción

ETSI y UIT-T definen tres niveles de control para proporcionar el servicio Vídeo bajo Demanda:

Control de conexión, es el control básico para establecer una conexión simple entre dos puntos extremos. Este nivel de control siempre está ubicado en el plano de control y por tanto usa los mensajes de señalización.

Control de sesión, que engloba aquellas capacidades necesarias para gestionar varios recursos simultáneamente, así por ejemplo en una sesión de vídeo es necesaria la coordinación ente las diferentes conexiones implicadas en ella. Este nivel de control debería pertenecer al plano de control pero dadas las limitaciones actuales de los protocolos de señalización (monoconexión por llamada, etc.) en la mayoría de los casos se debe utilizar el plano de usuario.

Control de aplicación, es el control extremo a extremo del servicio, engloba aquellas acciones que el usuario realiza a fin de controlar el servicio (navegación, control de la imagen, etc.). Este nivel de control está localizado en el plano de usuario.

Adicionalmente a estos tres niveles de control se establecerá en el plano de usuario la conexión para la transmisión de información entre desde el Centro de Información al usuario.

Dando por hecho que cualquier servicio de consulta requiere los mismos niveles de control y que el control de sesión va a necesitar al menos de una conexión en el plano de usuario, se tienen tres conexiones establecidas en el plan de usuario por medio del plano de control utilizando la señalización.

De estas conexiones, dos son bidireccionales para control (una entre el usuario y el Acceso al Servicio y otra entre el usuario y el Centro de Información) empleando sobre la capa MTA algún protocolo usuario a usuario ya estandarizado (por ejemplo el DSM-CC). Una tercera conexión es unidireccional para la transmisión de datos desde el Centro de Información al usuario.

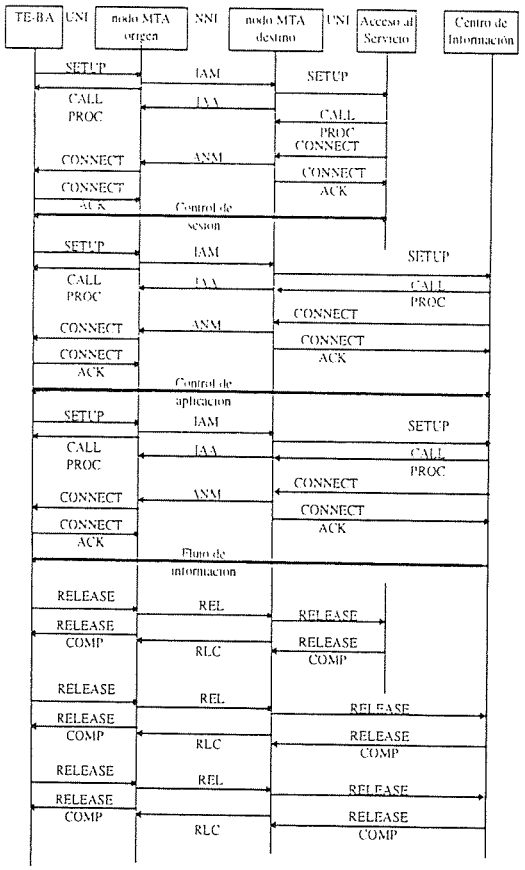


Figura 8: Principales flujos de comunicación asociados a un servicio de consulta

5 Estimación del tráfico de señalización para servicios de BA

Una vez seleccionados y clasificados los servicios de banda ancha (en cuanto a sus requisitos de señalización se refiere) la metodología propuesta para estimar el tráfico de señalización de los mismos es la siguiente:

- Identificación y caracterización de los mensajes de señalización (en las interfaces NNI)
- Estimación del tráfico de señalización por llamada para cada servicio en cada interfaz NNI
- Estimación del tráfico de señalización por cada usuario para cada servicio e interfaz NNI
- Estimación del tráfico total de señalización

En los siguientes apartados se detalla cada uno de los pasos de caracterización de los mensajes de señalización enunciados y se toma como ejemplo numérico el caso del servicio conversacional punto a punto de videotelefonía.

5.1 Caracterización de los mensajes de señalización

Los mensajes de señalización usados en la interfaz NNI para el tipo de servicios descritos que intervienen en el establecimiento y liberación son:

- Mensaje inicial de dirección (IAM)
- Mensaje de acuse de IAM (IAA)
- Mensaje de rechazo de IAM (IAR)
- Mensaje de respuesta (ANM)
- Mensaje de liberación (REL)
- Mensaje de liberación completa (RLC)

Para cada uno de estos mensajes se va a determinar su longitud teniendo en cuenta los parámetros de cada uno de ellos. El resultado de este proceso es la longitud de los mensajes a nivel PTM-3. Posteriormente se calcula la longitud total a nivel PTM-2 (señalización por red N° 7) y a nivel del número de células MTA (señalización por red MTA).

Independientemente del número de parámetros cada mensaje posee una longitud fija debida a las cabeceras comunes a todos los mensajes. Esta longitud es de 8 octetos.

La longitud de cada mensaje se dará en octetos.

Mensaje inicial de dirección: IAM

- Velocidad de célula MTA: 21 octetos
- Capacidad portadora BA: 11 octetos
- N° de la parte llamada: 15 octetos
- N° de la parte llamante: 15 octetos
- Categoría de la parte llamante: 6 octetos
- Contador del tiempo de propagación: 7 octetos
- Parámetros AAL: 22 octetos
- Identificador del elemento de conexión: 9 octetos
- Retardo de extremo a extremo: 7 octetos

Total longitud de parámetros: 112 octetos

Total longitud de mensaje: 120 octetos

Mensaje de acuse de recibo del IAM: IAA

- Identificador de elemento de conexión: 9 octetos

Total longitud de parámetros: 9 octetos
Total longitud de mensaje: 17 octetos

Mensaje de respuesta: ANM

- Indicador de tarificación: 6 octetos

Total longitud de parámetros: 6 octetos
Total longitud de mensaje: 14 octetos

Mensaje de liberación: REL

- Indicadores de causa: 34 octetos

Total longitud de parámetros: 34 octetos
Total longitud de mensaje: 42 octetos

Mensaje de liberación completa: RLC

Total longitud de parámetros: 0 octetos
Total longitud de mensaje: 8 octetos

Longitud de los mensajes de señalización a nivel PTM-2 (el nivel PTM-2 añade 6 octetos de cabecera)

IAM: 126 octetos

IAA: 23 octetos

ANM: 20 octetos

REL: 48 octetos

RLC: 14 octetos

Longitud de los mensajes de señalización a nivel ATM

IAM: 3 células

IAA: 1 células

ANM: 1 células

REL: 2 células

RLC: 1 células

5.2 Tráfico de señalización por llamada.

Una vez estimada la longitud de cada mensaje el siguiente paso es establecer el tráfico de señalización por llamada de cada servicio considerado. Para ello se ha de partir de los flujos de señalización definidos para cada servicio. A continuación se estima el tráfico de señalización por llamada para servicios conversacionales punto a punto. Para el resto de servicios el procedimiento es similar.

5.2.1 Servicios conversacionales punto a punto

Se obtendrá el tráfico de señalización por llamada partir del flujo de comunicaciones de la figura 4.

Se supone que la liberación de la llamada en este tipo de servicios la realiza un 50% de veces cada uno de los extremos. El sentido de ida se considera el correspondiente al del inicio de la llamada (sentido del IAM), el contrario se considera el sentido de vuelta.

En la Tabla 1 se muestran los resultados para el sentido de ida.

Tabla 1: Tráfico de ida para servicios conversacionales punto a punto

Mensaje	Longitud SS7	Longitud MTA
IAM	126	159 (3)
REL/2	24	53 (1)
RLC/2	7	53 (1)
Total	157	265 (5)

En la Tabla 2 se muestran los resultados en el sentido de vuelta

Tabla 2: Tráfico de vuelta para servicios conversacionales punto a punto

Mensaje	Longitud SS7	Longitud MTA
IAA	23	53 (1)
ANM	20	53 (1)
REl/2	24	53 (1)
RLC/2	7	53 (1)
Total	74	212 (4)

5.3 Tráfico de señalización generado por cada usuario

Una vez estimado el tráfico de señalización que genera una llamada el siguiente paso es estimar el tráfico de señalización generado por cada usuario de un servicio. Para ello se debería llevar a cabo una labor de identificación de tipos de usuarios atendiendo a los usos y demandas de los servicios considerados.

En el ejemplo que se viene siguiendo y con los servicios considerados se han identificado dos tipos de usuarios:

Usuarios residenciales

Usuarios de negocios

El tráfico de señalización por usuario se estima a partir del uso que de los diferentes servicios considerados hace cada tipo de abonado. Este uso se mide en llamadas por segundo en la hora cargada. Se presentan a continuación la estimación del uso de diferentes servicios por cada tipo de usuario

Usuarios residenciales:

Videotelefonía $1.389\ 10^{-4}$

Videoconferencia $8.33\ 10^{-5}$

Video bajo demanda $5.55\ 10^{-6}$

Emulación de circuitos 0

Red privada virtual 0

Usuarios de negocios:

Videotelefonía $5\ 10^{-4}$

Videoconferencia $1.389\ 10^{-4}$

Video bajo demanda 0

Los datos anteriores junto con el tráfico de señalización por llamada dan como resultado el tráfico de señalización por usuario y por cada servicio S A^I_S en cada interfaz I. Para el caso del servicio de videotelefonía se obtienen los siguientes resultados de tráfico de señalización por usuario tanto en el caso de señalización por la RSCC#7 como en el caso de señalización por la propia red ATM en la interfaz NNI entre nodos de tránsito.

Los resultados se dan en Octetos por segundo y por usuario.

Tráfico por usuario (videotelefonía). RSCC#7

Tráfico de ida. Residencial: 0.0218
Tráfico de vuelta. Residencial0.0102
Tráfico de ida. Negocios: 0.0785
Tráfico de vuelta. Negicios0.037

Tráfico por usuario (videotelefonía). MTA
Tráfico de ida. Residencial: 0.0368
Tráfico de vuelta. Residencial: 0.0294

Tráfico de ida. Negocios: 0.1325
Tráfico de vuelta. Negocios: 0.106

5.4 Tráfico total de señalización

En este capítulo se hace una estimación del tráfico de señalización que debe soportar cada interfaz considerada en la estructura genérica de red MTA.

Para ello los datos necesarios son:

- Tráfico de señalización por abonado y servicio.
- Número total de usuarios: N (millones)
- Proporción de tipos de usuarios. En el ejemplo que se viene siguiendo se ha considerado la siguiente proporción: 60% residenciales y 40% de negocios.
- Penetración de los diferentes servicios según el tipo de usuario
- Distribución del número de usuarios (N_E) según la estructura de red considerada (elementos de red centralizados o distribuidos, número de nodos de tránsito) para cada servicio

5.5 Penetración de servicios

Cada servicio considerado tendrá una diferente penetración en cada tipo de usuario identificado. Al mismo tiempo se han de establecer plazos temporales en los que tal penetración sea válida, ya que la penetración de un servicio no es constante con el tiempo.

En el ejemplo y con los servicios considerados el dato de penetración se va a considerar en tres periodos de tiempo:

- A corto plazo
- A medio plazo
- A largo plazo

Los datos de penetración mostrados en las siguientes tablas se han obtenido de [3]

A continuación se muestran las cifras de penetración estimadas a corto plazo para los servicios demandados por abonados del tipo residencial.

Videotelefonía: 7%

Videoconferencia: 4%

Video bajo demanda: 4%

A continuación se muestran las cifras estimadas de penetración a corto plazo para los servicios demandados por abonados del tipo negocios

Videotelefonía: 6%

Videoconferencia: 4%

Video bajo demanda: 0%

El tráfico total de señalización generado por los servicios por un tipo de usuario en una interfaz NNI y en cada sentido vendrá dado por la siguiente expresión:

$$A^I_R = \sum_S P_{RS} \cdot A^I_{RS} \cdot N_E \cdot 0.6$$

$$A^I_N = \sum_S P_{NS} \cdot A^I_{NS} \cdot N_E \cdot 0.4$$

Donde:

A^I_R: Tráfico total de señalización de un servicio para usuarios residenciales para la interfaz I

A^I_N: Tráfico total de señalización de un servicio para usuarios de negocios para la interfaz I

P_{RS}: Penetración (en tanto por uno) del servicio S para usuarios residenciales.

P_{NS}: Penetración (en tanto por uno) del servicio S para usuarios de negocios

A^I_{RS}: Tráfico de señalización por usuario residencial para el servicio S en la interfaz I

A^I_{NS}: Tráfico de señalización por usuario de negocios para el servicio S en la interfaz I

N_E: Número de usuarios que dependen del elemento E (nodo de tránsito, servidor, etc) según la interfaz considerada. Por ejemplo si se está considerando la estimación del tráfico para un servicio en la interfaz entre nodos de tránsito y la estructura de red contempla sólo dos de estos elementos N_E será N/2. Si lo que se considera es la interfaz entre un nodo de tránsito y un único servidor (por ejemplo un servidor de video bajo demanda) N_E será N.

Por lo tanto es en el factor N_E donde se considera la estructura de red (centralizada, distribuida) a la hora de hallar el tráfico total de señalización para un servicio y en una interfaz.

Las fórmulas anteriores suponen el caso peor de horas cargadas coincidentes para los servicios considerados.

5.6 Evaluación de la capacidad de la RSCC#7

Una vez estimado el tráfico total de señalización generado por los diferentes servicios en una determinada interfaz se habrá de evaluar si este tráfico puede ser o no cursado por la actual RSCC#7.

La capacidad de señalización entre dos nodos de la RSCC#7, C, viene determinada por los siguientes factores:

- Velocidad del enlace de señalización: 8 Kbytes/seg
- Carga máxima de dimensionado de un enlace de señalización. Este parámetro está comprendido en el rango 0.2-0.4.
- Número máximo de enlaces de señalización entre dos nodos: 16 enlaces.

Con las anteriores premisas la capacidad máxima queda en

$C = 51.200$ bytes/seg.

Así pues la evaluación se reduce a comprobar la siguiente desigualdad:

$$\max (A^I_R, A^I_N) < C$$

6 Conclusiones

Según se ha mostrado en este artículo la planificación de la red de señalización de una RDSI-BA basada en tecnología MTA requiere de muy diferentes consideraciones (servicios que se han de señalizar, estructura e interfaces, protocolos soportados en cada interfaz, tráfico de señalización que se ha de transferir entre elementos de red).

Se ha intentado elaborar una estrategia de tal forma que se puedan evaluar diferentes soluciones mostrando los pasos a seguir:

Selección de servicios

Agrupación de los mismos atendiendo a requisitos comunes de señalización

Descripción de las necesidades de señalización de cada grupo de servicios (componentes funcionales y flujos de señalización)

Identificación y caracterización de los mensajes de señalización

Estimación del tráfico de señalización generado en cada interfaz considerada a partir de datos de demanda y penetración teniendo en cuenta las diferentes estructuras de red (centralización de elementos, descentralización, distribución de usuarios, etc)

El tráfico de señalización así obtenido junto con el estado de definición de los protocolos de señalización en cada interfaz permitirá evaluar, con cierto criterio, las diferentes opciones barajadas a la hora de señalizar servicios de Banda Ancha.

Referencias

[1] Network Architecture. EURESCOM project EU-P515. May 1996.

[2] Proyecto INSIGNIA. (IN and B-ISDN signalling integration on ATM platform.

[3] RACE. Network Evolution, CFS M100. Dec 1994

[4] ATM Forum. ATM User-Network Interface Specification , Version 3.1.

[5] Recomendación UIT-T F.811

[6] Recomendación UIT-T F.812

[7] Recomendación UIT-T I.210

[8] Recomendación UIT-T I.211

[9] Recomendación UIT-T I.375

[10] Recomendación UIT-T I.580

[11] Recomendación UIT-T Q.2010

[12] Recomendación UIT-T Q.2210

[13] Recomendación UIT-T Q.2650

[14] Recomendación UIT-T Q.2721

[15] Recomendación UIT-T Q.2722

[16] Recomendación UIT-T Q.2764

[17] Recomendación UIT-T Q.2768

[18] Recomendación UIT-T Q.2971

Multicast

Transmisión en tiempo real y multicast sobre redes ATM

Enrique Arcizaga, Alicia San Millán
ROBOTIKER, LÍNEA DE REDES DE DATOS
PARQUE TECNOLÓGICO DE ZAMUDIO. EDIF.202 48170 ZAMUDIO
Correo electrónico : enrique@robotiker.es

Abstract :
The goal of this document is to analyse the different alternatives to transmit multimedia traffic to multiple receivers over broadband networks. IP multicast is the core technology for mutlimedia services over the Internet. The multimedia traffic needs a lot of bandwidth and the guarantee of the QoS. Within this document, we propose to integrate ATM networks, which provide high bandwidth and QoS, with IP multicast to get "Multicast Multimedia Services".

1. Transmisión en tiempo real y multicast sobre una red ATM

La mayoría de las aplicaciones multimedia desarrolladas hoy en día se centran en el envío de audio y video (y en muchos casos de datos) en tiempo real y con múltiples destinatarios a través de la red de transmisión.

En este documento se analiza y propone el uso de una infraestructura de red ATM [1] y el uso de IP multicast para cubrir las necesidades de los servicios multimedia.

2. Transmisión en tiempo real

Previo a la introducción sobre las características de ATM que hacen de ella una opción óptima para la transmisión de datos en tiempo real, es interesante hacer una breve referencia a las necesidades en cuanto a ancho de banda de los servicios que la red deberá soportar y cómo ATM es capaz de hacer frente a las mismas.

Entre los servicios más típicos se encuentra la transmisión de televisión de calidad estándar. Esta requiere un ancho de banda que varía entre 1,5 y 15 Mbps. La televisión de alta definición (HDTV, High Definition Television) que tiene una calidad superior a la anterior, requiere un ancho de banda que varía entre los 15 y los 150 Mbps. Un tercer ejemplo podría ser la videotelefonía, que requiere un ancho de banda entre 0,2 y 2 Mbps.

La Tabla 1 muestra los índices de error y el retardo global aceptados por diferentes servicios según el consorcio 1022 (Tecnología para ATM) del RACE (Research on Advanced Communications in Europe) [1].

2.1 Ventajas de ATM para los servicios en tiempo real

Con ATM no hay problemas de ancho de banda. El usuario contrata con la red el ancho de banda que necesita según la aplicación de la que se trate, teniendo como único límite los 622 Mbps que el ATM Forum define como máxima velocidad a contratar.

El usuario puede reservar el ancho de banda bien bajo demanda (SVC, Switched Virtual Circuit) o bien por subscripción (PVC, Permanent Virtual Circuit). Si la conexión se establece bajo demanda es necesario el intercambio de mensajes de señalización entre el usuario y la entidad de señalización.

La señalización se lleva a cabo a través del UNI (User - Network Interface). La versión de este protocolo implementada en los equipos actuales es el UNI 3.1 aunque el ATM Forum ha estandarizado ya el UNI 4.0 para señalización [2], [3].

A través de la señalización se negocia entre otras cosas, el ancho de banda requerido por la aplicación y la calidad de servicio asociado a la misma, asegurando los requisitos de los servicios multimedia.

Pero centrando el tema en la transmisión bidireccional o multidireccional de video y audio en tiempo real a través de la red, el punto clave reside en el control que ésta ejerce sobre el retardo global de transmisión (delay) y sobre la variación de retardo o CDV (Cell Delay Variation).

Las características de ATM que controlan y minimizan tanto el retardo global como el CDV son las siguientes :

a) El usuario exige a la red una calidad de servcio o QoS (Quality of Service) que ésta ha de mantener en todo momento.

b) La red ATM está formada por conmutadores que realizan la conmutación de celdas por hardware basándose en los identificadores VPI o VPI / VCI (Virtual Path Identifier, Virtual Channel Identifier). Esta simplicidad supone una mayor rapidez en el procesamiento de las

celdas y por tanto, un ahorro en el retardo global.

c) Las celdas son de tamaño fijo, 53 bytes. Esto supone que el tiempo de retardo introducido por los conmutadores de la red es constante (CDT del orden de unos 15μseg. por cada nodo de conmutación [4]), con una pequeña fracción variable debido al encolamiento de las celdas en los buffers de los mismos (CDV del orden de 6μseg. por nodo [4]). Todo esto hace que el retardo introducido por la red en la transmisión de las celdas esté entre límites controlados y por tanto, la red sea capaz de mantener los requisitos de la conexión en cuanto a retardos y CDV.

d) Los conmutadores de red disponen de buffers de entrada y/o salida para almacenar las celdas que esperan ser conmutadas. Como se ha dicho en el párrafo anterior, esto supone una variación variable en el retardo que sufren las celdas de una determinada conexión. Este retardo está controlado ya que los conmutadores son capaces de reaccionar frente a congestiones descartando celdas de baja prioridad para dar salida rápida a las de alta prioridad (esta prioridad viene indicada en la cabecera de la celda, en el campo CLP, Cell Loss Priority).

Estos últimos puntos concluyen que ATM es perfectamente válida para la transmisión de datos en tiempo real puesto que controla los retardos introducidos por la red, de forma que se satisfacen los requisitos de los servicios multimedia tal y como aparecen en la Tabla 1.

2.2 Aspectos de red relativos a los servicios en tiempo real

En la Fig.1 se muestran las causas, los efectos y la propagación a través de las capas [5].

Pero además del retardo global y del CDV existen otros aspectos que afectan de forma importante a los servicios en general y a los servicios en tiempo real en particular : congestión y llegada de celdas erróneas.

2.2.1 Celdas erróneas

La red ATM desecha todas aquellas celdas con cabecera errónea. Normalmente el error está en los valores VPI/VCI. Estos valores pueden ser mal asignados por un conmutador por ejemplo, y eso produce la pérdida de la celda para una conexión y la aparición de una celda "misinserted" (mal insertada) para otra. Estas celdas son desechadas a nivel físico (comprobación del HEC, Header Error Checksum) y no llegan a la capa ATM.

Por otra parte, también es posible la aparición de errores en los datos de usuario. La capa AAL (ATM Adaptation Layer) comprueba la validez de los mismos mediante el campo CRC (Cyclic Redundancy Check) pero no implementa ningún mecanismo de retransmisión, esta función deberá ser implementada por capas superiores, si es el caso.

2.2.2 Congestión en la red

El estado de congestión se define como aquél en que la red no es capaz de establecer nuevas conexiones ni de satisfacer el QoS contratado por las ya establecidas.

En una red ATM es posible llegar a un estado de congestión por fallos en algún punto de la red o por fluctuaciones inesperadas en el flujo de datos de alguna conexión establecida.

Existen varios algoritmos que los conmutadores ATM pueden ejecutar para solventar esta situación.

Servicio	BER	CLR	CMR	Retardo global (mseg.)
Telefonía	10E-7	10E-3	10E-3	25 / 500
Transmisión de datos	10E-7	10E-6	10E-6	1000 (50)
Video broadcast	10E-6	10E-8	10E-8	1000
Sonido HIFI	10E-5	10E-7	10E-7	1000
Control remoto de procesos	10E-5	10E-3	10E-3	1000

Nota : En el caso de la transmisión de datos aparecen en la tabla dos valores para el retardo global. Los 50 mseg. de retardo han sido introducidos para tener en cuenta sistemas distribuidos que exigen un retardo global menor.

BER : Ratio de error de bit (Bit Error Rate)
CLR : Ratio de pérdida de celda (Cell Loss Rate)
CIR : Ratio de inserción errónea de celda (Cell Misinsertion Rate)

Tabla 1. Indices de error y retardo global aceptados por diferentes servicios

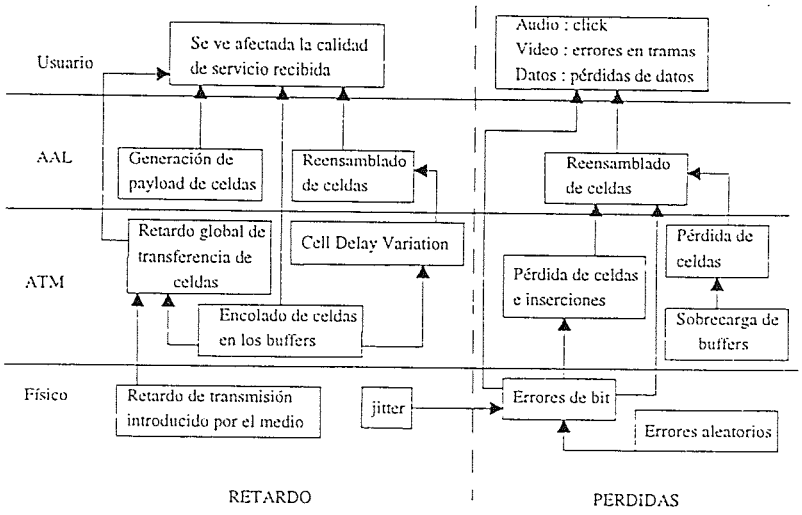


Fig. 1. Causas y efectos de los retardos y pérdidas

Los más usuales son EFCI (Explicit Forward Congestion Indicator), EPD (Early Packet Discard) y PPD (Partial Packet Discard) [4].

EFCI se basa en el bit CLP de la cabecera ATM. Toda celda con CLP = 1 es de baja prioridad. Un conmutador que entre en estado de congestión comenzará por descartar estas celdas para salir de tal estado.

EFCI es un algoritmo muy sencillo que desecha celdas de una forma aleatoria, sin preocuparse de si éstas pertenecen al mismo o a diferentes paquetes originales de datos. Este comportamiento es normal puesto que la red ATM entiende de celdas y no de paquetes pero el rendimiento obtenido por la red utilizando EFCI es muy bajo, dado que el paquete original no puede ser recuperado en recepción a partir de las celdas generadas en el proceso de segmentación.

Esto es lo que tratan de solventar tanto EPD como PPD. Se trata de descartar celdas que pertenezcan a un mismo paquete original de datos.

Con **EPD** el conmutador descarta celdas cuando detecta que se va a llegar a un estado de congestión. Los buffers del conmutador tienen definido un nivel umbral y el descarte de celdas se produce según el nivel de ocupación del buffer con respecto a este nivel umbral.

Si al comienzo de un nuevo paquete el buffer está por encima del nivel umbral, se procede al descarte de todas las celdas que pertenezcan a ese paquete.

Con **PPD** el conmutador actúa cuando la congestión ya está presente. Si una celda de un paquete es descartada, el resto de las celdas pertenecientes a ese paquete son también descartadas.

Con este algoritmo, las celdas anteriores a aquella primera que fue descartada, sí son enviadas a destino porque el algoritmo no predice la congestión, sino que actúa en el momento en que ésta se presenta.

Con EPD el rendimiento obtenido por la red es mejor que con PPD [6].

CLR	CDT (mseg.)	CDV (microseg.)	CER	CMR
10E-9	metropolitano : 2 nacional : 12 internacional : 24	metropolitano : 50 nacional : 150 internacional : 900	10E-8	10E-5

CLR : Indice de pérdida de celdas (Cell Loss Ratio)
CDT : Retardo de transferencia de celda (Cell Transfer Delay)
CDV : Variación de retardo de celda (Cell Delay Variation)
CER : Indice de celdas erróneas (Cell Error Ratio)
CMR :Indice de celdas mal insertadas (Cell Misinsertion Rate)

Tabla 2. Rendimiento GIGACOM

3. Ejemplo real de una red comercial ATM (GIGACOM)

GIGACOM es el servicio portador ATM de telefónica. Su cobertura comprende la totalidad del territorio nacional [7].

Actualmente la red GIGACOM sólo ofrece conexiones PVC. Aunque el UNI 3.1 está implementado en todos los equipos disponibles en el mercado, la red no lo soporta.

Por otra parte, transporta sólo conexiones CBR. Las conexiones VBR son atendidas como conexiones CBR con un ancho de banda igual al máximo contratado, PCR.

La conexión a GIGACOM debe ser a través de interfaces a 155 Mbps o a 34 Mbps.

La Tabla 2 refleja los valores de los parámetros de rendimiento (ver 4) obtenidos para la red GIGACOM (el CDT está medido sobre una distancia máxima de 4000 km).

El usuario que accede a GIGACOM ha de pagar una cuota mensual en concepto de abono (proporcional al número de accesos contratados) más una cantidad variable por cada conexión que establezca. Esta parte variable consta a su vez de una parte fija en concepto de establecimiento de conexión, más una parte variable dependiente de la distancia del usuario al nodo de acceso, es decir se tratará de llamada metropolitana o bien de llamada nacional, y de la duración de la conexión.

4. Servicios y clases de QoS en ATM

Como se ha comentado anteriormente, entre las ventajas que ATM ofrece está la posibilidad de exigir una ancho de banda y una calidad de servicio determinada a la red. Esto se realiza a través del descriptor de tráfico y de las clases de QoS.

El usuario especifica el ancho de banda requerido a través del descriptor de tráfico ATM incluido en el mensaje de establecimiento de conexión. Este descriptor contiene los siguientes parámetros :

a) PCR : Peak Cell Rate. Máxima velocidad a la que se transmitirá en la conexión.

b) SCR : Sustainable Cell Rate. Valor medio de velocidad sostenible por la conexión.

c) MBS : Maximum Burst Size. Máximo intervalo durante el que se transmitirá a la velocidad PCR.

d) MCR : Minimum Cell Rate. Mínima velocidad aceptada por la conexión.

No se especifican siempre todos los parámetros, depende del tipo de servicio exigido a la red. El ATM Forum ha definido cinco clases de servicios a soportar por la red ATM. Son los siguientes :

a) CBR. El flujo de datos está caracterizado por una velocidad de transmisión constante, PCR, garantizada para todas las celdas pertenecientes a la conexión y durante todo el tiempo que ésta permanezca activa. La red garantiza los valores exigidos de velocidad, retardo, variación de retardo e índice de pérdida de celdas. Este tipo de servicio es usado para transportar tráfico en tiempo real, es decir, altamente sensible a los retardos.

b) rtVBR. El flujo de datos se genera a velocidad variable caracterizada por PCR, SCR y MBS. La red garantiza los valores exigidos de velocidad, retardo, variación de retardo e índice de pérdida de celdas. Este tipo de servicio es usado para transportar tráfico en tiempo real, es decir, altamente sensible a los retardos.

c) nrtVBR. El flujo de datos se genera a velocidad variable caracterizada por PCR, SCR y MBS. Al contrario que en los anteriores, no se trata de tráfico en tiempo real. La red no garantiza un valor del retardo global ni variaciones de retardo. Sí garantiza la velocidad y el índice de pérdida de celdas exigidos.

d) ABR (Available Bit Rate). Es un servicio tipo Best Effort. La red garantiza una velocidad de transmisión entre MCR y PCR. No se trata de tráfico en tiempo real por lo que la red no garantiza ni el retardo global ni variaciones de retardo. Garantiza un índice de pérdidas bajo.

e) UBR (Unspecified Bit Rate). Es un servicio tipo Best Effort. La red no ofrece ninguna garantía de velocidad, de retardo global, de variaciones de retardo o de pérdida de información.

Cada tipo de servicio requiere una calidad de servicio diferente. En el UNI 3.1, se definen clases de QoS. Cada clase es un conjunto de valores asociados a los parámetros de rendimiento que se definen para la red ATM.

Los parámetros de rendimiento que definen la clase de QoS son los siguientes :

a) CLR (Cell Loss Ratio). Indice de pérdida de celdas.

b) máxCTD (Cell Transfer Delay). Retardo de transmisión máximo para cada celda de la conexión.

c) MCTD (Mean Cell Transfer Delay). Media de los CTDs de un número determinado de celdas.

d) CDV (Cell Delay Variation). Variación del retardo.

e) CER (Cell Error Ratio). Indice de celdas erróneas recibidas.

f) SECBR (Severely Errored Cell Block Ratio). Número de bloques severamente dañados. Un bloque está formado por M celdas, donde M es normalmente el número de celdas de datos transmitidas entre dos celdas OAM (Organization and Maintenance) consecutivas. Un bloque se considera severamente dañado cuando el número de celdas erróneas es igual o mayor que N, valor aún por definir.

g) CMR (Cell Misinsertion Rate). Indice de celdas mal insertadas recibidas.

Las clases de QoS definidas en el UNI 3.1 y los servicios soportados por la red se relacionan como sigue :

a) Clase 1. Garantiza los parámetros de rendimiento necesarios para la transmisión de audio y video a velocidad constante (CBR).

b) Clase 2. Garantiza los parámetros de rendimiento necesarios para la transmisión de audio y video a velocidad variable (rtVBR).

c) Clase 3. Garantiza los parámetros de rendimiento necesarios para la transmisión de datos orientados a conexión (nrtVBR).

d) Clase 4. Garantiza los parámetros de rendimiento necesarios para la transmisión de datos no orientados a conexión (nrtVBR).

e) Clase 0. Garantiza los parámetros de rendimiento exigidos por los servicios ABR y UBR.

La red está sólo obligada a ofrecer QoS Clase 0. El soporte del resto de las clases depende del proveedor de red.

La clase de QoS elegida se especifica, al igual que el descriptor de tráfico, en el mensaje de establecimiento de la conexión.

En el UNI 4.0 aparece la posibilidad de fijar valores individuales a los parámetros de rendimiento : CLR, máxCTD y CDV. Se mantiene la posibilidad de indicar la clase de QoS deseada por compatibilidad con el UNI 3.1 (para el caso de atravesar redes que no soporten el UNI 4.0) y como indicación de valores de QoS alternativos a los indicados individualmente.

Por otra parte, el ATM Forum define cuatro clases de capa AAL diferentes : AAL1, AAL2 (aún no definida), AAL3/4 y AAL5.

Cada capa es óptima para el transporte de un determinado servicio pero eso no quiere decir que esa capa sea el única que puede transportarlo.

Por ejemplo, la capa AAL1 se ha definido pensando en el servicio CBR. Este servicio es normalmente usado para la transmisión de datos síncronos a velocidad constante. La cabecera introducida por la capa AAL1 lleva información de control que garantiza la sincronización entre los dos extremos de transmisión. Ahora bien, nada prohibe al usuario seleccionar un servicio CBR (con QoS clase 1) y encapsular los datos mediante AAL5. La sincronización la conseguirá mediante protocolos adicionales.

Dado que nuestro propósito es la transmisión de información altamente sensible a los retardos de comunicación y en menor medida a las pérdidas de paquetes, las alternativas lógicas son CBR y rtVBR.

El establecer una conexión CBR supondría pagar durante todo el tiempo de conexión un ancho de banda que en muchos momentos permanecería inactivo. La opción lógica es elegir un servicio rtVBR.

Al elegir servicio rtVBR, la solución más lógica es utilizar la capa AAL2 del protocolo pero como aún no está definida, se elige AAL5 (Fig. 5).

5. Transmisión multicast

De todo lo anterior se deduce que ATM satisface los requisitos en cuanto a ancho de banda, retardos e índices de pérdidas que exige la transmisión de grandes cantidades de datos en tiempo real.

Ahora bien, como se ha comentado al principio, una de las aplicaciones multimedia que mayor auge tiene entre los usuarios finales es la videoconferencia. En ella, video, audio y en algunos casos datos, son transmitidos en tiempo real a través de la red hacia generalmente varios receptores (un solo receptor sería en este escenario un caso particular).

ATM no soporta direcciones de grupo o multicast, es decir, en los estándares tan sólo se definen

direcciones ATM individuales.

Hoy por hoy, las aplicaciones multicast utilizan el protocolo IP multicast. La red MBone (Multicast Backbone) utiliza este protocolo y ya se realizan videoconferencias sobre ella con una buena calidad.

Tanto el ATM Forum como el IETF (International Engineering Task Force) han desarrollado y continúan desarrollando estándares relativos a la transmisión de paquetes IP (unicast y multicast) a través de una red ATM. Los estándares son los siguientes :

a) El ATM Forum ha estandarizado el LANE (LAN Emulation) para que aplicaciones diseñadas para otros protocolos (IP, IPX, AppleTalk ...) puedan correr sobre una red ATM de forma transparente. Utilizando LANE y mediante la creación de VLANs (Virtual Local Area Network) es posible llevar a cabo comunicaciones multicast sobre una red ATM. El ATM Forum ha comenzado a desarrollar un segundo estándar, MPOA (Multiprotocol Over ATM). Se espera que su primera versión, MPOA 1.0, esté finalizada en Julio de 1997. [8]

b) El IETF ha creado por su parte estándares paralelos al anterior. Por una parte, el "Classical IP over ATM" (RFC 1577) [9] para aplicaciones IP sobre ATM y el "Multiprotocol Encapsulation Over ATM Adaptation Layer 5" (RFC 1483) [10] para aplicaciones IP, IPX y otras sobre una red ATM.

De todos ellos, el LANE es el más utilizado actualmente. El problema de Classical IP es que solo soporta transmisión de IP unicast sobre ATM. El estándar MPOA no está aún finalizado.

5.1 Transmisión de IP multicast sobre ATM

La transmisión de paquetes IP multicast sobre la red ATM exige la existencia de un sistema intermedio que permita el paso de direcciones IP de grupo a conexiones virtuales ATM y que encapsule los datagramas IP en celdas ATM.

Existen diferentes alternativas : LANE, MPOA y RSVP.

5.1.1 LANE, LAN Emulation

LANE es la capa adaptadora IP multicast - ATM utilizada actualmente para la transmisión de datos.

LANE es transparente a la aplicación multicast, es decir, la aplicación desconoce que está corriendo sobre una red ATM y no sobre una red IP. La desventaja de este método está precisamente ahí, al desconocer la existencia de ATM, la aplicación no se beneficia de una de sus principales ventajas, el soporte de diferentes QoSs.

En las redes IP no se garantiza el envío fiable y en orden de los datos. Es más, ni siquiera se garantiza el envío de los datos. Este servicio se denomina Best Effort.

LANE puede emular una red IP tipo Ethernet o tipo Token Ring y para ello el servicio ofrecido es también tipo Best Effort (ABR o UBR).

LANE permite la difusión de paquetes broadcast a través de su red virtual asociada. Para transmitir estos paquetes a participantes que pertenezcan a otras redes virtuales, es necesario el paso por el router multicast.

Una red virtual o VLAN es una red IP lógica definida sobre una red IP real.

LANE es adecuada para el envío de datos IP a través de una red ATM pero no para el envío de audio y/o video IP sobre la misma pues el servicio ABR, típicamente utilizado por LANE, no proporciona la calidad de servicio que este tipo de tráfico requiere.

5.1.2 MPOA, Multiprotocol Over ATM

MPOA no está todavía estandarizado y por tanto, no está implementado en los equipos de red existentes. No obstante, algunos de ellos implementan versiones pre-estándares propias del fabricante.

El uso de MPOA requiere el uso los siguientes estándares : UNI (3.0, 3.1, 4.0), LANE 2.0 y NHRP (Next Hop Resolution Protocol).

Las ventajas que MPOA ofrece son las siguientes :

a) Con MPOA es posible transportar toda clase de protocolos (IP, IPX, DecNET, etc.) sobre ATM. Esto se realiza utilizando la encapsulación LLC/SNAP con AAL5 [10].

b) Con MPOA se establece un circuito virtual ATM, con un QoS determinado, entre los terminales finales. Este circuito se denomina "shortcut".

c) Los servidores encargados de la resolución de direcciones dialogan entre sí a través del protocolo NHRP.

Como se muestra en Fig. 2, en MPOA existen dos componentes lógicos : MPC (MPOA Client) y MPS (MPOA Server).

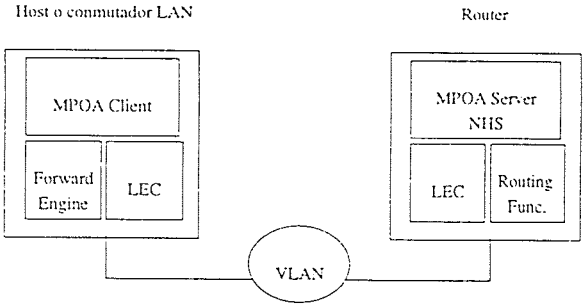


Fig. 2. Componentes MPOA

Cuando se intentan comunicar MPCs pertenecientes a VLANs diferentes, el utilizar LANE o un shortcut depende de la cantidad de tráfico transmitido. Así, se diferencia entre flujos "short-lived" y flujos "long-lived". Un flujo se denomina long-lived cuando el número de celdas transmitidas por la aplicación en un intervalo de tiempo determinado supera un umbral fijado a priori. En caso contrario, se trata de un flujo short-lived.

Si el flujo es short-lived se utiliza LANE para la comunicación entre ambos terminales. Es decir, los paquetes de datos seguirán el camino hallado por los routers (denominado "routed path"). Si el flujo es long-lived, el MPC origen resolverá la dirección ATM destino y establecerá una conexión ATM o shortcut.

En este último caso, la resolución de direcciones IP - ATM se realiza a través de MPOA Servers (MPS) y del protocolo NHRP, que a su vez está formado por dos componentes lógicos : NHC (Next Hop Client) y NHS (Next Hop Server).

Cada VLAN lleva asociado su MPS y su NHS. Todos los dispositivos pertenecientes a esa VLAN resolverán cualquier dirección a través de ellos.

Si el MPC fuente y la dirección IP destino a resolver pertenecen a la misma VLAN, el MPS de ésta será capaz de enviar directamente la dirección ATM pedida.

Si por el contrario, la dirección IP destino a resolver pertenece a una VLAN distinta a la del MPC fuente, la resolución de direcciones se llevará a cabo mediante el intercambio de mensajes NHRP entre los NHSs de las distintas VLAN.

El NHRP actúa sólo en un entorno unicast. La resolución de direcciones IP multicast a múltiples direcciones ATM pasa por el uso de servidores MARS (Multicast Address Resolution Server) que aún no han sido estandarizados por el ATM Forum.

5.1.3 RSVP, Resource Reservation Protocol

RSVP es el protocolo utilizado por la MBone para transmitir datos en tiempo real a través de una red IP multicast [11].

Este protocolo se implementa tanto en los terminales finales de comunicación como en los routers IP multicast.

Mediante paquetes RSVP denominados *"Resv"*, la aplicación receptora le indica al primer router el QoS necesario para la recepción de los paquetes IP de datos. En una red de transmisión IP, los paquetes *Resv* son encaminados por los routers IP hasta el terminal emisor, de forma que cada router implicado en el encaminamiento de los paquetes de datos realiza la reserva de recursos necesaria. El terminal emisor utiliza la información contenida en los paquetes *Resv* que le llegan para configurar sus funciones policía.

Una vez hecho esto, el terminal emisor emite paquetes *"Path"* hacia el terminal receptor que contienen información importante : dirección IP del router anterior (ya que los paquetes *Resv* han seguido el mismo camino que deberán seguir los paquetes de datos, pero en sentido contrario), formato de los paquetes de datos que se transmitirán, características del tráfico a transmitir, etc.

A partir de aquí, el terminal emisor podrá transmitir los paquetes de datos a través de la red IP con un QoS garantizado.

En RSVP se definen cuatro tipos de QoS para el transporte de tráfico IP :

a) Garantizado, que ofrece un retardo garantizado matemático.

b) Predictivo, que ofrece un límite de retardo probabilístico.

c) Controlado, que trata de ofrece varios niveles de QoS entre los cuales la aplicación escogerá el que considere más adecuado. Sólo garantiza un retardo mínimo.

d) Best Effort.

La reserva RSVP puede borrarse explícitamente a través de un paquete *"ResvTear"* o de un paquete *"PathTear"*. Esto no es en realidad necesario pues RSVP se basa en los paquetes de refresco para mantener o borrar las reservas realizadas.

Los paquetes *Resv* y los paquetes *Path* se transmiten a través de la red cada cierto tiempo. Esto de denomina "refresco" y el intervalo de tiempo transcurrido entre dos paquetes *Resv* (o *Path*) sucesivos se denomina "tiempo de refresco".

Por otra parte, los routers tienen determinado un tiempo de espera máximo o "timeout" para los paquetes de refresco. Si el timeout expira porque el refresco no ha llegado, la reserva de recursos se borra.

Ahora bien, en este documento se ha elegido una infraestructura de red ATM. En ella, el QoS de la conexión se negocia a través del protocolo de señalización UNI 3.1.

Es decir, el router multicast que recibe los paquetes *Resv* de la aplicación receptora debe ser capaz de establecer una conexión ATM con un QoS determinado basándose en la información contenida en esos paquetes.

En este punto hay que tener cuenta ciertas diferencias entre RSVP y ATM :

a) En RSVP es el receptor el que solicita un determinado QoS mientras que en ATM lo hace el emisor a través de la señalización UNI.

b) En RSVP es posible modificar el valor de QoS mediante el refresco. En ATM el valor de QoS se mantiene durante todo el tiempo de conexión.

c) En RSVP se borran las reservas de recursos a través del timeout. En ATM es necesario el envío del correspondiente paquete de señalización (*"Release"*).

d) En RSVP la reserva de recursos y la generación de la ruta a seguir por los paquetes de datos se realiza en fases diferentes. En ATM la solicitud del QoS es concurrente con el establecimiento de la ruta para las celdas de datos.

e) RSVP permite QoSs diferentes para cada receptor de un entorno multicast. ATM ofrece el mismo QoS a todos los receptores del grupo y en concreto, aquel solicitado por el primer participante del grupo.

f) RSVP realiza una reserva de recursos unidireccional mientras que ATM garantiza, en conexiones punto a punto, un QoS para ambos sentidos de la comunicación. En un entorno multicast, ATM también reserva recursos unidireccionalmente.

5.2 Solución elegida

De las tres opciones anteriores la más adecuada es sin duda la última, RSVP. MPOA no está definido aún y LANE no es adecuado por soportar sólo servicios ABR.

Con el mapeado de calidades de servicio RSVP-ATM, la aplicación IP puede pedirle a la red el QoS que le convenga según sus necesidades.

Pero los routers disponibles en el mercado no son capaces de realizar este mapeado. Tan sólo transportan IP sobre ATM utilizando LANE (o Classical IP) de forma que el servicio solicitado a la red es ABR que como se ha explicado anteriormente no es válido para el envío de datos en tiempo real.

Con RSVP la aplicación exige a la red los recursos necesarios para satisfacer sus requisitos de tiempo real. Ahora bien, la transmisión multicast engloba otra serie de problemas que habrá que resolver de una u otra forma. Son los siguientes :

a) Los usuarios deberán tener la capacidad de darse de alta o darse de baja en un grupo multicast.

b) La transmisión a través de la red de un mismo paquete de datos hacia múltiples receptores deberá sobrecargar mínimamente la misma.

c) En algún punto de la red se deberá resolver una dirección IP multicast a múltiples direcciones ATM individuales.

La resolución de estas cuestiones se realiza mediante el uso de tres protocolos : IGMP (Internet Group Multiast Protocol) , DVMRP (Distance Vector Multicast Routing Protocol) y ATMARP (ATM Address Resolution Protocol). A continuación se explican todos ellos :

IGMP es el protocolo utilizado para llevar cuenta de los participantes de cada grupo multicast establecido. Se implementa tanto en los routers multicast como en los terminales finales de la aplicación.

La primitiva *"Join_Local_Group"* es usada por el terminal final siempre que quiera unirse a un determinado grupo multicast. El router procesa esta primitiva y anota en su tabla la nueva dirección IP individual participante en el grupo multicast.

La primitiva *"Leave_Local_Group"* es usada por el terminal final siempre que quiera borrarse de un determinado grupo multicast. El router procesa esta primitiva y borra la línea correspondiente en su tabla interna.

De esta forma el router multicast conoce las direcciones IP de los participantes del grupo multicast en su red local. Algunos de ellos estarán conectados al router a través de un interfaz Ethernet, otros serán terminales finales ATM. En Fig. 4 se muestra una posible infraestructura de red.

Ahora bien, puede que el grupo multicast sea externo, es decir, que terminales pertenecientes a otras redes locales y asignados a otros routers multicast, formen parte también de ese grupo.

Los paquetes IP multicast han de ser transmitidos hacia todos los participantes del grupo, ya sean externos o locales. Un router multicast ha de saber a qué otro router o routers multicast ha de enviar los paquetes de datos para que éstos lleguen a todos los destinatarios.

Por otra parte, esta transmisión deberá hacerse de forma que la red se sobrecargue mínimamente, cuestión b) planteada anteriormente.

La forma de obtener esta información de encaminamiento es el protocolo DVMRP, derivado del RIP (Routing Information Protocol) y usado actualmente por la MBone.

DVMRP se ejecuta en cada router multicast. Se basa en el intercambio de datagramas DVMRP encapsulados en paquetes IP entre routers multicast vecinos. Dichos datagramas están compuestos por una cabecera de tamaño fijo (la misma que en IGMP) y una parte de datos que contiene bien una petición de ruta o bien una respuesta a una petición determinada.

Un router multicast determina la ruta adecuada (camino mínimo) para el envío de paquetes multicast a partir de la información recibida en los datagramas DVMRP de respuesta provenientes de sus routers vecinos.

El uso de DVMRP sobre ATM supone el establecimiento de conexiones virtuales entre routers multicast vecinos. Lo lógico es encapsular los datagramas IP con LLC/SNAP y utilizar el servicio ABR.

Una vez el router conoce las direcciones IP participantes del grupo (terminales finales locales y routers vecinos) ha de traducirlas a direcciones ATM, cuestión c) planteada anteriormente.

El router recibe un paquete IP multicast que ha de transmitir a varios receptores a través de la red. El paquete IP multicast lleva como dirección de destino la dirección IP multicast o de grupo. El router debe traducir de alguna forma esta dirección de grupo en las direcciones ATM individuales de todos y cada uno de los participantes en el grupo.

Una solución posible es crear un servidor que traduzca la dirección IP multicast en las direcciones ATM individuales. Una vez realizado esto, a través del UNI 3.1 se establecería el enlace punto - multipunto adecuado.

Este servidor de direcciones se denomina MARS (Multicast Address Resolution Server). Los primeros estudios realizados sobre este tema están reflejados en la RFC 2022 del IETF. Además, el MARS está en fase de desarrollo en los laboratorios Bellcore.

Tabla de encaminamiento

IP destino	Vía
144.254.37.X	Directo
144.254.65.X	Directo
144.254.23.X	144.254.10.3
144.254.51.X	144.254.10.3
144.254.49.X	144.254.10.2
144.254.73.1	144.254.10.3

Tabla de resolución de direcciones

IP destino	ATM destino
144.254.10.3	a
144.254.10.2	b

Fig. 3. Tablas internas del router

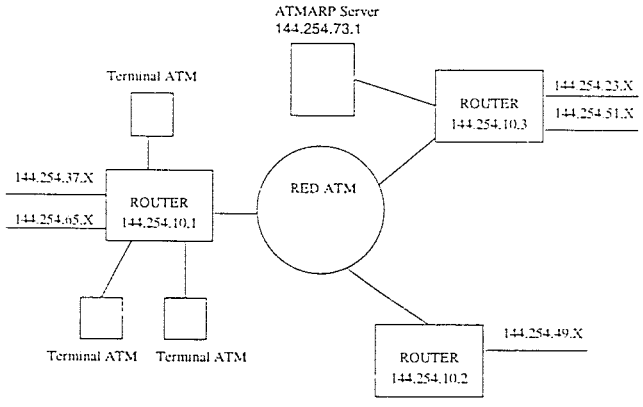


Fig.4. Red global ATM

Actualmente, la resolución de direcciones se realiza mediante el protocolo **ATMARP** que consiste en un servidor de direcciones IP unicast - direcciones ATM que sirve a un dominio determinado dentro la red global. Este dominio se suele referir como LIS (Logical Internet Subnet).

El servidor va aprendiendo las direcciones IP y direcciones ATM de todos los nodos de su LIS. Todo terminal se configura con la dirección IP / dirección ATM del servidor. Cuando el terminal se da de alta en la red, envía un mensaje al servidor indicándole su dirección IP. El servidor toma nota de ello, pidiéndole además su dirección ATM.

Cuando un terminal desea conocer la dirección ATM correspondiente a una determinada dirección IP, envía la consiguiente petición al servidor (*"ATMARP_Request"*) y éste le responde con la dirección pedida (*"ATMARP_Reply"*).

Este protocolo es muy simple pero tiene una limitación importante : la comunicación entre nodos pertenecientes a distintas LISs se da a través de routers por defecto.

Por ejemplo, en Fig.4 aparece dibujado el servidor de una LIS determinada. Supongamos que su router por defecto es aquel con dirección IP 144.254.10.3. También supongamos que la Fig. 3 muestra las tablas internas del router 144.254.10.1.

La tabla de encaminamiento contiene información sobre como llegar hasta otros terminales de la LIS. Si se quiere transferir información a un terminal con dirección 144.254.49.5, se hará a través del router 144.254.10.2 como indica la tabla. Además, la tabla de resolución de direcciones indica la dirección ATM de este último router. Sin embargo, si la dirección del terminal destino no aparece en la tabla de encaminamiento, los datos son enviados al router por defecto, 144.254.10.3, cuya dirección ATM aparece también en la tabla de resolución de direcciones.

Exactamente lo mismo ocurre cuando se desee resolver una dirección IP a una dirección ATM. Si la dirección IP a resolver pertenece a la LIS, es decir, aparece en la tabla de encaminamiento, una *ATMARP_Request* es enviada al servidor. Sin embargo, si esta dirección IP destino no aparece en la tabla, la petición es enviada al router por defecto. Este router a su vez, reenviará la petición al router por defecto de otra LIS.

5.3 Resumen de los protocolos multicast

En resumen, los terminales finales se dan de alta o de baja en un grupo multicast a través de primitivas IGMP.

Un router multicast sabe perfectamente como direccionar los paquetes de datos destinados a un determinado grupo multicast. En primer lugar y gracias al protocolo IGMP, puede difundir el paquete por los participantes del grupo pertenecientes a su red local. En segundo lugar y gracias al protocolo DVMRP, conoce las direcciones de los routers hacia los cuales ha de transmitir los paquetes para que éstos lleguen al resto de los participantes del grupo.

Una vez conocidas las direcciones IP de interés, la traducción a direcciones ATM se hace a través del protocolo ATMARP.

A través del UNI 3.1 se establecerá la conexión punto - multipunto adecuada.

6. Sincronización entre terminales finales

Como se ha comentado en el capítulo 4, el ATM Forum ha estandarizado una serie de tipos de servicio, de clases de QoS y de capas AAL.

En ese apartado se ha comentado también que para la transmisión de video y audio en tiempo real a través de la red ATM es lógico optar por un servicio rtVBR con QoS Clase 2. La capa AAL óptima sería la AAL2 pero al no estar aún definida se elige la AAL5.

La capa AAL5 es mucho más sencilla que la AAL1. Su generación y su procesado son mucho más rápidos. Ahora bien, esta sencillez radica en la escasez de información de control transportada por su cabecera.

La Fig. 5 muestra el formato de trama AAL5. Como paso previo a la encapsulación en AAL5, es conveniente insertar una cabecera LLC/SNAP para que el sistema diseñado sea compatible con la especificación "Multiprotocol Encapsulation Over ATM Adaptation Layer 5" (RFC 1483) del IETF, aceptada también por el ATM Forum para el estándar MPOA. Esta encapsulación sirve para distinguir la clase de protocolo que se transporta sobre ATM. Los valores de la cabecera LLC/SNAP que aparecen en la figura son los correspondientes al datagrama IP.

Como se ve en la figura, la cabecera insertada por AAL5 no contiene información alguna de sincronización. Esta deberá ser transportada por algún otro protocolo.

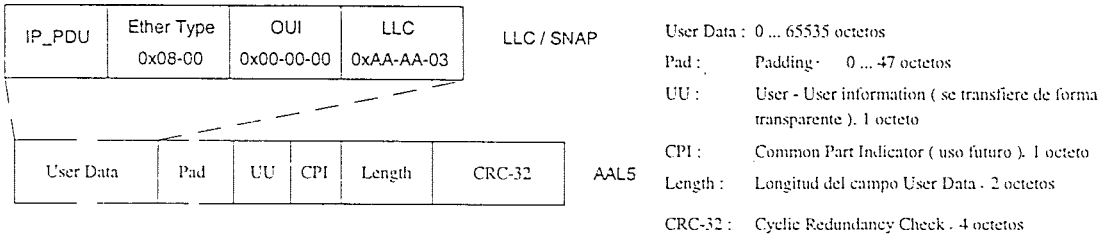


Fig 5. Encapsulación AAL5

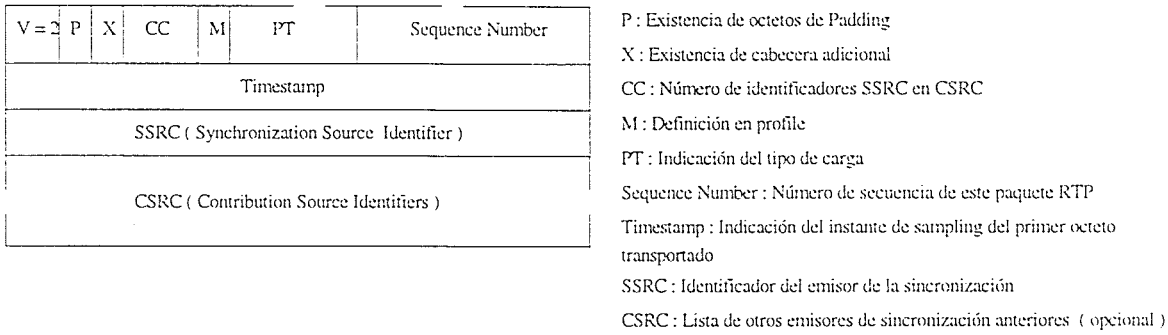


Fig. 6. Paquete RTP

RTP (Real Time Transport Protocol) es el protocolo usado en la MBone para este fin y está definido en la RFC 1889 del IETF.

RTP actúa junto con RTCP (Real Time Control Protocol) de forma que RTP transporta los datos de usuario y RTCP transporta los datos de control. Ambos protocolos corren sobre UDP utilizando puertos de comunicación distintos, RTP un puerto par y RTCP el puerto impar inmediatamente superior.

El video y el audio se transportan en sesiones separadas con lo que en una sesión de videoconferencia se utilizarían cuatro puertos lógicos diferentes.

La Fig.6 muestra la cabecera RTP que transporta la información de sincronización necesaria para la transmisión de datos síncronos. El Payload Type indica el tipo de carga transportada. El número de secuencia sirve para la detección de pérdidas y la recuperación en orden de paquetes. El timestamp sirve para conocer los instantes de tiempo del stream de video o audio original al que corresponden los datos llegados.

Cada participante de una sesión multicast genera paquetes de control RTCP que permiten realizar estadísticas sobre la calidad de emisión y/o recepción.

Uno de los campos de este paquete lleva la identificación de la fuente de datos, CNAME (Canonical Name), que identifica el terminal emisor tanto del audio como del video. Este valor no se ve modificado por sistemas intermedios ni por reseteos del terminal, al contrario que el SSRC de la cabecera RTP, que es un número aleatorio elegido por el terminal, diferente para los streams de audio y video y que puede verse modificado por sistemas intermedios o por el reseteo del terminal.

La transmisión de paquetes RTCP tiene asignado un tanto por ciento del ancho de banda total muy pequeño, aproximadamente del 5%.

Todos los participantes generan los paquetes RTCP con el mismo intervalo de tiempo de forma que si pasado un número pequeño de estos intervalos, no se ha recibido ningún paquete RTP o RTCP de un determinado participante, se considera que este ha dejado de formar parte del grupo y todos los demás o lo borran de su tabla o lo marcan como inactivo.

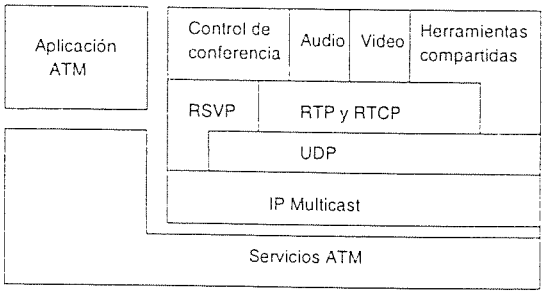


Fig. 7. Pila de protocolos

7. Comunicación extremo a extremo

La mayoría de los sistemas se basan en la arquitectura cliente - servidor. El sistema cliente desea acceder a cierta información contenida en el sistema servidor y para ello, es necesario que ambos se comuniquen a través de un mismo protocolo.

Una alternativa posible es el protocolo RTSP (Real Time Streaming Protocol), que posibilita el envío controlado de datos en tiempo real.

En RTSP se distinguen dos tipos de mensajes : mensajes de conexión y mensajes objeto. Los mensajes de conexión están directamente relacionados con el establecimiento y liberación de la conexión. Los mensajes objeto identifican el stream de datos a recuperar y parámetros relacionados con el mismo.

RTSP permite enviar los datos de la aplicación mediante diferentes protocolos de transporte : TCP, UDP, UDP multicast, etc. Normalmente es el cliente el que le indica al servidor mediante un mensaje *"SetTransport"* el tipo de protocolo de transporte a utilizar. Existe un sólo caso en el que el servidor puede enviar este mensaje : informar al cliente de que el stream de datos tiene capacidad multicast.

Conclusiones

En definitiva, para poder transmitir aplicaciones multimedia que incorporen las facilidades de multicast sobre una red ATM, hace falta recurrir al protocolo IP. Sin embargo, dadas las características de este protocolo (Best Effort), obliga a utilizar un conjunto de protocolos adicionales (RTP, RTCP, RSVP, etc.) para poder satisfacer los requisitos de tiempo real de los servicios multimedia.

Actualmente, se está trabajando a nivel de proyectos europeos (ACTS) para que el "gap" entre el mundo Internet y el mundo ATM desaparezca en los próximos años.

Referencias

[1] De Prycker, Martin : "Asynchronous Transfer Mode. Solution for Broadband ISDN". 1995

[2] ATM Forum : "ATM User - Network Interface Specification". Versión 3.1. Septiembre 1994.

[3] ATM Forum : "ATM User - Network Interface Specification". Versión 4.0. Julio 1996.

[4] Bradner, Scott, Mcclellan, Rolf : "The 1997 Strategic Networks ATM Switch Evaluation Results". Mayo 1997.

[5] Cuthbert, L.G, Sapanel, J.C : "ATM. The Broadband Telecommunications Solutions ". 1993

[6] Romanov, Allyn, Floyd, Sally : "Dynamics of TCP Traffic Over ATM Networks ".

[7] Alonso, A., Gómez, O. : "Gigacom. El servicio portador de Telefónica". Global Communications. Pág. 56-61. Abril 1997.

[8] Fredette, Ander N. : "Multiprotocol Over ATM Version 1.0. MPOA Baseline". Mayo 1997.

[9] Laubach, M. : "Classical IP Over ATM". RFC 1577. Hewlett Packard Laboratories. Enero 1994.

[10] Heinamen, Juha : "Multiprotocol Over AAL5". RFC 1483. Telecom Finland. Julio 1993.

[11] Braden, R. : "Resource Reservation Protocol". Internet Draft. Noviembre 1996.

Conexiones Multipunto en Arquitecturas de Conmutación MTA. Una Solución para Entornos de Área Local

Josepmaria Malgosa Sanahuja[1], Joan García Haro[2] y Alberto Gimeno Cardo[1]
[1]DEPARTAMENTO DE INGENIERÍA ELECTRÓNICA Y COMUNICACIONES
CENTRO POLITÉCNICO SUPERIOR, UNIVERSIDAD DE ZARAGOZA
MARÍA DE LUNA 3, 50015 ZARAGOZA
[2]DEPARTAMENTO DE MATEMÁTICA APLICADA Y TELEMÁTICA
UNIVERSIDAD POLITÉCNICA DE CATALUÑA
C/JORDI GIRONA, 1 Y 3. MÓDULO C-3, CAMPUS NORD, 08034 BARCELONA
Correo electrónico: jms@posta.unizar.es, teljgh@mat.upc.es

Abstract:

Most of the services that an ATM Broadband Integrated Services Digital Network (B-ISDN) has to provide have a multicast and broadcast nature. For this type of services, a source of traffic transmits information to multiple destinations. In consequence, the ATM switching nodes have to support multi-point connections and therefore require some mechanism to allow the replication of ATM multicast cells to subsequently route them. In this paper a novel ATM switching architecture providing multicast capabilities is presented. This ATM switch operates in a very simple manner and achieves a notable reduction in the hardware complexity. However, the system becomes unstable for some range of operation values. To increase the stability margin of the system, along with the incorporation of speed-up techniques, a new cell selection policy adaptive to the input traffic is proposed. This cell selection algorithm improves the switch performance under bursty traffic and keeps low the hardware cost.

Finally, the performance of the entire system is evaluated by using computer simulation. An approximate mathematical analysis that justifies and validates the obtained results is also developed in this paper.

1. INTRODUCCIÓN

En este artículo se ofrece primeramente un enfoque general de la problemática multipunto, describiendo las arquitecturas de conmutación MTA (Modo de Transferencia Asíncrono) más representativas que lo soportan. Se aporta una clasificación de las propuestas según dos grandes grupos: basadas en red de copia [1][2][3][4] y basadas en etapas de memoria compartida [6][7][8][9][10]. Se analizan algunos de los problemas que presentan ambas metodologías y se proponen posibles soluciones.

A continuación se propone una nueva arquitectura de conmutación MTA con capacidad multipunto de muy baja complejidad *hardware* [11]. En contrapartida, la arquitectura presenta tramos de inestabilidad y un crecimiento limitado de su estructura, lo cual no es un inconveniente para su aplicación en redes de área local MTA (ATM LAN). Para aumentar el margen de estabilidad del sistema, se sugiere la incorporación de técnicas de aceleración (*speed-up*). Las prestaciones de esta arquitectura básica se evalúan exhaustivamente bajo distintas caracterizaciones de tráfico multipunto. En particular se estudia el comportamiento bajo tráfico multipunto a ráfagas.

Este tráfico es de vital importancia pues se admite que la aplicación más inmediata de los conmutadores MTA en la industria se llevará a cabo mediante la emulación de redes de área local (LANE). En dichas redes, las tramas generadas por el subnivel de acceso al medio (MAC) deben segmentarse en grupos de 48 octetos, cada uno de los cuales está asociado con una celda MTA. Este conjunto de celdas determinan el tamaño de una ráfaga, mientras que el tiempo transcurrido entre la transmisión de dos tramas consecutivas establece la duración del período de silencio.

Debido a la inherente capacidad *broadcast* que presentan las redes LAN, es fundamental que el conmutador empleado soporte tráfico multipunto. El objetivo es mejorar de nuevo la arquitectura básica propuesta mediante el diseño de una nueva política de selección que consiga incrementar sus prestaciones manteniendo la simplicidad *hardware*. La eficacia de esta nueva política se fundamenta en el hecho de que todas las celdas que conforman una ráfaga tienen el mismo conjunto de direcciones destino.

Finalmente, se evalúan mediante simulación las prestaciones del sistema y se desarrolla un análisis matemático que, de forma aproximada, justifica y valida los resultados obtenidos.

2. SOLUCIONES MULTIPUNTO

Es este apartado se describen brevemente las aportaciones más significativas que se han realizado con el objeto de adaptar los conmutadores MTA al tráfico multipunto. Dichas aportaciones pueden dividirse en dos grandes grupos: basadas en red de copia y basadas en elementos de memoria compartida.

2.1. Arquitecturas multipunto basadas en red de copia

En este tipo de arquitecturas, el tratamiento del tráfico multipunto se consigue concatenando una

red de copia al nodo conmutador (Fig. 1). La misión de la red de copia es la de realizar físicamente todas y cada una de las réplicas de las celdas multipunto. El conmutador MTA conectado a su salida es el encargado de encaminar las réplicas hacia sus correspondientes puertos de destino. En conclusión, se puede admitir que una red de copia simplemente convierte el tráfico punto-multipunto en tráfico punto a punto (que es tratado posteriormente por el conmutador propiamente dicho).

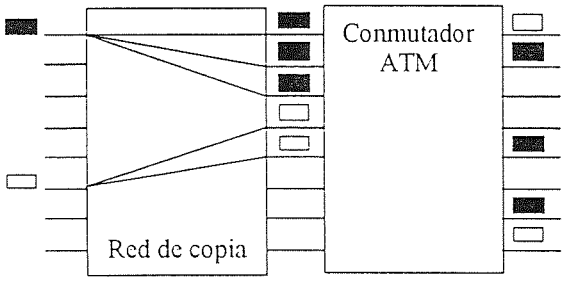


Fig. 1. Estructura general de un conmutador multipunto basado en red de copia.

Muchos de los sistemas de conmutación multipunto pueden ajustarse dentro de esta categoría de arquitecturas, de entre las que destacamos las expuestas en [1][2][3][4].

Las arquitecturas de conmutación multipunto basadas en red de copia presentan la ventaja de que no condicionan el diseño del conmutador MTA concatenado a su salida, puesto que ambos sistemas trabajan de forma independiente. En contrapartida, los inconvenientes más relevantes que presentan son:

- La posibilidad de saturar la red de copia si el número de réplicas demandado por las celdas multipunto sobrepasa el tamaño de la red. En consecuencia, las celdas que no puedan replicarse se perderán, o bien, serán demoradas y tratadas de nuevo.
- Debido al algoritmo que controla el proceso de réplica, las celdas multipunto presentes en la entrada deben servirse en un orden estrictamente monótono (creciente o decreciente). En situaciones de carga elevada, es posible que las primeras celdas en servirse consuman la mayor parte de la capacidad total de la red. En consecuencia, las subsiguientes celdas multipunto con un número de copias elevado no podrán servirse provocando el bloqueo a las celdas con un número de copias más reducido. Este fenómeno recibe el nombre de *HOL fanout blocking*.

En la Fig. 2 se ejemplifica de forma gráfica los problemas de la saturación y del *HOL fanout blocking*.

El primer problema se puede solventar incorporando colas en la entrada de la red de copia. De esta manera, si en una determinada ranura temporal se supera la capacidad total de la red, las celdas que no puedan servirse se almacenarán en cola a la espera de ser replicadas en las siguientes ranuras temporales, evitando las pérdidas que presentaba la estructura original.

El segundo problema se acostumbra a solucionar con una técnica denominada división de celdas (*cell splitting*). Esta técnica permite que el total de copias que requiere una celda puedan hacerse en ranuras temporales consecutivas y no forzosamente en una misma ranura temporal. En consecuencia, la capacidad total de la red es aprovechada al máximo aunque en ambos casos el retardo que sufren las celdas se incrementa.

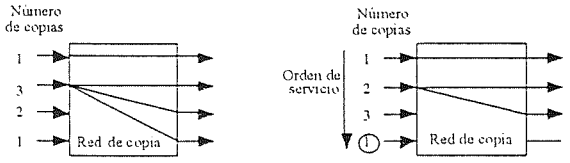


Fig. 2. Ejemplo de saturación y HOL fanout blocking en una red de copia.

2.2. Arquitecturas multipunto basadas en elementos de memoria compartida

En este tipo de arquitecturas [5][6][7][8][9][10], todos los puertos de entrada-salida comparten de forma dinámica un único módulo de memoria como se muestra en la Fig. 3. Aunque físicamente el *buffer* constituye una unidad de memoria *hardware*, su capacidad total se subdivide en N colas lógicas (listas encadenadas), cada una de ellas asociada a un puerto de salida. En la memoria *buffer* se almacenan las celdas MTA y en el módulo *memoria de direcciones* se almacenan los punteros que permiten localizar cada una de las celdas en su respectiva cola lógica. El módulo *controlador* es el encargado de gestionar y mantener dichas colas lógicas.

Durante la fase de escritura, las celdas presentes en los puertos de entrada son multiplexadas en un único enlace y almacenadas en la memoria compartida. Con el objeto de reducir la velocidad de proceso de las memorias, es habitual incorporar a la salida del multiplexor un convertidor serie-paralelo. La longitud del BUS de salida L, puede coincidir con la longitud en bits de una celda MTA.

De lo expuesto hasta el momento, se desprende que en esta arquitectura, el proceso de conmutación se lleva a cabo gracias al *controlador*, almacenando cada celda entrante en la lista encadenada que le corresponde (en función de su dirección destino).

En la fase de lectura, se extraen celdas de la memoria para trasladarlas hacia sus correspondientes puertos de salida. Para ello, basta con leer secuencialmente las celdas ubicadas en la cabecera de

cada una de las colas lógicas. Finalmente, el demultiplexor síncrono es el encargado de distribuirlas hacia sus correspondientes puertos de destino.

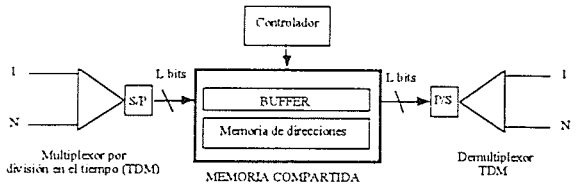


Fig. 3. Estructura de un conmutador multipunto basado en unidades de memoria compartida.

De entre las numerosas ventajas que presentan las arquitecturas basadas en etapas de memoria compartida destacamos su facilidad de adaptación al tráfico multipunto. En este entorno, la única discrepancia con respecto a la metodología anterior es que ahora, la incorporación de una celda multipunto en el *buffer* conlleva actualizar todas las colas lógicas a las que van dirigidas cada una de sus copias. A diferencia de lo que ocurre en las redes de copia, aquí el proceso de réplica no es físico sino lógico (es decir, en vez de replicar celdas se replican direcciones de memoria, con el consiguiente ahorro de memoria física).

Las prestaciones de este tipo de arquitecturas son idénticas a las conseguidas en los conmutadores con colas en la salida en cuanto a los retardos de conmutación. Sin embargo, el uso de la capacidad de almacenamiento del sistema es más eficiente que el obtenido en arquitecturas de conmutación con colas en la entrada y en la salida [12]. Además, se trata de arquitecturas más fáciles de adaptar al tráfico multipunto.

El inconveniente más notable que presenta este tipo de sistemas es que necesitan utilizar memorias muy rápidas. La capacidad de las memorias depende del número máximo de copias que puede llegar a demandar una celda multipunto y en consecuencia, del tamaño del conmutador N, limitando su capacidad de crecimiento como conmutadores de una sola etapa. En [13][14][15][16] se propone el diseño de conmutadores de gran tamaño basados en etapas de memoria compartida a partir de elementos de tamaño reducido, interconectándolos a través de una red de tres etapas (Clos).

3. ARQUITECTURA DE CONMUTACIÓN BASADA EN MÓDULOS GESTIONADOS POR PALABRAS DE ACTIVIDAD (MAW)

La arquitectura de conmutación MAW (*Managed by Activity Words*) [11] está constituida por N puertos de entrada y N puertos de salida trabajando todos ellos a la misma velocidad. El eje de tiempo se divide en ranuras temporales cuya duración coincide con el tiempo de transmisión de una celda.

El sistema está formado por una red de conmutación por división en el espacio (*banyan*) y por N módulos de entrada. Los elementos de la red de conmutación son sin memoria y el encaminamiento de los paquetes hacia los respectivos puertos de salida se realiza mediante técnicas de auto-encaminamiento [17].

Cada módulo de entrada está conectado a un enlace entrante de la red de conmutación. Los módulos de entrada están formados básicamente por una cola FIFO donde se van almacenando las celdas multipunto. En la Fig. 4 se muestra el diagrama de bloques de un módulo de entrada.

3.1. Descripción del módulo de entrada

Cuando una celda MTA se incorpora al módulo de entrada, el bloque *generador de palabras* es el encargado de generar la palabra de actividad asociada a dicha celda.

La palabra de actividad esta formada por un total de N bits. Cada uno de estos bits está asociado a uno de los N puertos de salida. La activación del bit i-ésimo significará que la celda multipunto desea encaminar una de sus réplicas hacia el puerto de salida i. Lógicamente, el generador de palabras utilizará tanto los identificadores de camino y circuito virtual como la información de la tabla de encaminamiento (asignada durante la fase de establecimiento de la conexión) para determinar los bits que se deben activar.

Paralelamente al proceso anterior, el módulo *generador de direcciones (escritura)* es el encargado de generar la dirección de la memoria BUFFER donde se almacenará el campo de información de la celda (*payload*) y su palabra de actividad. Puesto que la disciplina de esta memoria es FIFO, el generador de direcciones puede construirse a partir de un contador cíclico módulo M, siendo M el número máximo de elementos que pueden almacenarse en la memoria. El valor de M debe elegirse de forma que se garantice un valor aceptable de la probabilidad de pérdida de celda y de retardo de conmutación.

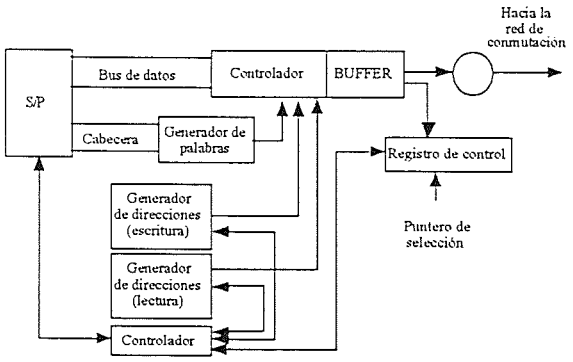


Fig. 4. Diagrama de bloques de un módulo de entrada de la arquitectura MAW.

El *generador de direcciones (lectura)* proporciona la dirección de memoria para la lectura de la primera celda almacenada en el BUFFER y de su palabra de actividad asociada. Este generador de direcciones también puede implemetarse a partir de un contador cíclico módulo M.

En el *registro de control* se almacena la palabra de actividad asociada a la celda en servicio. El *puntero de selección* apunta a un determinado bit de la palabra de actividad. Puesto que cada uno de los bits de la palabra de actividad está asociado a un puerto de salida, el valor del *puntero de selección* junto con el estado del bit, indica si la celda en servicio requiere una copia hacia un determinado puerto de salida.

Finalmente, cada bloque funcional incorpora su correspondiente controlador para implemetar tareas de gestión y control.

3.2. Proceso de copia

En esta arquitectura, el proceso de copia de una celda multipunto consiste simplemente en la generación de la palabra de actividad. En las arquitecturas basadas en etapas de memoria compartida, el sistema tiene que almacenar, para cada una de las réplicas, la dirección de memoria donde se ubica la celda multipunto. Si se desea que el sistema realice todas las copias en una misma ranura temporal, incluso en situaciones donde el número de copias es muy elevado, no queda otra alternativa que la de utilizar memorias muy rápidas (y en consecuencia, costosas).

Sin embargo, en la arquitectura MAW, el proceso es independiente del número de copias requerido por la celda multipunto. De esta forma, se consigue que las exigencias en las prestaciones de las memorias sean menos restrictivas. Esta propiedad permite que la aceleración del sistema también sea más fácil de implementar y menos costosa en comparación con otro tipo de arquitecturas.

3.3. Política de selección

La política de selección es la encargada de decidir si las celdas ubicadas en los servidores de cada uno de los módulos de entrada pueden ser encaminadas hacia la red de conmutación. Dicha política de selección se lleva a cabo siguiendo un algoritmo cíclico módulo N. Al inicio de cada ranura temporal, el *puntero de selección* de cada módulo de entrada apunta a uno de los bits de la palabra almacenada en el *registro de control*. Supóngase que en la ranura temporal t, el *puntero de selección* del módulo de entrada j apunta al bit i de la correspondiente palabra de actividad. Entonces:

- En el módulo de entrada $(j+1) \bmod N$, durante la ranura temporal t, el puntero de selección apunta al bit $(i+1) \bmod N$.
- En el módulo de entrada j, durante la ranura temporal $(t+1)$, el puntero de selección apuntará al bit $(i+1) \bmod N$.

De acuerdo con la política de selección anterior, todas las celdas seleccionadas tienen siempre direcciones destino distintas, con lo que evitamos el problema de la contención a la salida. Además, puesto que las direcciones destino están ordenadas monótonamente (mod N) también eliminamos el bloqueo interno en la red de auto-encaminamiento utilizada.

La selección de una copia para su posterior encaminamiento hacia la red de conmutación sigue el siguiente algoritmo: en cada ranura temporal, el *puntero de selección* inspecciona el bit i-ésimo del *registro de control*:

- Si el bit está activo, el módulo de entrada libera hacia la red de conmutación una copia de la celda en servicio (notar que su dirección de salida coincide con el valor del *puntero de selección*). El bit es puesto a cero y se comprueba la presencia de más bits activos. Si la comprobación es negativa (ya se han realizado todas las copias), la celda en servicio y su correspondiente palabra de actividad pueden ser liberadas.
- Si el bit está desactivado, no se realiza ninguna acción.

En cualquier caso, en la siguiente ranura temporal, el *puntero de selección* se incrementa en una unidad (mod N) y se repite el proceso anterior. En la Fig. 5 se muestra un ejemplo del funcionamiento de la política de selección durante 4 ranuras temporales consecutivas.

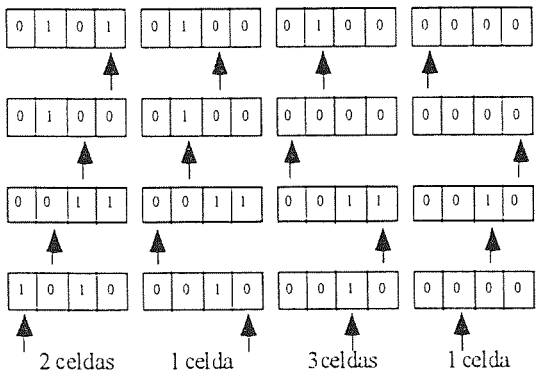


Fig. 5. Ejemplo del funcionamiento de la política de selección durante cuatro ranuras temporales consecutivas ($N = 4$).

Para liberar una celda del módulo de entrada basta con incrementar el contenido del *generador de direcciones (lectura)* en una unidad (mod N). Automáticamente, la celda ubicada en el servidor es

reemplazada por la celda almacenada en la cabecera de la memoria BUFFER. Paralelamente, su palabra de actividad asociada (almacenada también en la misma posición dentro del BUFFER) es transferida al registro de control.

Para determinar el estado de ocupación de las colas, hay que comparar el contenido de los dos generadores de direcciones. Dicha función se lleva a cabo mediante el módulo *controlador*. Si las colas están completamente llenas, el convertidor Serie/Paralelo deberá rechazar las nuevas celdas que aparezcan a su entrada.

3.4. Análisis de estabilidad

El aspecto positivo de la simplificación conseguida en esta arquitectura es la facilidad que presenta su implementación *hardware*. En contrapartida, y debido a la propia simplificación, el sistema es inestable a partir de una determinada carga efectiva. El objetivo ahora es determinar el valor de la carga efectiva máxima, a partir de la cual la longitud de las colas de entrada crece indefinidamente.

Con objeto de simplificar el estudio analítico, se asumen las siguientes hipótesis de partida:

- Todos los módulos de entrada son estadísticamente iguales (tráfico balanceado). En consecuencia, el estudio se centra en un único módulo de entrada.
- El módulo de entrada se modela mediante una cola FIFO de celdas multipunto en espera de servicio junto con un servidor (Fig. 6). La celda que contenga el servidor permanecerá en servicio hasta que finalice el encaminamiento de todas sus copias.
- Longitud de las colas de entrada supuesta infinita.
- El patrón estadístico del tráfico de entrada será de Bernoulli con parámetro p.
- El número de copias demandado por las celdas multipunto es constante e igual a n.
- Las direcciones destino siguen una ley estadística uniforme entre todos los puertos de salida.

Definimos la variable aleatoria K como el número total de llegadas que aparecen en un intervalo de tiempo de T ranuras temporales (sigue una ley binomial)

$$P(K=k)=\binom{T}{k}p^k(1-p)^{T-k} \qquad (1)$$

el intervalo de tiempo medio transcurrido entre dos llegadas consecutivas es

$$t=\frac{T}{E[K]}=\frac{1}{p} \qquad (2)$$

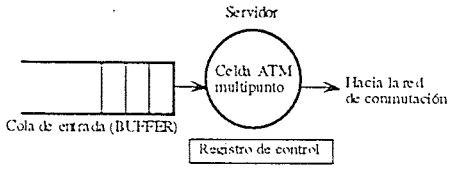


Fig. 6. Modelo utilizado para los módulos de entrada.

definiendo $1/\mu$ como el tiempo medio de servicio y aplicando el criterio de estabilidad se obtiene

$$\frac{1}{p}>\frac{1}{\mu} \qquad (3)$$

donde nuestra única incógnita resulta ser $1/\mu$.

En cada ranura temporal la política de selección examina una única posición del *registro de control*. Al tratarse de una política cíclica en el tiempo, puede asegurarse que, en el peor de los casos, al cabo de N ranuras temporales la celda finalizará el proceso de réplica y por tanto el servicio. Además, ya que en una misma ranura temporal el sistema sólo es capaz de encaminar una única copia, también se observa que la celda permanecerá en servicio como mínimo n ranuras temporales. Por lo tanto, el valor del tiempo de servicio r está acotado por

$$n \le r \le N \quad n \in \{1,\ldots,N\}$$

y en consecuencia

$$\frac{1}{\mu}=\sum_{r=n}^{N} rP_r(r) \qquad (4)$$

Donde fijada una ranura temporal, $P_I(i)$ se define como la probabilidad de que la política de selección visite el bit i del registro de control. Dicha probabilidad condicionará el valor que puede tomar el tiempo de servicio

$$P_r(r)=\sum_{i=0}^{N-1} P_r(r|i)P_I(i) \qquad (5)$$

Puesto que la política de selección es cíclica en el tiempo, fijada una ranura temporal cualquiera, todos los bits del registro de control tienen la misma probabilidad de ser atendidos

$$P_I(i)=\frac{1}{N} \qquad (6)$$

y sustituyendo en (5)

$$P_r(r)=\frac{1}{N}\sum_{i=0}^{N-1} P_r(r|i) \qquad (7)$$

Además, puesto que la distribución de las

direcciones destino es uniforme, la probabilidad $P_r(r|i)$ no depende del valor inicial que tome el *puntero de selección*

$$P_r(r|i) = P_r(r|j) \quad \forall i, j \tag{8}$$

y en consecuencia

$$P_r(r) = \frac{1}{N} N P_r(r|i) = P_r(r|i) \tag{9}$$

Para el cálculo de (9) se utilizará la definición clásica de probabilidad (casos favorables dividido por el número de casos posibles) y a partir de (8) se supone, sin pérdida de generalidad, que el *puntero de selección* se sitúa inicialmente en la posición *0* del *registro de control*.

Los casos posibles vienen determinados por el conjunto de combinaciones de n bits activados sobre el total de los N bits que conforman *el registro de control*

$$C_N^n = \binom{N}{n} \tag{10}$$

Para que el retardo valga r, el bit situado en la posición r-1 debe estar activado y todos los bits situados en posiciones posteriores deben de estar inactivos. El resto de los bits activados (n-1) deben distribuirse de cualquier orden en las posiciones anteriores a la r-1 (fig. 7). En consecuencia, los casos favorables son

$$C_{r-1}^{n-1} = \binom{r-1}{n-1} \tag{11}$$

y por lo tanto

$$P_r(r) = \frac{C_{r-1}^{n-1}}{C_N^n} = \frac{\binom{r-1}{n-1}}{\binom{N}{n}} \tag{12}$$

Con ello, el valor del tiempo medio de servicio

$$\frac{1}{\mu} = \sum_{r=n}^{N} r \frac{C_{r-1}^{n-1}}{C_N^n} \tag{13}$$

finalmente, aplicando la condición de estabilidad

$$\rho_{efec} < \frac{n}{\sum_{r=n}^{N} r \frac{C_{r-1}^{n-1}}{C_N^n}} \tag{14}$$

Siendo ρ_{efec} la carga efectiva de las líneas de salida, definida como la carga real de las líneas de entrada multiplicada por el número medio de copias.

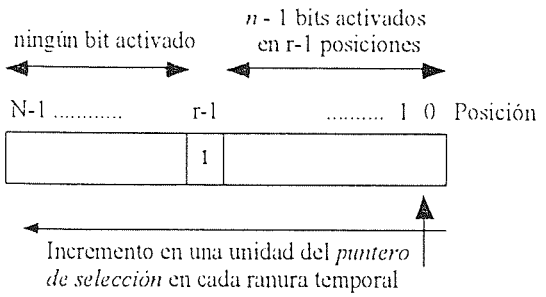


Fig. 7. Condición para que el retardo tome el valor r. Notar que el último bit activado debe situarse en la posición r-1.

En la tabla 1 se resumen los valores de carga efectiva máxima conseguida en esta arquitectura en función del número de copias. Se ha supuesto un tamaño del conmutador de $N = 16$.

3.5. Técnicas estabilizadoras: factor de aceleración (*speed-up*)

Como puede apreciarse en la tabla anterior, los valores máximos de carga efectiva que acepta el sistema básico son inaceptables. Así, por ejemplo, para un número de copias de 2, la carga efectiva máxima en las líneas de entrada no puede superar ni siquiera el 20% para que el sistema sea estable. Por ello se hace necesario la incorporación de técnicas que permitan aumentar el margen de estabilidad del sistema.

Una de estas técnicas es el uso de aceleración (*speed-up*), que consiste en aumentar en un factor S tanto la velocidad de proceso de los módulos de entrada como la velocidad de conmutación de la red. De esta forma, ahora, cada módulo de entrada está capacitado para encaminar, en una misma ranura temporal, un máximo de S copias hacia sus respectivos puertos de salida.

El factor de aceleración obliga a modificar la referencia temporal utilizada hasta el momento en el análisis. La ranura temporal se divide en S partes iguales que se denotan como ciclos y que constituirán el nuevo patrón temporal. La política de selección sigue siendo la misma, sólo que ahora, el valor del *puntero de selección* se incrementará en una unidad (mod N) en cada ciclo.

Tabla 1. Tiempos de servicio (expresados en ranuras temporales) y las correspondientes cargas efectivas máximas permitidas en la arquitectura MAW (N=16).

	Número de copias (n)					
	1	2	3	4	5	6
1/μ	8.5	11.33	12.75	13.6	14.16	14.57
ρef	0.117	0.176	0.235	0.294	0.353	0.418

Nótese que en el supuesto de que la velocidad de transmisión de las líneas de entrada-salida coincidan, la incorporación de la aceleración nos obliga a utilizar colas también en los puertos de salida. Por lo tanto, la estabilidad del sistema vendrá condicionada por la estabilidad de ambas colas. Pruebas realizadas mediante simulación confirman que la inestabilidad dominante es la que se produce en las colas de entrada, por lo que ahora se centrará nuestro estudio en el análisis de los módulos de entrada. Un estudio analítico y por simulación sobre el comportamiento de las colas de salida puede encontrarse en [5][6].

Para facilitar el cálculo, se supone que tanto el tamaño del conmutador N como el factor de aceleración S son potencias naturales de 2. La metodología utilizada en este apartado es la misma que la empleada en el apartado anterior, sólo que ahora, puesto que cada módulo de entrada es capaz de encaminar en una misma ranura temporal un total de S réplicas, en vez de considerar los bits del *registro de control* de forma individualizada, se trabajará con grupos de S bits (Fig. 8).

Las cotas superior e inferior del tiempo de servicio son

$$r_1 \le r \le r_2$$
$$r_1 = \left\lceil \frac{n}{S} \right\rceil \text{ y } r_2 = \frac{N}{S} \qquad (15)$$

Donde $\lceil x \rceil$ indica el entero superior a x. Las posibles combinaciones de n bits activados en el *registro de control* siguen siendo las mismas que para el caso anterior. Los casos favorables pueden calcularse a partir de la siguiente expresión

$$\sum_{i=1}^{Q} C_S^i C_{(r-1)S}^{n-i} \qquad (16)$$

$$Q = \begin{cases} n & n \le S \\ S & n > S \end{cases}$$

y en consecuencia

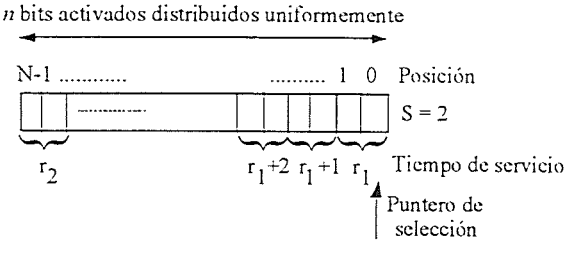


Fig. 8. Ejemplo de estabilidad con factor de aceleración S=2.

$$P_r(r) = \frac{\sum_{i=1}^{Q} C_S^i C_{(r-1)S}^{n-i}}{C_N^n} \qquad (17)$$

finalmente

$$\frac{1}{\mu} = \sum_{r=\lceil n/S \rceil}^{N/S} r \frac{\sum_{i=1}^{Q} C_S^i C_{(r-1)S}^{n-i}}{C_N^n} \qquad (18)$$

En la tabla 2 se resumen los valores máximos de carga efectiva en función del número de copias y del factor de aceleración. Se ha supuesto que el tamaño del conmutador es de $N = 16$. Se puede observar que, por ejemplo, con un factor de aceleración S=4 el sistema ya es totalmente estable a partir de un número de copias superior a 3.

Mencionar, por último que los resultados expuestos en las tablas 1 y 2 se han contrastado con los obtenidos mediante simulación. En ambas situaciones (con y sin aceleración) los valores teóricos se ajustan fielmente con los resultados de las simulaciones [18].

3.6. Técnicas estabilizadoras: política de selección adaptativa

Si al sistema original se le incorpora el valor de aceleración adecuado, las prestaciones que ofrece la arquitectura bajo la presencia de tráfico uniforme son aceptables. Sin embargo, cuando el tráfico en la entrada es a ráfagas, las prestaciones vuelven a empeorar drásticamente.

Dicho tráfico es de vital importancia en el entorno de las redes de alta velocidad, puesto que se admite que la incorporación de arquitecturas de conmutación MTA en los ámbitos comerciales e industriales se llevará a cabo mediante la emulación de redes LAN (LANE). En dichas redes, las tramas Ethernet-Token Ring deben segmentarse en grupos de 48 octetos, cada uno de los cuales está asociado con una celda MTA. Este conjunto de celdas determinarán el tamaño de la ráfaga, mientras que el tiempo transcurrido entre la transmisión de dos tramas LAN consecutivas determinará la duración del periodo de silencio.

Tabla 2. Cargas efectivas máximas que se obtienen gracias a la incorporación del factor de aceleración. El guión indica que el sistema es estable para cualquier carga efectiva inferior a la unidad.

copias	Factor de aceleración		
	2	4	8
1	0.22	0.4	0.67
2	0.34	0.63	-
3	0.45	0.86	-
4	0.57	-	-
5	0.69	-	-
6	0.8	-	-
7	0.92	-	-
8	-	-	-

En situaciones de muy baja carga, el retardo medio sufrido por una celda MTA es aproximadamente igual al tiempo medio de servicio. A partir de las ecuaciones encontradas en los apartados anteriores se observa que para tráfico punto a punto y con un factor de aceleración igual a la unidad, el valor de este retardo, expresado en ranuras temporales, es de $(N+1)/2$ (siendo N el número de puertos de entrada/salida del conmutador). El motivo por el cual este retardo es tan elevado es debido a que la celda debe esperar a que el *puntero de selección* apunte hacia la dirección de salida demandada. Cuando el tráfico es a ráfagas la situación empeora, puesto que al tener todas las celdas que pertenecen a una misma ráfaga la misma dirección destino, para que la política de selección vuelva a seleccionar la misma dirección destino, deberán transcurrir N ranuras temporales. De aquí que las prestaciones de esta arquitectura caigan en picado cuando es sometida a este tipo de tráfico.

Debido a la inherente capacidad *broadcast* que presentan las redes LAN, es de vital importancia que el conmutador empleado disponga de capacidad multipunto. El objetivo ahora es diseñar una nueva política de selección que consiga mejorar las prestaciones de esta arquitectura cuando el tráfico de entrada es a ráfagas y simultáneamente, que siga manteniendo una cierta simplicidad *hardware*. La eficacia de esta nueva política de selección se fundamenta en el hecho de que todas las celdas que conforman una ráfaga tienen el mismo conjunto de direcciones destino (puesto que, por ejemplo, todas ellas forman una misma trama LAN).

La política de selección empleada hasta el momento (clásica) tiene un comportamiento cíclico en el espacio y en el tiempo. La nueva política de selección sigue siendo cíclica en el espacio pero deja de serlo en tiempo. Con el objeto de maximizar el número de celdas que se encaminan hacia sus respectivos puertos de salida, en cada ranura temporal el sistema estima la posición óptima del *puntero de selección* en el primer módulo de entrada (módulo 0). Puesto que la política de selección sigue siendo cíclica en el espacio, la posición del *puntero de selección* en el resto de los módulos de entrada se calcula incrementando sucesivamente en una unidad (módulo N) el valor encontrado para el módulo *0*.

En la Fig. 9 se muestra un ejemplo de funcionamiento. En ella se observa que durante la ranura temporal k, la posición óptima del puntero de selección en el primer módulo de entrada es la 2 puesto que es la que consigue encaminar un total de 4 celdas. En la siguiente ranura temporal $(k+1)$ la posición óptima del puntero pasa a ser la 0 puesto que ahora esta posición es la que maximiza el número de celdas encaminadas. No existe, por tanto, una relación cíclica entre ranuras temporales y posiciones del *puntero de selección*.

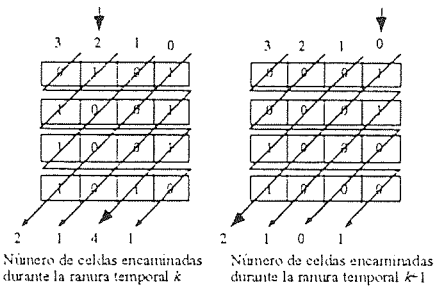


Fig. 9. Ejemplo del funcionamiento de la política adaptativa durante dos ranuras temporales consecutivas (N=4).

Para determinar la posición óptima se tiene que calcular el número total de celdas que se pueden encaminar en cada una de las posibles posiciones del *puntero de selección*. Una de las posibles formas de determinar dicha posición es mediante una cadena de sumadores y un comparador, como se muestra en la Fig. 10. En caso de empate (es decir, cuando no hay una única posición óptima sino varias) el sistema escoge aquella que lleva más tiempo sin ser atendida. Ello obliga a incorporar un contador adicional para cada posición del *registro de control* con el objeto de contabilizar el tiempo en que una determinada posición permanece sin ser atendida. Cuando dicho contador llegue a su valor máximo, la réplica asociada a esa posición debe servirse de forma inmediata, independientemente de la posición óptima calculada por el sistema. Con la incorporación de estos contadores se consiguen evitar posibles situaciones de injusticia y simultáneamente se mantiene la simplicidad *hardware*.

4. RESULTADOS DE SIMULACIÓN

En esta sección se compararán y se analizarán los resultados de simulación obtenidos con la arquitectura de conmutación MAW utilizando tanto la primera política de selección cíclica como la nueva política adaptativa.

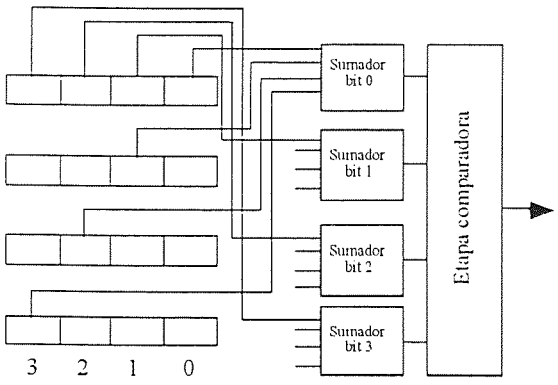


Fig. 10. Implementación *hardware* para determinar la posición óptima *del puntero de selección* (N = 4).

La figura de prestaciones que se presenta es el retardo medio de conmutación, definido como el tiempo medio (expresado en ranuras temporales) desde que una celda multipunto entra a un determinado módulo de entrada hasta que su última copia abandona el sistema.

Se han utilizado los modelos de tráfico uniforme (con generación Bernoulli) y a ráfagas (siguiendo un modelo ON-OFF) y en ambos casos, las direcciones de salida de las celdas multipunto siguen una ley uniforme. El número de copias n se ha supuesto constante, por ser éste el caso más desfavorable frente a un número de copias distribuido con arreglo a una función estadística truncada al valor que se toma como constante.

En cualquier caso, el número de copias nunca excede del número de puertos de salida. El tamaño del conmutador se ha fijado al valor de N=16 ya que para la construcción de conmutadores de mayor tamaño es preferible utilizar la metodología expuesta en [13]. El factor de aceleración es de S=4 por ser este el valor que maximiza la relación entre las prestaciones del sistema y la capacidad de implementar dicha aceleración de forma factible y realista.

4.1. Tráfico uniforme

La primera consideración que se observa es que a medida que el número de copias aumenta, mejoran las prestaciones de la arquitectura, independientemente de la política de selección empleada. En efecto, como se ha visto en los apartados anteriores, el tiempo de servicio medio aumenta a medida que crece el número de copias. Sin embargo, la carga real de las líneas de entrada (definida como el cociente entre la carga efectiva y el número medio de copias) disminuye en una proporción inversa todavía mayor. Así, por ejemplo, y en referencia a los valores calculados en la tabla 1, se observa que el tiempo de servicio medio con un número de copias igual a 6 es 14,57 unidades de tiempo (ranuras temporales). Si comparamos éste valor con el calculado para 3 copias (12,75) se observa claramente que la relación entre ambas magnitudes es cercana a la unidad. Por contra, la carga real de las líneas de entrada (fijada una carga efectiva cualquiera) para el caso $n = 6$ es exactamente la mitad de la utilizada para el caso $n = 3$.

Además, con la política de selección adaptativa se consigue un doble resultado: en primer lugar, el margen de estabilidad aumenta (sobre todo para el caso $n = 2$) y en segundo lugar, el retardo medio de conmutación para cargas bajas es también bajo.

También se observa que cuando la carga efectiva de las líneas de entrada es muy alta, la política de selección adaptativa ya no aporta ninguna ventaja con respecto la política cíclica. Ello es debido a que, bajo estas condiciones, el número de empates que se producen para poder determinar la posición óptima del *puntero de selección* es demasiado elevado.

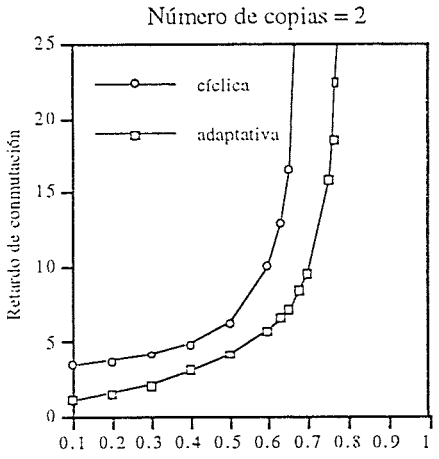


Fig. 10. Retardo de conmutación medio (expresado en ranuras temporales) en función de la carga efectiva para tráfico aleatorio con un número de copias constante e igual a 2.

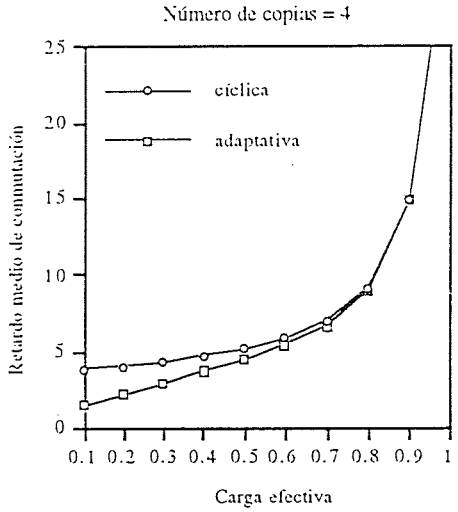


Fig. 11. Retardo de conmutación medio (expresado en ranuras temporales) en función de la carga efectiva para efectiva para tráfico aleatorio con un número de copias constante e igual a 4.

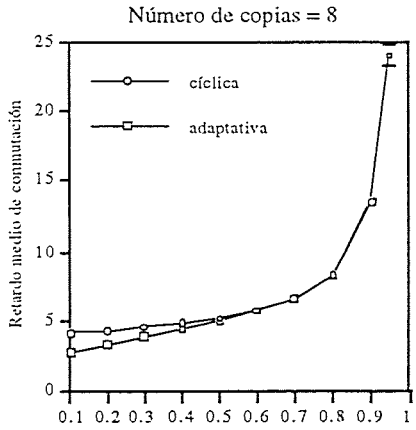


Fig. 12. Retardo de conmutación medio (expresado en ranuras temporales) en función de la carga efectiva para efectiva para tráfico aleatorio con un número de copias constante e igual a 8.

4.2. Tráfico a ráfagas

Para este caso se ha supuesto una duración del periodo de actividad (T_{on}) de 5 ranuras temporales. El valor del periodo de silencio se determina a partir de la carga efectiva empleada utilizando la siguiente expresión:

$$T_{off} = \frac{T_{on}(\rho_{on} - \frac{\rho_{efec}}{n})}{\frac{\rho_{efec}}{n}}$$

Siendo ρ_{on} la probabilidad de generación de celdas en una ranura temporal incluida dentro del periodo de actividad. En todas las simulaciones que se presentan se ha supuesto $\rho_{on} = 1$.

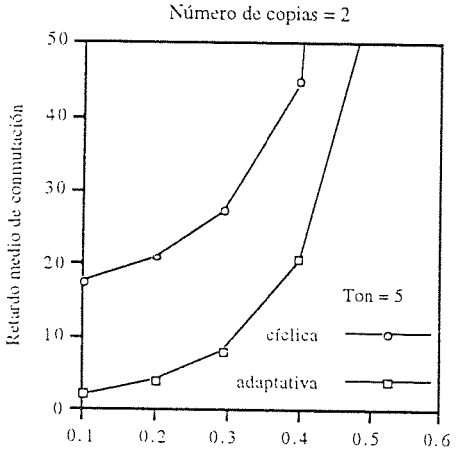


Fig. 13. Retardo de conmutación medio (expresado en ranuras temporales) en función de la carga efectiva para tráfico a ráfagas (Ton = 5) con un número de copias constante igual a 2.

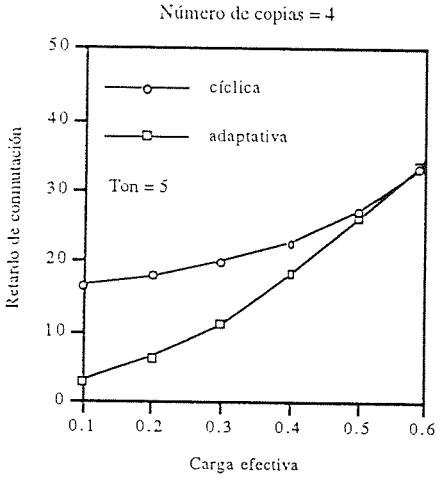


Fig. 14. Retardo de conmutación medio (expresado en ranuras temporales) en función de la carga efectiva para tráfico a ráfagas (Ton = 5) con un número de copias constante igual a 4.

A partir de las gráficas se observa que, al igual que en el caso anterior, el retardo medio disminuye a medida que el número de copias aumenta. Además, gracias a la presencia de la política adaptativa, el valor de dicho retardo toma valores aceptables en el margen de cargas efectivas bajas e intermedias.

5. CONCLUSIONES

En este artículo se ha presentado una arquitectura de conmutación MTA con capacidad multipunto cuyas características principales son:

- Simplicidad *hardware*.

En efecto, en esta arquitectura, la aceptación de tráfico multipunto se lleva a cabo mediante la generación de una palabra de actividad asociada a la celda multipunto. En consecuencia, las exigencias en la tecnología empleada son menos restrictivas que las que presentan los sistemas basados en el uso de módulos de memoria compartida (replicación lógica) o en red de copia (replicación física).

- Adaptación natural al tráfico multipunto.

Como se observa en las simulaciones que se han presentado, las prestaciones de retardo medio de conmutación mejoran cuando aumenta el número de copias.

- Estabilidad

Aunque la arquitectura básica presenta problemas de inestabilidad, la adición de técnicas de aceleración junto con la incorporación de una política de selección adaptativa al tráfico de entrada permiten aumentar el margen de estabilidad del sistema. A partir de los resultados obtenidos en las simulaciones, podemos concluir que con una aceleración de orden 4, se consigue aumentar el margen de estabilidad de la arquitectura hasta cotas razonables.

- Aceptación de tráfico a ráfagas

Con la incorporación de la política de selección adaptativa, el sistema es capaz de reducir sensiblemente el retardo medio de conmutación para cargas bajas e intermedias y simultáneamente, se consigue aumentar el margen de estabilidad. Ello permite que bajo la presencia de tráfico a ráfagas, las prestaciones de la arquitectura se mantengan dentro de cotas aceptables.

6. LÍNEAS FUTURAS

- Desarrollar un análisis matemático similar al expuesto en este artículo, que permita modelar el comportamiento del sistema cuando éste utiliza la política de selección adaptativa.

- Con el objeto de aumentar el tamaño del conmutador sin que ello repercuta de forma drástica en sus prestaciones, se propone el estudio y evaluación de diversas técnicas de crecimiento basadas en la interconexión de redes de tres etapas (redes de Clos).

Agradecimientos

Este trabajo ha sido parcialmente financiado por el proyecto de investigación SIGLA (CICYT TEL96-1452).

Referencias

[1] T. T. Lee, "Nonblocking Copy Network for Multicast Packet Switching", *IEEE J. Selected Areas on Comm.*, Vol. 6, No 9, pp 1455-1467, Dec 1988

[2] C. J. Chang and C. J. Ling, "Overflow Controller in Copy Network of Broadband Packet Switch", *Electronic Letters*, Vol. 27, No 11, pp 937-939, 23rd, May 1991.

[3] R. P. Bianchini Jr and H. S. Kim, "Design of a Nonblocking Shared-Memory Copy Network for ATM", *Proc. of IEEE Infocom'92*, pp 876-885. May 1992.

[4] W. De Zhong, Y. Onozato and Kaniyil, "A Copy Network with Shared Buffers for Large Scale Multicast ATM Switching", *IEEE/ACM Trans. on Networking*, Vol. 1, No 2, pp 157-165. Apr. 1993.

[5] Coudreuse, J. P. , and M. Servel, "Prelude: An Asynchronous Time Division Packet Switched Network", *ICC'87*, 1987, pp 769-773.

[6] N. Endo, T. Kozaki, T. Ohuchi, H. Kuwahara, S. Gohara, "Shared Buffer Memory Switch for an ATM Exchange". *IEEE Trans. on Comm*, Vol 41, No 1, Jan 1993.

[7] T. Kozaki, N. Endo, Y. Sakurai, O. Matsubara, M. Mizukami, "32x32 Shared buffer type ATM switch VLSI's for B-ISDN's", *IEEE J. Selected Areas on Comm*, Vol. 9, No 8, pp 1239-1247, Oct 1991.

[8] Y. Shobatake, M. Motoyama, E. Shobatake, T. Kamitake, S. Shimizu, M. Noda, K. Sakaue, "A One-Chip Scalable 8x8 ATM Switch LSI Employing Shared Buffer Architecture". *IEEE J. Selected Areas on Comm*, Vol. 9, No 8, pp 1248-1254, Oct 1991.

[9] Kuwahara, H. et. al., "Shared Buffer Memory Switch for an ATM Exchange", *ICC'90*, Boston 1989.

[10] J. García-Haro, Andrzej-Jajszczyk, "ATM Shared-Memory Switching Architectures", *IEEE Network Magazine*, Vol 8, No. 4, Jul/Aug. 1994, pp 18-26.

[11] J. García-Haro, J. Malgosa-Sanahuja, J. L. Melús-Moreno, "Multicasting facilities in ATM switching architectures. A study of several approaches". *IEEE Pacific RIM Conference on Comm*, May 1995.

[12] M. J. Karol, M.G. Hluchyj, S. P. Morgan, "Input Versus Output Queueing on a Space-Division Packet Switch", *IEEE Trans. Comm.*, Vol. COM-35, pp 1347-1356, Dec. 1987.

[13] K. Y. Eng, M.J. Karol, Y.S. Yeh, "A Growable Packet (ATM) Switch Architecture: Design Principles and Applications", *GLOBECOM'89 Conf. Rec.*, Nov. 1989, pp 1159-1165.

[14] M. J. Karol, I. Chih-Lin, "Performance Analysis of a Growable Architecture for Broadband Packet (ATM) Switching", *IEEE Trans. Comm.*, Vol. 40, No. 2, pp 431-439.

[15] Soung C. Liew, "Multicast Routing Algorithms for 3-Stage Clos ATM Switching Networks", *GLOBECOM'91*, Jan. 1991, pp 1916-1925.

[16] D. J. Marchok, C. E. Rohrs, "First Stage Multicasting in a Growable Packet (ATM) Switch", *ICC'91*, March 91, pp 1007-1013.

[17] X. Chen, "A Survey of Multistage Interconnection Networks in Fast Packet Switches", *International Journal of Digital and Analog Communication Systems*, Vol 4, 1991, pp 33-59.

[18] Alberto Gimeno Cardo, "Estudio y Evaluación por Simulación de la Arquitectura de Conmutación multipunto MAW con Política de Selección Adaptativa al Tráfico de Entrada", *Proyecto Fin de Carrera*, Universad de Zaragoza, 1996.

ABC´96: Un Servicio de Teleeducación sobre ATM de Ambito Intercontinental.

Juan Quemada Vives, Tomas de Miguel Moro, Arturo Azcorra, Santiago Pavón, Joaquin Salvachúa, Manuel Petit, David Larrabeiti, Tomas Robles, Gabriel Huecas.
DIT - UPM

Pedro Chas, Roberto de Isidro, Carmen Cesar, Carlos Leon.
TELEFONICA I+D

La existencia de redes de comunicaciones de banda ancha de ambito europeo como es la Red Piloto Paneuropea ATM, permite la creación de nuevos servicios con un enorme potencial en el mundo de la educación. Las serie de Escuelas de Verano sobre Comunicaciones Avanzadas de Banda Ancha que comenzó en 1993 con ABC´93, y cuya ultima edición ABC´96 se realizo en Julio de 1996, ha tenido por objeto mostrar las posibilidades que la tecnología actual tiene en este campo. Su objetivo fué desde un principio la interconexión de auditorios o aulas remotas creando aulas virtuales de ambito Europeo, o intercontinental, donde los asistentes pueden comunicarse sin las limitaciones que impone la separación fisica. La Universidad Politécnica de Madrid y TELEFONICA I-D han participado en su realización de desde el comienzo.

Estas escuelas de verano han supuesto un hito en la realización de eventos distribuidos y han creado una infraestructura que tiene multiples utilidades en el entorno universitario. Permite distribuir clases entre centros, realizar tele-reuniones, compartir congresos y seminarios, ...

Pasamos a describir la ultima escuela de verano distribuida ABC´96 como ejemplo de creación de un nuevo tipo de servicio de tele-educación.

ABC´96

La IV "Escuela de Verano Internacional Distribuida sobre Comunicaciones de Banda Ancha - ABC´96" ha sido un evento distribuido donde mas de 20 auditorios situados en 16 países se han interconectado en un enorme auditorio virtual de cobertura intercontinental. Se realizó del 9 al 12 de Julio de 1996.

ABC'96 ha tenido cinco sedes principales desde donde los conferenciantes hicieron sus presentaciones y desde donde existió acceso abierto a cualquier participante. Los auditorios principales estuvieron situados en Madrid, Aveiro, Bruselas, Berlín y Nápoles. En Madrid la sede principal estuvo situada en la ETSI de TELECOMUNICACION de la UPM y fue organizada en colaboración con TELEFONICA de España y TELEFONICA I+D.

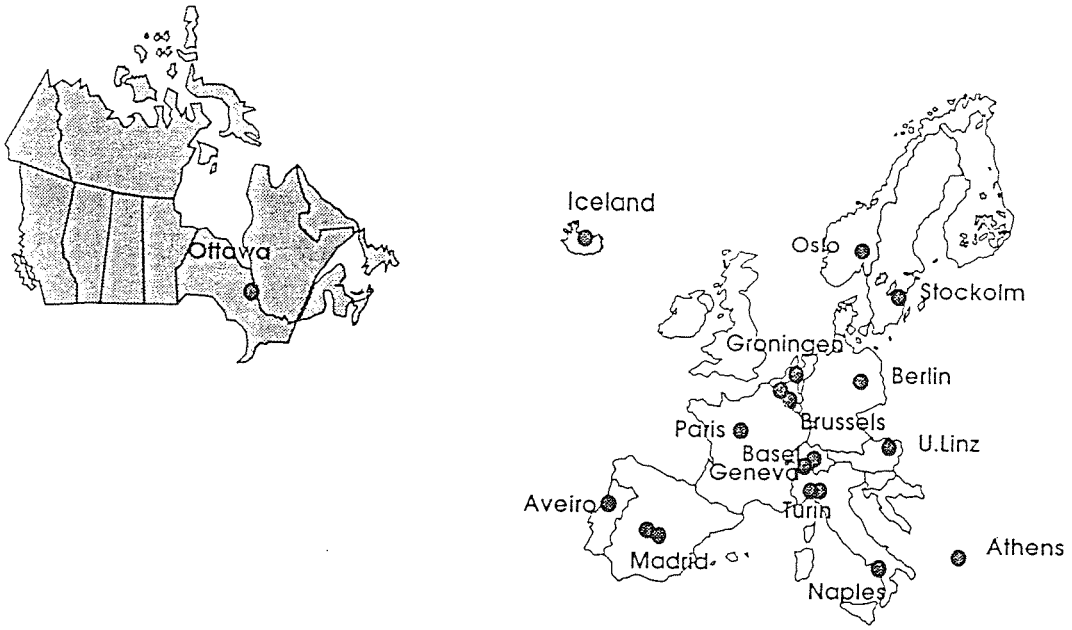
Se puede encontrar mas información en:
http://www.dit.upm.es/~proy/nice/abc96

Las otras 4 sedes principales estuvieron situadas en: Belgacom - Bruselas, CRIAI - Nápoles, Universidad de Aveiro - Portugal, Universidad Técnica de Berlín - Alemania.

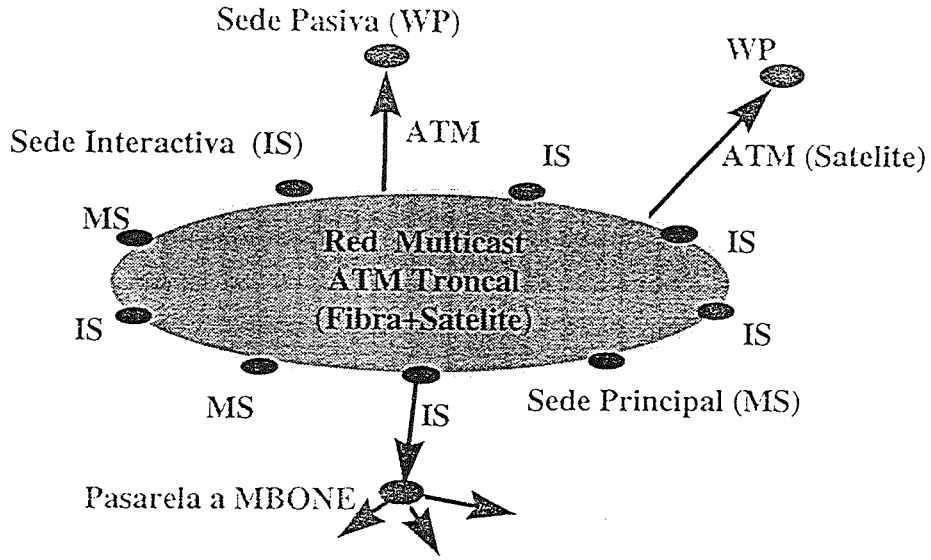
Adicionalmente, existieron otras 16 sedes subsidiarias en ABC '96 donde pudo asistir audiencia local. Estas fueron:

CERN	Ginebra, Suiza
CNET	París, Francia
CRC - BADLAB	Ottawa, Canadá
CSELT	Turín, Italia
Euro-National Host	Bruselas, Bélgica
KPN	Groningen, Holanda
Politecnico de Torino	Turín, Italia
PTI	Reiquiavic, Islandia
SICS	Estocolmo, Suecia
Swiss National Host	Basilea, Suiza
TELEFONICA I+D	Madrid, España
Telenor	Oslo, Noruega
Technical University of Prague	Praga, República Checa
University of Demokritos	Atenas, Grecia
University of Linz	Linz, Austria

La figura siguiente describe la cobertura geográfica de ABC'95.



El modelo de evento en que se basa una conferencia distribuida como es una escuela de verano se describe en la figura siguiente.



Existe un núcleo troncal compuesto por sedes interactivas. Entre estas se incluyen las sedes principales, donde se garantiza una calidad y una continuidad en el servicio. Esto se consigue con una red adicional de seguridad que puede entrar en funcionamiento si la primera falla. Los ponentes presentan sus ponencias de estas sedes principales. El acceso regular solo se permite en estas sedes, que son las que garantizan una continuidad del servicio.

Además existen sedes pasivas, en las cuales es posible seguir el desarrollo del evento, pero donde no es posible interaccionar con el resto.Dentro de este capitulo se pueden diferenciar tres categorías.

1) En primer lugar están las sedes pasivas conectadas a través de enlaces terrestres de tipo ATM en las cuales se tiene una calidad de presentación similar a las sedes interactivas.
2) En segundo lugar están las sedes pasivas conectadas a través de un enlace satélite de tipo ATM en las cuales se tiene una calidad de presentación similar a las sedes interactivas. El interés de este tipo de sedes estriba en que un solo canal ATM sirve para crear un numero ilimitado de sedes pasivas, con lo cual el impacto del coste de la comunicación en la cuota de asistencia se puede reducir hasta limites despreciables.
3) Por ultimo esta la distribución que se realizo a través de la red MBONE con las herramientas VIC y VAT, con las cuales se consiguió una difusión mundial, pudiendo conectarse cualquier estación de trabajo o PC que tuviese instaladas dichas herramientas y estuviese conectado a la red MBONE.

LA PLATAFORMA DE SERVICIO

La realización del evento distribuido ABC´96 se basa en una aplicación multimedia de tele-educación denominada ISABEL. Esta aplicación permite

interconectar los auditorios de todas las sedes transformando el conjunto en una Unica Sala de Conferencias Virtual.

La aplicación ISABEL ha sido desarrollada por un equipo de profesores y estudiantes del Departamento de Ingeniería de Sistemas Telemáticos de Universidad Politécnica de Madrid. ISABEL es una aplicación de CSCW (Computer Supported Co-operative Work).

La aplicación ISABEL tiene tres componentes :

- Un componente de tele-presencia que transmite los videos de las sedes activas a todos los auditorios.

- Un componente de espacio de trabajo compartido con el presentador puede controlar su presentación que es visualizada en todas las sedes.

- Un componente de gestión del evento que permite controlar lo que pasa en todos los auditorios de forma centralizada.

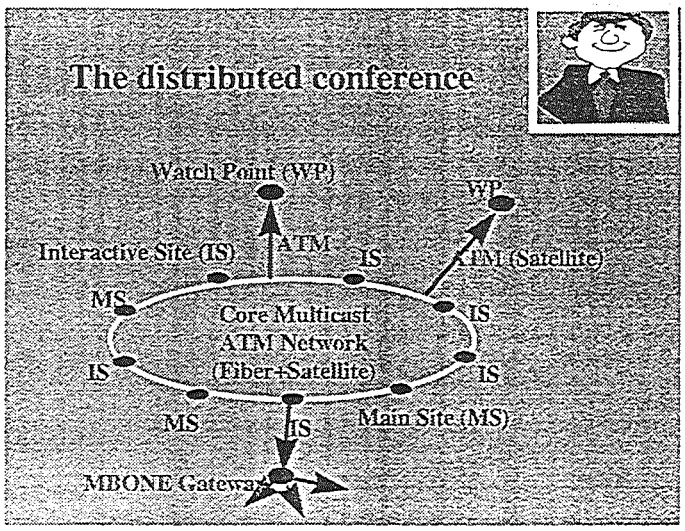
Estos tres componentes son típicos en aplicaciones de trabajo cooperativo, pero en ISABEL adquieren un nivel de integración que no se da en aplicaciones similares.

La Red de Comunicaciones utilizada es un embrión de las futuras Superautopistas de la Información sucesoras de la Internet actual. Incluye conexiones transnacionales por fibra óptica basadas en ATM, así como conexiones vía satélite e interconexión con la Internet. Esto ha sido posible gracias a una compleja colaboración entre diversas Instituciones, National Hosts, Operadores Públicos y proyectos de I+D.

Modos de Interacción

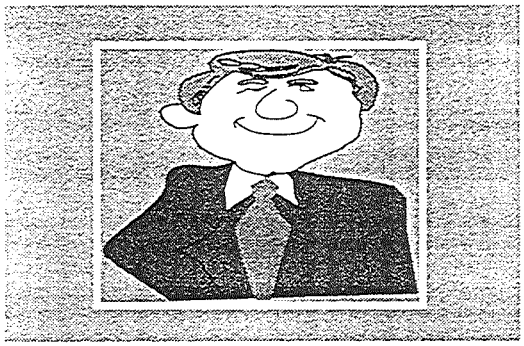
La realización de eventos distribuidos utilizando ISABEL se basa en el concepto de modo de interacción. Cada actividad distribuida tiene un modo de interacción en el cual unas sedes están activas y otras son pasivas. Para reflejar correctamente cada actividad en curso es conveniente utilizar para cada tipo de interacción entre sedes una configuración particular. Esto permite centrar mejor la atención de los participantes.

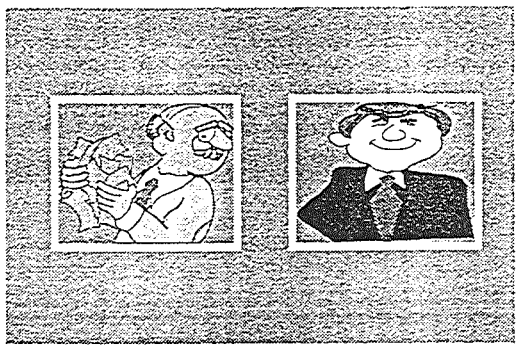
Como ejemplo vamos a mostrar algunos modos de interacción típicos. La primera figura representa el modo presentación, que se utiliza cuando un ponente esta presentando su ponencia ayudado de una transparencia. En él, solo se presenta la imagen del ponente en tamaño reducido y sus transparencia.



En este modo la atención debe centrarse en la transparencia y el vídeo del ponente tiene como misión crear sensación de presencia del ponente en todos los auditorios.

En la figura siguiente se presentan otros dos modos de interacción en los cuales solo se presenta el vídeo. El primer modo esta pensado para cuando un ponente habla sin ayudarse de otro material gráfico. El segundo para el turno de preguntas o para un debate entre dos, de modo que los videos de ambos interlocutores se vean en todos los auditorios.

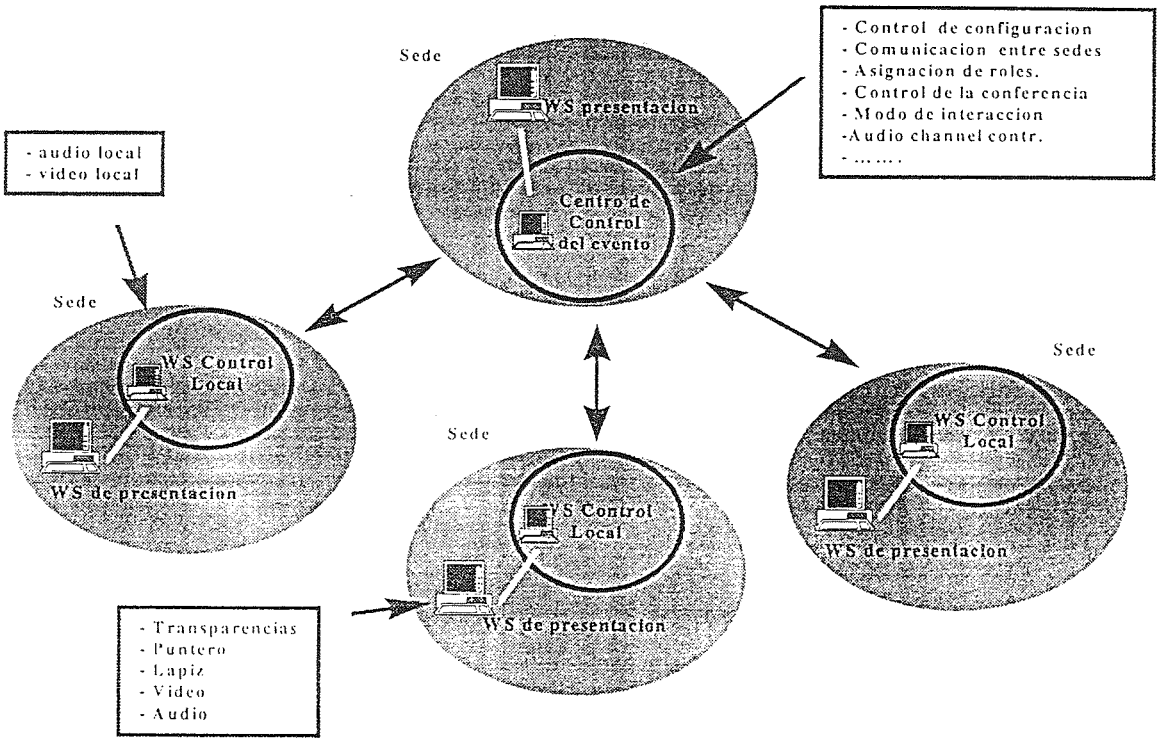




El Control de la Conferencia Distribuida

Es necesario disponer de un componente de gestión capaz de controlar el componente de tele-presencia y el espacio de trabajo compartido para que un evento distribuido se desarrolle con fluidez. Por eso en ISABEL se ha desarrollado un componente de gestión de una conferencia que permite controlar en todo momento el estado de presentación de los otros componentes en todas las sedes, principales, interactivas y pasivas.

El componente de gestión se ha diseñado siguiendo el modelo que se representa en la figura siguiente. En este modelo existe un centro de control del evento que se sitúa normalmente en una de las sedes principales. La aplicación ISABEL permite que desde este centro de control se decidan las configuraciones de las pantallas proyectadas en todos los auditorios así como los roles y capacidades de cada sede.



Ejemplos de los roles o capacidades que se pueden asignar a un auditorio son :

- Capacidad de controlar la realización de una presentación, es decir la selección y paso de transparencias, el puntero y el lápiz.

- Capacidad para hacer preguntas, es decir la inclusión del audio y vídeo de la sede seleccionada.

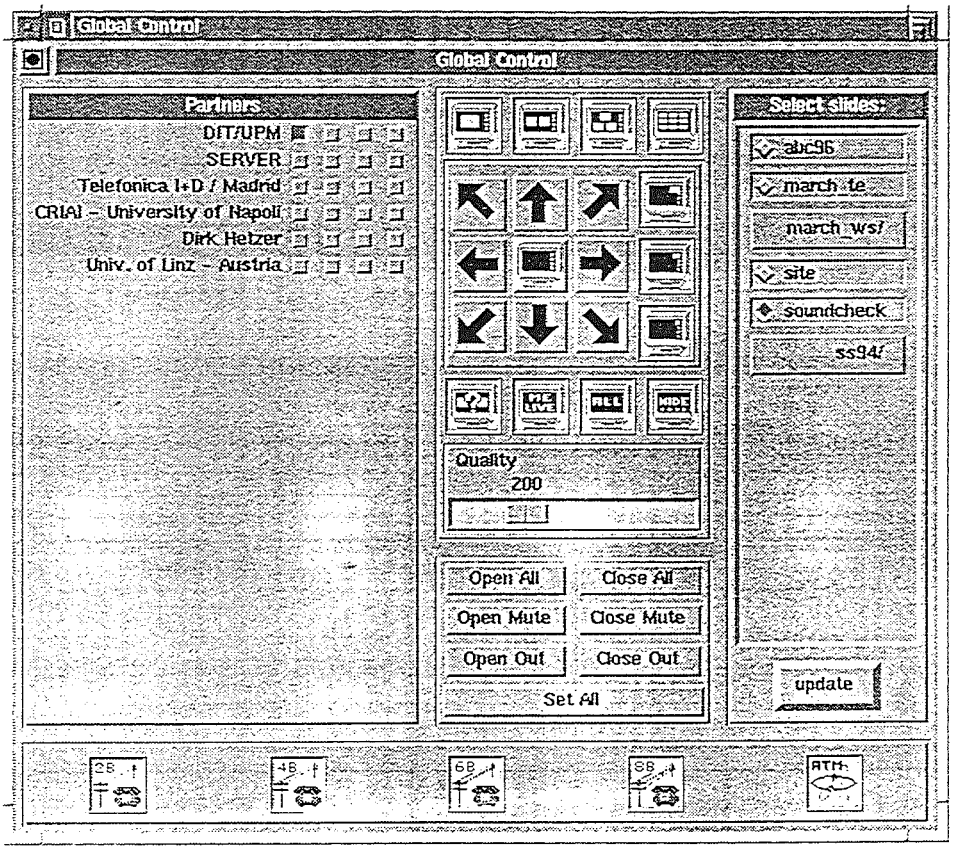
- La capacidad de participar en una discusión.

- La capacidad de hacer una demostración practica.

- .....

El componente de control permite controlar el modo de interacción, así como las capacidades asignadas a cada sede a través del panel de control que se representa en la figura siguiente. Este panel de control solo esta disponible en la estación de trabajo de control que se encuentra en el centro de control del evento. Desde él se puede controlar lo que ocurre en todas las sedes.

Cada botón selecciona un modo de interacción, un rol, una caracteristica o un elemento asociado a estos. Los detalles se pueden encontrar en el manual (ver : http://www.dit.upm.es/~proy/isabel).



Agradecimientos

La realización de una Escuela de Verano de esta magnitud no es posible sin la colaboración de una gran cantidad de personas e instituciones. Además hay que incluir a las personas e instituciones que participaron en la realización de las tres escuelas de verano anteriores. Queremos agradecer aquí a todos los participantes de los proyectos NICE, BRAIN, IBER e ISABEL de los programas ACTS y RACE de la Unión Europea. Todos ellos han contribuido al éxito de ABC´96. Deseamos mencionar, en particular, el apoyo de la compañia TELEFONICA de España en la realización de estos eventos.

BIBLIOGRAFÍA

"Synchronous Distributed Multimedia Services: Concepts, Case Studies and Challenges" by J. Quemada, S. Pavón, J. Salvachúa, D. Larrabeiti, T. Robles, PROMS : 3rd INTERNATIONAL WORKSHOP ON PROTOCOLS FOR MULTIMEDIA SYSTEMS October 15-18, 1996, Madrid, Spain

"ISABEL: A CSCW Application for the Distribution of Events", J. Quemada, T. de Miguel, A. Azcorra, S. Pavon, J. Salvachua, M. Petit, D. Larrabeiti, T. Robles, G. Huecas. BOOK: Multimedia Telecomunications and Applications, Springer Verlag, LNCS 1185, 1996, Editors: G. Ventre, J. Domingo-Pascual, A. Danthine. pp. 137-153, ISSN 0-7923-9529-8.

"Integracion de Componentes en la Aplicacion de Trabajo Cooperativo ISABEL", S. Pavón, T.P. de Miguel, M. Petit, J. Salvachua, J. Quemada, L. Rodriguez, P.L. Chas, C. Acuna, V. Lagarto, J. Bastos, Jornadas Telecom 94, Noviembre, 1994.

"ISABEL - Experimental Distributed Cooperative Work Application over Broadband Networks", T.P. de Miguel, S. Pavón, J. Salvachua, J. Quemada, P.L. Chas, J. Fernandez-Amigo, C. Acuna, L. Rodríguez, V. Lagarto, J. Bastos, pp 353--362, Springer-Verlag - *Lecture Notes in Computer Science*, Volume 868, September 1994.

"Distance Learning: Networks and Applications for RACE Summer School '94", A. Azcorra, T. Miguel, J. Quemada, S. Pavon, Department of Telematics from UPM, P. Chas, C. Acuña, P. Aranda, Telefonica I + D, V. Lagarto, J. Bastos, J. Domingues: Centro de Etudos de Telecomunicacoes, *The ATM Forum Newsletter* September, 1994 - Volume 2 Issue 3.

"Distribution of ABC´95 over the European ATM Pilot Network with the ISABEL Application", J. Quemada, T. Miguel, A. Azcorra, S. Pavon, J. Salvachua, M. Petit, J. I. Moreno, P. L. Chas, C. Acuna, L. Rodriguez, V. Lagarto, J. Bastos, J. Fontes, J. Domingues, Broadband Islands Conference, Dublin September 1995.

"Tele-education Experiences with the ISABEL Application", J. Quemada, T. Miguel, A. Azcorra, S. Pavon, J. Salvachua, M. Petit, J. I. Moreno, P. L. Chas, C. Acuna, L. Rodriguez, V. Lagarto, J. Bastos, J. Fontes, J. Domingues, High Performance Networking for Tele-teaching - IDC´95, Madeira November 1995.

Gestión

Uso de SDL para la especificación de sistemas de gestión OSI

M. Rodríguez Cayetano, E. Fernández López
DEPARTAMENTO DE TEORÍA DE LA SEÑAL Y COMUNICACIONES E INGENIERÍA TELEMÁTICA
ETSIT de Valladolid, UNIVERSIDAD DE VALLADOLID
REAL DE BURGOS S/N, 47011 VALLADOLID
Correo electrónico: manrod@tel.uva.es

Abstract:

An alternative method of a mapping from GDMO tamplates to SDL-92 is proposed in this work, which is based on the direct translation of each template into SDL constructions. That feature will allow simplify the proccess of automatic translation between both languages. The use of a natural language for the specification of GDMO behaviour templates is mapped into a SDL formal specification of the various parts of a managed object class. Types of characterizations of that behaviour and the results of its respective constructions in formal language are also exposed here.

1. Introducción

Una de las tareas más importantes y complejas en el desarrollo de aplicaciones de gestión basadas en el modelo OSI es la representación de los aspectos de gestión de los recursos de red. Con el fin de facilitar esta tarea y la estructuración de la información de gestión asociada, ITU-T ha desarrollado las recomendaciones *Directrices para la Definición de Objetos Gestionados* [3] (GDMO) y el *Modelo de Información de Gestión* (MIG) [1], ambas siguen un modelo orientado a objetos.

En el modelo de información de gestión se estructura la información de gestión en un conjunto de *objetos gestionados* que son abstracciones de los recursos de procesamiento y comunicación de datos. Cada objeto gestionado se define en términos de los atributos que posee, las operaciones que pueden realizarse sobre él, las notificaciones que pueden emitir y su relación con otros objetos gestionados. Los objetos que comparten una definición común se agrupan en *clases de objetos gestionados*.

GDMO introduce un conjunto de *plantillas* para la descripción (sintáctica) de clases y los elementos que las forman (atributos, operaciones, etc.). Los atributos en GDMO son representados mediante la notación ASN.1 [5], mientras que el comportamiento se puede expresar de manera informal mediante lenguaje natural o formalmente, si bien GDMO no impone ninguna estructura a la hora de expresar los comportamientos. De hecho, otras recomendaciones de ITU-T que incluyen plantillas GDMO (como la recomendación X.721, Definición de la Información de Gestión [2]) utiliza la descripción del comportamiento mediante lenguaje natural.

Esta descripción informal del comportamiento puede introducir ambigüedades que impidan conocer el comportamiento exacto del objeto gestionado, necesario para simplificar la especificación y el desarrollo de aplicaciones de gestión. Se manifiestan estas ambigüedades en la interacción de diferentes implementaciones de objetos gestionados con los agentes que acceden a ellos y los gestores que los controlan.

Las técnicas de descripción formal [6] permitirían aprovechar la definición del comportamiento de los objetos gestionados para la especificación de agente y gestor. El uso de estas técnicas está justificado por las ventajas que aportan, como son:

- permitir una especificación no ambigua y precisa del sistema
- proporcionar una base para el análisis de la especificación que permita comprobar su completitud, corrección y conformidad con los requisitos que debe cumplir el sistema
- proporcionar una base para analizar y simular distintas realizaciones del sistema antes de escoger la implementación definitiva del mismo. Esta característica permite reducir los costes de depuración del sistema.

El lenguaje de especificación formal SDL (recomendado por ITU-T) ha sido hasta ahora considerado como un buen candidato para ser utilizado como técnica de descripción formal para sistemas en tiempo real. En su versión de 1992 (SDL-92) [6] incluye extensiones orientadas a objetos, y ha sido posteriormente ampliado (recomendación Z.105 [7]) para permitir utilizar la notación ASN.1 en la definición de tipos de datos. Esto supone una adaptación del lenguaje para poder ser integrado completamente con GDMO.

En [8] se investiga el uso de SDL-92 para la especificación formal de objetos gestionados, cubriendo además la correspondencia del lenguaje con plantillas GDMO. En dicho artículo se muestran puntos cruciales en el establecimiento de la correspondencia, en cuanto a propiedades en los objetos MIG no soportadas por SDL-92: los atributos están solamente representados por tipos de atributos, la aceptación de todas las operaciones para cualquier estado de los objetos y la herencia múltiple. En [9] se alcanzan resultados de una integración completa entre ambos lenguajes, sustituyendo el lenguaje natural por descripciones formalizadas SDL.

Se propone en este trabajo un nuevo método de establecimiento de la correspondencia de plantillas GDMO a SDL-92, que servirá, a su vez, para la especificación estática y dinámica de los objetos gestionados. Esto permitirá obtener una especificación completa en SDL de las clases de objetos gestionados y de determinados procedimientos que podrán ser llamados por los procesos agentes. Este método partirá de la estructura de un objeto gestionado genérico dada en [9], y estará basado en la traducción directa de cada plantilla en construcciones SDL. Esta característica permite simplificar el proceso de traducción automática de GDMO a SDL-92.

La estructura seguida será la siguiente: en la sección 2 se describe el MIG y las plantillas GDMO, las extensiones orientadas a objetos en SDL-92 se describen en la sección 3, en la sección 4 se propone el método de traducción de las plantillas GDMO a construcciones SDL-92 y, por último, en la sección 5 se muestra un ejemplo del proceso de traducción.

2. Características del Modelo de Información de Gestión y GDMO

Dentro del entorno OSI, las aplicaciones de gestión realizan sus actividades mediante el procesamiento e intercambio de información de gestión, que se lleva a cabo mediante el uso de los servicios y protocolos de información de gestión común (CMIS y CMIP) [11], [12]. El Modelo de Información de Gestión (MIG) [1] estructura la información de gestión en objetos gestionados, que son definidos mediante las plantillas desarrolladas en la recomendación Directrices para la Definición de Objetos Gestionados o GDMO [3].

El MIG define atributos como propiedades de los objetos gestionados, a los que se asocian valores que se representan mediante tipos ASN.1 [5]. Estos atributos pueden estar agrupados mediante grupos de atributos, que son referidos como una sola entidad.

Se definen dos tipos de operaciones en los objetos gestionados: aquellas que se dirigen a un objeto gestionado para ser aplicadas a sus atributos, y aquellas que se aplican a los objetos gestionados como a un todo. La realización o no de una operación en un objeto gestionado está sujeta a restricciones que se expresan en la definición de la clase de objeto gestionado.

Los objetos gestionados pueden emitir notificaciones cuando se produce algún evento relevante para ellos. La definición de clase de objeto gestionado incluirá las notificaciones que puede emitir y las condiciones que determinan su envío.

Cada clase de objetos está caracterizada por un grupo de *lotes*, que son una colección de atributos, notificaciones, operaciones y/o comportamiento. Estos lotes pueden ser obligatorios (aparecen en cualquier instancia de la clase) o condicionales (su presencia en la instancia depende de que se verifique una condición asociada). Las condiciones asociadas a los lotes condicionales suelen estar relacionadas con capacidades del recurso representado.

Cada lote puede ser utilizado en la definición de una o más clases de objetos gestionados. Una vez que han sido incorporados a un objeto gestionado, los elementos que forman el lote se convierten en una parte integral del objeto y son accesibles solamente como una parte de ese objeto gestionado.

MIG establece una relación de contención en la que un objeto gestionado de una clase puede contener objetos gestionados de la misma o de diferentes clases. El objeto gestionado continente se denomina *objeto gestionado superior* de sus objetos gestionados subordinados. Las relaciones de contención entre objetos son independientes de las relaciones de herencia entre clases de objetos gestionados, y se representan dentro del denominado *árbol de contención*.

La relación de contención se utiliza para la denominación de instancias de objetos. Cada objeto (instancia) se nombra por la combinación del nombre de objeto superior y la información que identifica al objeto gestionado unívocamente dentro del ámbito de su objeto superior.

El Servicio de Información de Gestión Común (CMIS) [11] describe el mecanismo de comunicación que enlaza un gestor con un agente, que a su vez da acceso a un conjunto de objetos gestionados. Existen tres categorías de servicios: *servicio de asociación* (para el establecimiento de un asociación entre un agente y un gestor), *servicio de notificación de gestión* (para el envío de notificaciones del agente al gestor) *y servicios de operación de gestión* (para el envío de operaciones de gestión al agente por parte del gestor).

Se establecen también dentro de CMIS mecanismos para la selección de un conjunto de instancias de objetos sujetos a una operación de gestión. La *delimitación* permite seleccionar uno o varios objetos (instancias) dentro del árbol de nombrado a partir de un objeto gestionado base. Además, se permite la aplicación de *filtros* (comprobaciones sobre características de los atributos de un objeto) para la selección de un subconjunto de los objetos previamente seleccionados por el proceso de delimitación. En caso de que una operación de gestión vaya dirigida a más de un objeto gestionado, se puede especificar el tipo de sincronización: *atómica*, si la operación debe realizarse sobre todos los objetos o sobre ninguno (en el caso de que no sea

posible sobre alguno de ellos) o *la mejor posible* si la operación debe realizarse sobre aquellos objetos sobre los que sea posible.

La recomendación GDMO propone una notación basada en el uso de plantillas para la especificación de las características estáticas y dinámicas de los objetos gestionados. Los tipos de plantillas definidos son MANAGED OBJECT CLASS, PACKAGE, NAME BINDING, ATTRIBUTE, ATTRIBUTE GROUP, ACTION, NOTIFICATION y PARAMETER. En estas plantillas se especifican los siguientes aspectos:

a) identificación de la clase objeto, y descripción de la misma mediante la herencia múltiple de las características de un conjunto de superclases y/o la incorporación de lotes de características (estos no constituyen necesariamente por si solos objetos completos).

b) atributos o grupos de atributos, incluyendo los parámetros asociados, los valores por defecto que se aplican en la creación y cualquier restricción en la forma en que los valores de los atributos pueden ser modificados.

c) acciones, enumerando todas la posibles, los tipos de datos asociados a las solicitudes y/o confirmaciones de las mismas y los parámetros que pueden acompañarlas.

d) notificaciones, enumerando todas las posibles, los tipos de datos asociados al envío y/o confirmación de la misma y los parámetros que pueden utilizar.

e) normas de creación y borrado de objetos gestionados dentro del árbol de contención.

f) comportamiento, que indica cuándo las acciones o notificaciones son apropiadas y la forma en que se restringe su secuencia.

3. Modelado orientado a objetos en SDL-92

SDL (*Specification and Description Language*) es un lenguaje de especificación formal desarrollado y recomendado por ITU-T aplicable a todas aquellas áreas que requieran especificar el comportamiento de sistemas en tiempo real. Una especificación SDL define el comportamiento del sistema mediante un conjunto de estímulos y respuestas. El modelo del sistema está formado por un conjunto de procesos (máquinas de estados finitos extendidas) que se ejecutan concurrentemente e interactúan por medio del intercambio asíncrono de señales, y de forma síncrona mediante la llamada a procedimientos remotos. Cada proceso está caracterizado por un conjunto de variables que almacenan datos y un comportamiento, expresado en forma de estados y transiciones entre estados.

Las extensiones incluidas en SDL-92 permiten introducir mecanismos y conceptos orientados a objetos.

Una definición en SDL-92 es una definición de tipo (del que se pueden derivar instancias) o es una definición de instancia. SDL-92 permite la definición de tipos de sistema, de bloque, de proceso y de servicio.

El mecanismo de herencia se establece mediante especialización, agrupando conceptualmente los tipos en una jerarquía de tipos. La especialización permite a un tipo (el subtipo) estar basado en la propiedades de otro tipo (el supertipo). La especialización es soportada tanto en las estructuras de datos como en el comportamiento, permitiéndose además indicar en la especificación del comportamiento del (super)tipo dónde y cómo puede ser extendida. Se puede establecer la especialización, añadiendo propiedades al supertipo, o redefiniendo propiedades virtuales.

SDL-92 permite también la definición de tipos genéricos, que se obtienen a partir de tipos parametrizados para ser utilizados en diferentes contextos. Se definen propiedades para los parámetros de contexto (restricciones formales) cuando éstas sean requeridas por los tipos genéricos.

Las características del paradigma orientado a objetos se han incluido en SDL-92 con las siguientes particularidades:

- SDL-92 soporta encapsulación de varios niveles. Un sistema está formado por un conjunto de procesos que ofrece una interfaz de acceso basado en las señales que puede recibir y en los procedimientos que exporta y pueden ser invocados de forma remota. Además, los procesos pueden agruparse en bloques que esconden detalles internos de ellos a otros bloques. Cada bloque puede contener, a su vez, otros bloques.
- SDL-92 permite sólo herencia simple: un tipo (clase) sólo puede heredar las características de una única superclase y de las superclases de ésta dentro del árbol de herencia.
- SDL-92 soporta polimorfismo, que se establece mediante la petición de la ejecución de un procedimiento remoto en otro proceso, que tendrá resultado diferente dependiendo de la subclase a la que pertenezca el proceso en el que se va ejecutar el procedimiento.

Relacionado con el concepto de tipo se encuentra la construcción *lote*. Un lote SDL define una colección de tipos que pueden ser usados en diferentes especificaciones (p.e. por otros tipos). Puede ser utilizado cuando se definan instancias de un sistema o

cuando se definan otros lotes mediante la cláusula use.

Para simplificar la especificación en SDL de aplicaciones distribuidas en las que los tipos de datos intercambiados se definen mediante la notación ASN.1, se ha desarrollado la recomendación Z.105 [7] que es una ampliación de SDL para soportar la especificación directa de tipos de datos mediante ASN.1. Esta recomendación define además, una serie de operadores asociados a los tipos básicos de ASN.1. Esta ampliación de lenguaje SDL será utilizada como base en el método de traducción propuesto.

4. Especificación SDL del sistema gestionado a partir de plantillas GDMO

4.1 Traducción de las plantillas GDMO a construcciones SDL

El procedimiento seguido para obtener una especificación SDL de los objetos gestionados a partir de la información contenida en las plantillas GDMO que los definen, consiste en la traducción de los distintos elementos de dichas plantillas a construcciones SDL. Para almacenar toda la información de las plantillas se han definido estructuras de datos que contendrán las características (estáticas) de los atributos, grupos de atributos, acciones y notificaciones que forman un objeto gestionado. Puesto que cada uno de estos elementos puede ser incluido en distintas clases de objetos gestionados, se ha optado por definir un lote SDL por cada plantilla GDMO. Estos lotes SDL serán incorporados (*use*) en la especificación de los objetos gestionados.

Puesto que en un lote SDL no se pueden especificar directamente variables, aunque sí tipos de datos, tipos de procesos, tipos de servicios, tipos de bloques y procedimientos, definiremos uno o más procedimientos que almacenarán las características estáticas del elemento del objeto gestionado considerado en un tipo de datos de la especificación SDL.

Este método de traducción simplifica la generación automática de especificaciones SDL de los objetos gestionados, puesto que la obtención de características estáticas de los elementos que lo forman se reducirá al uso de los lotes SDL apropiados y la ejecución de uno o más procedimientos definidos en dichos lotes. En concreto, cada objeto gestionado tendrá que incorporar los lotes SDL correspondientes a los lotes GDMO obligatorios y condicionales que lo forman, y los lotes SDL que definen las características de las superclases de las que se deriva (estos lotes podrán, a su vez, incorporar otros lotes).

Entre las características estáticas de los objetos gestionados se encuentran la definición de los tipos de datos ASN.1 de los atributos, parámetros e información asociada a las acciones y notificaciones. Para no tener que definir una variable de un tipo distinto por cada tipo de datos ASN.1 que utilice cada uno de estos elementos, se utilizan tipos de datos CHOICE. Así, la especificación de la sintaxis ASN.1 utilizada por un determinado atributo, parámetro, etc., supondrá añadir un campo al tipo CHOICE que se utiliza para almacenar los valores de ese tipo de elemento.

Las características dinámicas de los elementos que constituyen el objeto gestionado, expresadas en la sección BEHAVIOUR de las distintas plantillas de forma textual, deberán ser sustituidas por uno o más procedimientos SDL. Considerando las descripciones de características de los comportamientos dadas en el modelo de información [1], GDMO [3], y el draft de la enmienda 4 a la recomendación GDMO (GDMO+, *Specifying The Behaviour of Managed Objects*) [4], podemos distinguir los siguientes tipos:

- **precondiciones** (o condiciones previas): son condiciones que deben verificarse inmediatamente antes de la realización de una determinada tarea, como puede ser la emisión de una notificación, la realización de una acción, el envío de un parámetro asociado a un atributo, acción o notificación en una PDU (Unidad de Datos del Protocolo) del protocolo de gestión o un cambio en el comportamiento del objeto gestionado. Para facilitar el proceso de traducción, estos procedimientos podrían tener un nombre de la forma <tipo_plantilla>Preconditions<nombre>, donde <tipo_plantilla> es el tipo de plantilla de donde hemos extraído el comportamiento (*action, notification, attribute*, etc.) y <nombre> es el nombre de la plantilla GDMO. Estos procedimientos devolverán un valor Booleano *True* si se cumple la precondición o *False* en caso contrario.
- **Invariantes**: son condiciones que deben permanecer ciertas durante un determinado plazo de tiempo, como puede ser todo el tiempo de existencia de un atributo o de un objeto gestionado, la duración de la realización de una acción, etc. Estos procedimientos tendrán un nombre de la forma <tipo_plantilla>Invariants<nombre> y devolverán un valor Booleano *True* si la condición invariante se mantiene.
- **Postcondiciones** (o condiciones posteriores): son condiciones que deben ser ciertas inmediatamente después de la realización de una determinada tarea, como puede ser el envío de un parámetro en una PDU de gestión, la realización de una acción, o el envío de una notificación. El nombre de estos procedimientos será de la forma <tipo_plantilla>Postconditions<nombre>, y deberá contener las operaciones necesarias para asegurar el cumplimiento de la postcondición.

- **Eventos** que pueden desencadenar la realización de una tarea (si se cumplen las precondiciones e invariantes, caso de que existan), como el envío de una notificación o cambios en el comportamiento del objeto gestionado. Este tipo de procedimientos tendrán por nombre <tipo_plantilla>Events<nombre>, y devolverán un valor Booleano cierto si se ha producido alguno de los eventos que puede desencadenar la realización de una tarea determinada.
- **Comportamiento de las acciones**: son la secuencia de operaciones que se realizan al invocar una acción definida en un objeto gestionado. Este tipo de comportamiento dará lugar a procedimientos del tipo action<action_label>, donde <action_label> es el nombre de la plantilla acción.
- **dependencias entre valores de atributos**: se representarán mediante procedimientos de nombres relationship<nombre>, con <nombre> el nombre de la relación. Estos procedimientos realizarán las acciones necesarias para el cumplimiento de la relación, y podrán ejecutarse tras la modificación del valor de un atributo o cuando el objeto gestionado no esté procesando ninguna solicitud de operación.
- **Reglas de creación y borrado del objeto**: indicarán aquellas precondiciones, invariantes y postcondiciones relacionadas con la creación y las precondiciones para el borrado del objeto, así como los eventos que pueden determinan la creación o borrado del mismo (en caso de existir) además de las operaciones de creación y borrado del protocolo de gestión. Las reglas de creación y borrado de objetos están determinadas por la ligazón de nombres que estemos utilizando para la denominación de los objetos de una clase. Estas reglas se traducirán a procedimientos SDL con nombres de la forma *<nombre_ligazón>CreatePreconditions<nombre_clase>*, *<nombre_ligazón>CreateInvariants<nombre_clase>*, *<nombre_ligazón>CreatePostconditions<nombre_clase>* y *<nombre_ligazón>DeletePreconditions<nombre_clase>*.
- **Reglas de concordancia de atributos**: indican cómo se aplican las distintas reglas de concordancia (*Matching Rules*) relacionadas con el proceso de filtrado a un atributo concreto. Estas reglas se traducirán a procedimientos SDL de nombre <tipo_regla><nombre_atributo>, donde <tipo_regla> es el tipo de regla de concordancia que se desea aplicar (*equality, ordering, substrings*, etc.) y <nombre_atributo> el nombre de la plantilla atributo de donde se ha extraído el comportamiento. Estos procedimientos tendrán como parámetro de llamada el valor con el cual se quiere hacer la comprobación, y devolverán un valor Booleano que indicará si el resultado de la comprobación es cierto o falso.

4.2 Especificación SDL del objeto gestionado

Con la traducción desarrollada en el párrafo anterior obtenemos una serie de procedimientos que contienen las características estáticas y dinámicas de los distintos elementos que constituyen un objeto gestionado. En esta sección se presenta la especificación del objeto gestionado en sí, que está basado en la traducción de cada clase de objeto gestionado a un tipo de proceso SDL. El cuerpo de este proceso consistirá en una serie de transiciones y estados deducidos del comportamiento de un objeto gestionado genérico, que responderá a operaciones de gestión y emitirá notificaciones, y del comportamiento característico del objeto, deducido de las plantillas GDMO que definen los elementos que lo constituyen (expresado mediante los procedimientos definidos en la sección anterior).

La característica de herencia múltiple que pueden presentar las clases de objetos gestionados no se traduce a herencia de tipos de procesos en SDL, sino que se utilizan procedimientos que combinan las características de las distintas clases heredadas (expresadas en las correspondientes estructuras de datos SDL) y las almacenan en estructuras de la subclase.

En el tipo de proceso SDL, los atributos, grupos de atributos, acciones y notificaciones se representan mediante matrices SDL que almacenan sus características. En concreto, los atributos se representan en una matriz con tantas posiciones como atributos posea el objeto. Cada elemento de la matriz contendrá los siguientes campos:

- identificador de objeto del atributo (nombre único globalmente)
- valor del atributo
- tipos de reglas de concordancia para filtrado que soporta
- lista de parámetros asociados
- posibilidad de aceptar operación *replace-with-default*
- especificación de valor inicial
- especificación de valor por defecto
- tipos de operaciones de gestión que soporta (*get, replace, add, remove*)

Los grupos de atributos se representarán en una matriz cuyos elementos tendrán los siguientes campos:

- identificador de objeto del grupo de atributos
- lista de atributos que lo forman

La matriz que representa las características de las acciones tendrá elementos con los siguientes campos:

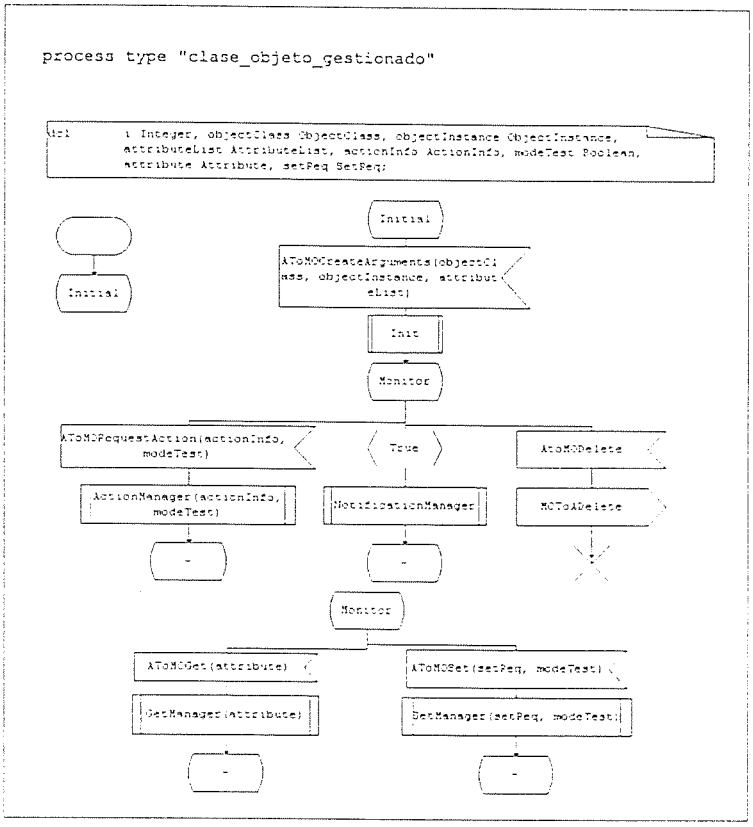


Fig. 1

- Identificador de objeto de la acción
- modo de confirmación (acción confirmada o no confirmada)
- lista de parámetros asociados

Por último, los elementos de la matriz de notificaciones contendrán los siguientes campos:

- Identificador de objeto de la notificación
- valor de los datos asociados a la notificación (si existen)
- lista de parámetros asociados

Un esquema general de la estructura de un objeto gestionado genérico (tipo de proceso SDL) se muestra en la Fig. 1. En esta figura se observan los distintos tipos de señales que puede recibir el objeto gestionado, y que se corresponden con los distintos tipos de operaciones de gestión que puede recibir del agente. También se pueden observar llamadas a procedimientos que se encargarán de las tareas relacionadas con las operaciones de gestión obtener valor de atributo, fijar valor de atributo, acción y notificación.

Como ejemplo de estos procedimientos, consideremos el caso del procedimiento *ActionManager*, encargado de procesar una solicitud de acción. La llamada a este procedimiento incluirá dos parámetros, la información relacionada con la acción (identificador de objeto de la acción y posible valor asociado), y un indicador de si la acción debe ejecutarse realmente o si sólo se desea comprobar si es factible su ejecución. Este último caso se producirá cuando el agente reciba una solicitud de operación acción que implique a varios objetos (seleccionados mediante reglas de delimitación y filtrado) que requiera sincronización atómica (la operación debe poder realizarse sobre todos los objetos gestionados; si no, no se realizará sobre ninguno). El procedimiento, en función del identificador de acción, comprobará las precondiciones e invariantes de la acción (ejecutando el procedimiento *actionPreconditions<nombre_acción>* y si el valor devuelto es cierto, el procedimiento *actionInvariants<nombre_acción>*). Si alguna de ellas no es cierta, se enviará una señal al agente indicando la imposibilidad de realizar la acción. Si son ciertas y sólo se requiere comprobar la posibilidad de ejecución de la acción, se enviará una señal al agente indicando tal circunstancia. Por último, si realmente se ha solicitado ejecutar la acción (y las precondiciones e invariantes se cumplen), se realizará una llamada al procedimiento *action<nombre_acción>* seguida de una llamada al procedimiento *actionPostconditions<nombre_accion>* y el envío al agente de los parámetros de respuesta de la acción, en caso de existir.

4.3 Especificación del agente

En esta sección se describen las funciones que debe desempeñar un agente OSI. Este agente se traducirá en un proceso SDL cuyo comportamiento se

expresará mediante una serie de estados y transiciones deducidos del comportamiento de un agente genérico y de las características de los objetos gestionados que pertenecen a su MIB.

Las tareas que debe desempeñar un agente OSI genérico independientemente de la MIB que gestione son:

- **Delimitación.** En el caso de que la solicitud de operación de gestión recibida por el agente incluya el campo de delimitación, será misión del agente seleccionar el conjunto de objetos dentro del árbol de nombrado según el tipo de delimitación y objeto base especificados. Esta tarea podría traducirse en el procedimiento SDL *scoping* que recibiese como parámetros el objeto base (clase e instancia) y el tipo de delimitación, y devolviese como resultado una secuencia de objetos gestionados (identificados por su clase e instancia).
- **Filtrado.** Si la operación de gestión solicitada incluye el campo de filtrado, el agente deberá encargarse de aplicar el filtro especificado a los objetos previamente seleccionados por el procedimiento de delimitación. Un filtro consiste en una combinación lógica de afirmaciones sobre los valores de los atributos y/o la presencia de los mismos en un objeto gestionado. Para comprobar si cada una de estas afirmaciones es cierta o no, se utilizarán los procedimientos que indican cómo se aplican las reglas de concordancia genéricas (*equality*, *ordering*, etc.) a cada atributo. (procedimientos *equality<nombre_atributo>*, *ordering<nombre_atributo>*, etc., definidos en la sección 4.1). Aquellos objetos para los que el filtro se verifica serán seleccionados para recibir la operación de gestión.
- **Sincronización.** En el caso de que se especifique sincronización atómica en una operación sobre múltiples objetos, el agente deberá comprobar la verificación de las restricciones asociadas a la operación enviando la petición a cada objeto en modo comprobación (parámetro *modeTest* con valor *True* en la señal enviada al objeto). Si la operación es factible en todos los objetos, se enviará la solicitud de operación a cada uno de ellos; si no es factible en alguno de ellos, no se realizará la operación sobre ninguno. Si la sincronización es del tipo *la mejor posible*, se enviarán las solicitudes de operación directamente a los objetos (algunos de ellos pueden rechazarla).
- **Procesamiento de operaciones destinadas a múltiples objetos.** En el caso de que la operación solicitada afecte a varios objetos gestionados (seleccionados mediante delimitación y/o filtrado), el agente deberá traducirla a una serie de operaciones dirigidas a cada uno de los objetos.
- **Gestión de grupos de atributos.** El agente deberá traducir cada operación dirigida a un grupo de atributos de un objeto a una serie de operaciones sobre cada uno de los atributos que forman el grupo.
- **Discriminación de eventos.** Cada gestor que dialoga con un agente debe indicarle qué tipos de notificación y de qué objetos está interesado en recibir. Esta información quedará registrada en el agente. De esta forma, al recibir una notificación, el agente sabrá a qué gestores debe enviarla, ya que la notificación enviada por los objetos gestionados no incluye ninguna información sobre los gestores destino de la misma.
- **Soporte de comunicación con los objetos gestionados.** El agente deberá utilizar la información aportada en una petición de operación de gestión para la identificación del objeto u objetos destino de la misma (identificador de clase e instancia) para poder identificar el elemento o elementos del sistema que los representan. En el modelo propuesto, el agente deberá traducir cada identificador de clase e instancia al identificador del proceso SDL (*PId*) que modela el objeto gestionado.

Una función que también debería desempeñar el agente es la gestión de clases alomórficas. Dado que el protocolo de gestión CMIP actualmente no soporta alomorfismo, no ha sido considerada en el presente trabajo.

Existen distintos aspectos del comportamiento que aparecen expresados en las plantillas GDMO y que no son en realidad características dinámicas de los objetos gestionados, sino del proceso agente. Entre esos aspectos, tenemos:

- **Restricciones para la creación de una instancia de objeto gestionado.** Puesto que el objeto todavía no existe, es misión del agente verificar si se cumplen las restricciones para la creación de dicho objeto. Estas restricciones están asociadas a la ligazón de nombres utilizada para la denominación de los ejemplares de la clase, por lo que dentro de una misma clase podrán existir tantas restricciones distintas como ligazones de nombres utilizados para denominación. Como ya se ha visto en la sección 4.1, estas restricciones se traducirán en los procedimientos SDL *<nombre_ligazón>CreatePreconditions<nombre_clase>*, *<nombre_ligazón>CreateInvariants<nombre_clase>* y *<nombre_ligazón>CreatePostconditions<nombre_clase>*. El agente deberá ejecutar los procedimientos *<nombre_ligazón>CreatePreconditions<nombre_clase>* y *<nombre_ligazón>CreateInvariants<nombre_clase>* y comprobar, por los valores que devuelven, si se cumplen las restricciones para la creación antes de crear realmente el objeto. El procedimiento *<nombre_ligazón>CreatePostconditions<nombre_clase>* deberá ser ejecutado por el objeto para asegurar que las postcondiciones se cumplen tras la creación. El agente también deberá encargarse de obtener los valores iniciales de los distintos atributos en el caso de que se especifique un objeto gestionado de valor inicial (*IVMO*) y enviárselos al objeto recién creado.

- **Restricciones para el borrado de objetos.** Antes de que pueda borrarse un objeto es necesario comprobar que se verifican las restricciones dadas en la ligazón de nombres utilizada para la denominación de los ejemplares de la clase. Estas restricciones se especificarán en el procedimiento *<nombre_ligazón>DeletePreconditions<nombre_clase>*.

Una vez vistas las distintas funciones que debe desempeñar un agente, comentaremos de manera informal el comportamiento que debe presentar. El proceso SDL que representa al agente recibirá peticiones de gestión provenientes de un gestor mediante el intercambio de señales. Al recibir la petición, el agente utilizará la información sobre el objeto base de la misma, el tipo de delimitación y el filtrado que se desea aplicar para obtener la lista de objetos a los que va dirigida la operación. Una vez obtenida, el agente enviará una señal, indicando el tipo de operación solicitada, a cada uno de los objetos de la lista, y esperará la respuesta de cada uno de ellos. La respuesta obtenida de cada objeto podrá ser utilizada para la confirmación de la operación al gestor en caso de que la operación se haya realizado en modo confirmado.

En el caso de que el agente reciba una notificación por parte de alguno de los objetos de su MIB, comprobará mediante el discriminador de eventos si existe algún gestor, con el que tenga establecida una asociación, que esté interesado en recibir una notificación de ese tipo y proveniente de

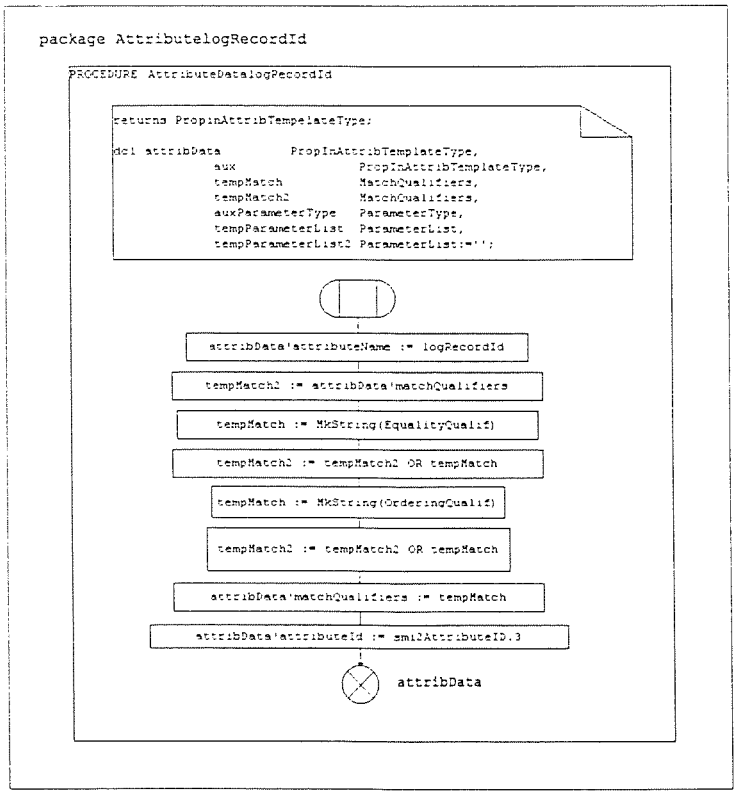


Fig. 2

ese objeto. Si es así, enviará la notificación a los gestores seleccionados.

5. Ejemplos del procedimiento de traducción

Como ejemplo del sistema de traducción de plantillas GDMO a SDL se considera el objeto *logRecord* definido en la recomendación X.721 de ITU-T [2]. La definición de dicho objeto y los elementos relacionados es la siguiente:

```
logRecord    MANAGED OBJECT CLASS
 DERIVED FROM    top;
 CHARACTERIZED BY
logRecordPackage    PACKAGE
   BEHAVIOUR
   logRecordBehaviour    BEHAVIOUR
    DEFINED AS "This managed object represents the
information stored in the logs";;
   ATTRIBUTES
   logRecordId    GET,
   loggingTime    GET;;;

REGISTERED AS        {joint-iso-ccitt ms(9) smi(3)
part2(2) managedObjectClass(3) 7}; -- changed by
Technical Corrigendum 2

logRecordId    ATTRIBUTE
        WITH    ATTRIBUTE    SYNTAX    Attribute-
ASN1Module.LogRecordId;
   MATCHES FOR EQUALITY, ORDERING ;

REGISTERED AS        {joint-iso-ccitt ms(9) smi(3)
part2(2) attribute(7) 3}; -- changed by Technical
Corrigendum 2

loggingTime    ATTRIBUTE
        WITH    ATTRIBUTE    SYNTAX    Attribute-
ASN1Module.LoggingTime;
   MATCHES FOR EQUALITY, ORDERING;
```

Cada plantilla de atributo dará lugar a un lote SDL de nombre *Attribute"attrib_label"* donde "attrib_label" es el nombre de la plantilla atributo. Dentro de este lote se definirá un procedimiento SDL que almacenará toda la información relacionada con el atributo en una estructura de datos SDL. El procedimiento correspondiente al atributo *logRecordId* se muestra en la Fig. 2. El campo de la plantilla GDMO que expresa el tipo ASN.1 del atributo se traducirá en una nueva entrada de la forma

attributeType"attrib_label" Type"attrib_label"

en el tipo ASN.1 CHOICE que contiene los posibles tipos de datos de los atributos (tipo *AttributeValue*) seguida de la definición:

Type"attrib_label" ::= "type_reference"

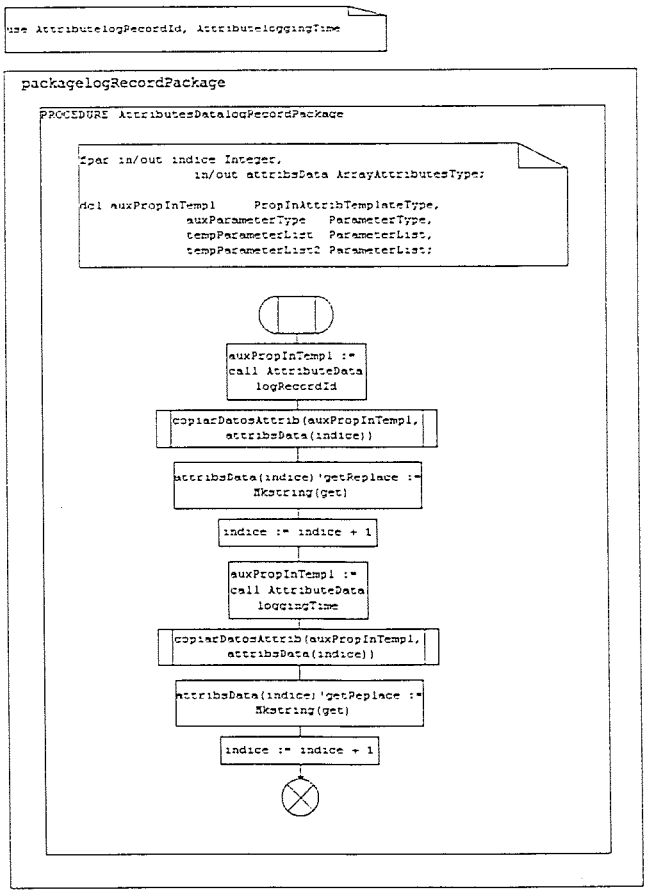


Fig. 3

donde type_reference es el tipo de datos ASN.1 del atributo. De esta forma, el tipo de datos *AttributeValue* y los tipos auxiliares tomarían la forma:

```
AttributeValue ::= CHOICE{
  attributeTypelogRecordId  TypelogRecordId,
  attributeTypelogginTime   TypeloggingTime}

TypelogRecordId ::= LogRecordId
TypeloggingType ::= LoggingTime
```

Los tipos de datos ASN.1 de los atributos se podrían haber incorporado directamente en el tipo *AttributeValue*, pero el método seguido simplifica la definición de nuevos atributos derivados de otros ya existentes.
Sería necesario incorporar el módulo *Attribute-ASN1Module* en nuestras definiciones de datos ASN.1.

El lote *logRecordPackage* se traducirá en un lote SDL con un nombre de la forma *package"package_label"* (lote *packagelogRecordPackage*) que contendrá un procedimiento que completará las definiciones de las propiedades de los atributos (procedimiento *AttributesDatalogRecordIdPackage*). Este procedimiento a su vez llamará a los procedimientos *AttributeDatalogRecordId* y *AttributeDataloggingTime* definidos en los lotes *AttributelogRecordId* y *AttributeDataloggingTime* respectivamente (estos lotes deberán ser usados desde el lote *packagelogRecordPackage*). La definición de este último lote se muestra en la Fig. 3.

Por último, consideraremos la plantilla de objeto gestionado, que se traducirá en un lote SDL de nombre *objectClass"class_label"* (lote *objectClasslogRecord*). Este lote usará los lotes donde se encuentren las definiciones de las características de las superclases de la clase dada, y los lotes obligatorios y condicionales que la forman. En este lote se definirá el procedimiento *objectClassDatalogRecord* que se encargará de combinar todas las definiciones de las superclases de la clase dada, lotes obligatorios y lotes condicionales en una única estructura de datos. La definición de este lote se muestra en la Fig. 4..

6. Conclusiones

El procedimiento que se acaba de describir para el establecimiento de la correspondencia entre ambos lenguajes pretende obtener una traducción directa de cada una de las plantillas GDMO a construcciones SDL-92, lo que significa una simplificación en el método de traducción y la posibilidad de realizarlo de forma automática. La estructuración del comportamiento presentado permite facilitar el diseño completo del sistema gestionado (objetos gestionados y agente).

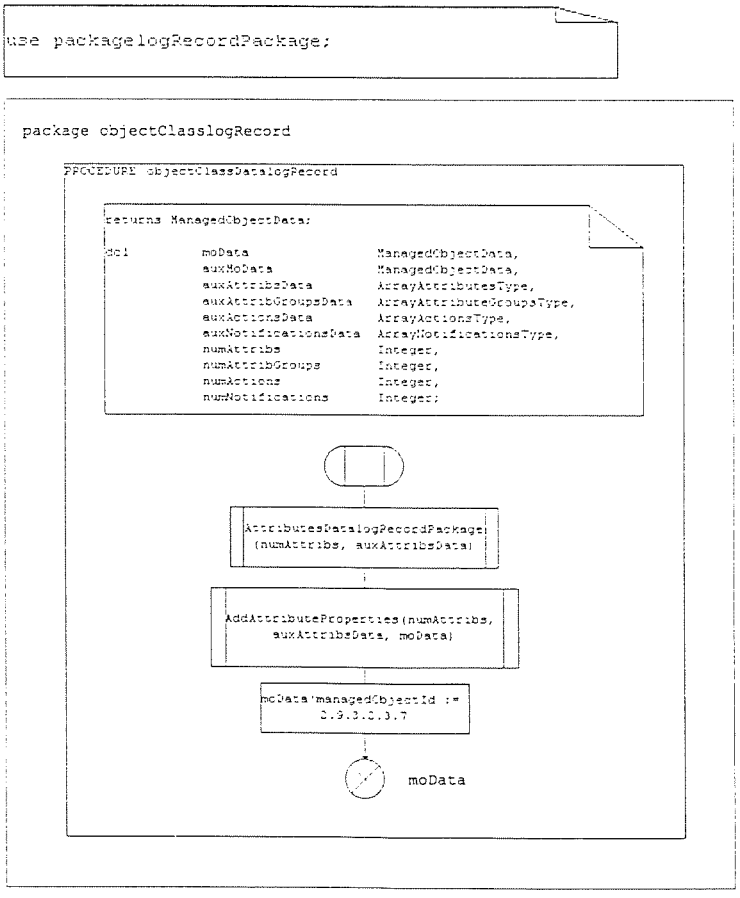


Fig. 4

Referencias

[1] "Tecnología de la Información - Estructura de la Información de Gestión - parte 1: Modelo de Información de Gestión" *Recomendación X.720 de CCITT / 10165-1 de ISO/IEC*, (1992).

[2] "Tecnología de la Información - Estructura de la Información de Gestión - parte 2: Definición de la Información de Gestión" *Recomendación X.721 de CCITT / 10165-2 de ISO/IEC*, (1992).

[3] "Tecnología de la Información - Estructura de la Información de Gestión - parte 4: Directrices para la Definición de Objetos Gestionados" *Recomendación X.722 de CCITT / 10165-4 de ISO/IEC*, (1992).

[4] "Information Technology - Open Systems Interconnection - Structure of Management Information: Guidelines for the Definition of Managed Objects - Amendment 4: GDMO+ Specifying the Behaviour of Managed Objects" *Enmienda 4 a la Recomendación X.722 de ITU-T*, (1997)

[5] "Especificación de una Notación de Sintaxis Abstracta Uno (ASN.1)" *Recomendación X.208 de CCITT*, (1992).

[6] "Lenguaje de Especificación y Descripción de CCITT". *Recomendación Z.100 de ITU-TSS*, (1992).

[7] "Lenguaje de Especificación y Descripción Combinado con Notación de Sintaxis Abstracta Uno". *Recomendación Z.105 de ITU-TSS*, (1995).

[8] Mazaher, S., y Moller-Pedersen, B. "On the Use of SDL-92 for the Spetification of Behaviuor in OSI Network". *Proceedings of the 1993 SDL Forum*, (1993).

[9] Bartocci, A., y Ferrero, A. "Integrated Use of SDL and GDMO". *Proceedings of the 1995 SDL Forum*, 279-290 (1995).

[10] Bartocci, A. y Ferrero, A. "SDL and GDMO Integration". *CSELT Technical Report* (1995).

[11] "Definición del Servicio de Información de Gestión Común para Aplicaciones CCITT" *Recomendación X.710 de CCITT*, (1991).

[12] "Especificación del Protocolo de Información de Gestión Común" *Recomendación X.711 de CCITT*, (1991).

Gestión Avanzada de una Red Corporativa de Telecomunicaciones

KOEHN, Rogério[1] - SER, Luis del[2] - SANCHEZ, David[3] - GUERRERO, Carmen[4] - CARNEIRO, Victor[5]
DEPARTAMENTO DE ELECTRONICA Y SISTEMAS
UNIVERSIDAD DE LA CORUÑA

Abstract :

Nowadays, an advanced telecommunications network management implies the use of standard solutions. The set of standards developed by ITU-T and known as TMN (Telecommunications Management Network), establishes a framework for building management systems that can cope with current situation of telecommunication networks: rapid technology evolution, coexistence of equipments from different vendors, demand of new services.

This paper makes a broad description of these standards and presents a pioneer implementation in a power utility telecommunications network: from the software platforms that make possible to develop such a system to the architecture of agents and new management applications that improve current network management processes.

1. Introducción

El crecimiento que han experimentado las redes de telecomunicaciones, la diversificación de tecnologías, las circunstancias de un entorno competitivo y liberalizado y el aumento de la complejidad de gestionar y proveer nuevos servicios en una red de telecomunicaciones conducen en la actualidad al desarrollo de sistemas de gestión estándares para las mismas.

La necesidad de gestionar algo tan complejo como una red de telecomunicaciones moderna lleva consigo cumplir los requerimientos de control y mantenimiento de la funcionalidad de las redes y proporcionar los medios para las facilidades y servicios de aprovisionamiento rápidos para el cliente final.

La definición de gestión de red es compleja y frecuentemente significa diferentes cosas para diferente gente. Por este motivo el ITU (*International Telecommunications Union*) define la red de gestión de telecomunicación (TMN - *Telecommunication Management Network*) en el conjunto de recomendaciones M.3000 que describen el marco para transporte y control de información de gestión para obtener los requerimientos operacionales de mayor calidad de la red y disponer una única visión de gestión de red de telecomunicaciones .

El nivel y la complejidad involucrada en la definición e implementación de TMN obliga a una tarea continua. A pesar de que los principales progresos ya se han realizado, parte del software de gestión de red y los interfaces asociados a redes de telecomunicaciones hoy en día contienen elementos propietarios. Y por lo tanto en un futuro próximo las dificultades de interfuncionamiento entre diferentes vendedores continuarán [1].

Por su parte, algunos suministradores han reconocido estas dificultades y se han marcado el objetivo de un interfaz independiente desde el principio. La plataforma de software orientada a objetos resultante es suficientemente flexible para adaptar las normas TMN actuales. Al mismo tiempo proporciona una arquitectura que se puede adaptar a la gestión de otros fabricantes de equipos PDH y SDH.

El SGRT - Sistema de Gestión de la Red de Telecomunicaciones - de Unión Fenosa es una implementación practica de un sistema de gestión que cumple la arquitectura TMN para su red de telecomunicaciones formada por elementos PDH (*Plesiochronous Digital Hierarchy*) y SDH (*Synchronous Digital Hierarchy*). Debido al componente de investigación de este proyecto se desbribe en esta publicación aspectos de su arquitectura, como caso practico de un sistema de gestión TMN.

2. El concepto TMN (M 3010 del ITU).

El concepto básico detrás de TMN es proporcionar una estructura de red organizada para lograr la interconexión entre los diferentes sistemas de operación y equipos de telecomunicación, utilizando una arquitectura conveniente y una serie de interfaces normalizados [2].

Dentro de la arquitectura TMN general existen tres aspectos básicos de ésta que pueden ser considerados por separado al planificar y diseñar un sistema de gestión basado en TMN. Estos tres aspectos son los siguientes:

[1] Actualmente en el Grupo Unión Fenosa - email : ib001059 *a* uef.es
[2] Actualmente en el Grupo Unión Fenosa - email : ns000404 *a* uef.es
[3] Actualmente en el Departamento de Electrónica y Sistemas de la Universidad Alfonso X, El Sabio - email : dsanchez *a* uax.es
[4] email : clopez *a* des.fi.udc.es
[5] email : victor *a* des.fi.udc.es

- arquitectura funcional TMN;
- arquitectura de información TMN;
- arquitectura física TMN.

2.1. Arquitectura Funcional TMN

La arquitectura funcional podría considerarse como los bloques funcionales que permiten construir sistemas complejos. La definición de bloques funcionales y puntos de referencia entre bloques da origen a los requisitos aplicables a las especificaciones de interfaz recomendadas para el TMN. Los bloques funcionales se comunican a través de la función de comunicaciones de datos (DCF).

Los bloques funcionales del TMN son :

- Bloque Funcional de Sistema de Operaciones (OSF) que procesa informaciones relacionadas con la gestión de las redes de telecomunicaciones con el objetivo de monitorizar, cordinar y controlar las funciones de los elementos de red.
- Bloque Funcional de Elemento de Red (NEF) que se comunica con TMN para ser monitorizado y o controlado.
- Bloque Funcional de Estación de Trabajo (WSF) que se responsabiliza de la interpretación de toda la información TMN para los usuarios, incluyendo un interfaz hombre-maquina.
- Bloque Funcional Adaptador Q (QAF) que tiene como función la conexión de las entidades no TMN a la red TMN.
- Bloque Funcional de Mediación (MF) es responsable por la compatibilidad de la información intercambiada por los bloques funcionales OSF y NEF o OSF y QAF.

Los bloques funcionales intercambian información a través de los puntos de referencias q, f, x, como ilustra la Fig. 1.

2.2. Arquitectura de Información TMN

La arquitectura de información, basada en un planteamiento orientado el objeto, proporciona el fundamento de aplicación de los principios de gestión de sistemas de interconexión de sistemas abiertos (OSI, *open systems interconnection*) a los principios del TMN [3,4].

Además de esto TMN añade algunos conceptos de modo que permita que el modelo de información atienda a otros requerimientos. Uno de estos conceptos es la Arquitectura Lógica en Niveles (LLA - *logical layered architecture*) que consiste en definir la arquitectura de gestión como una serie de niveles. La arquitectura lógica en niveles usa un planteamiento de la descomposición de una actividad de gestión en una serie de dominios funcionales. Cada dominio funcional es mapeado en un dominio de gestión bajo el control de un Función de Sistema de Operación (OSF).

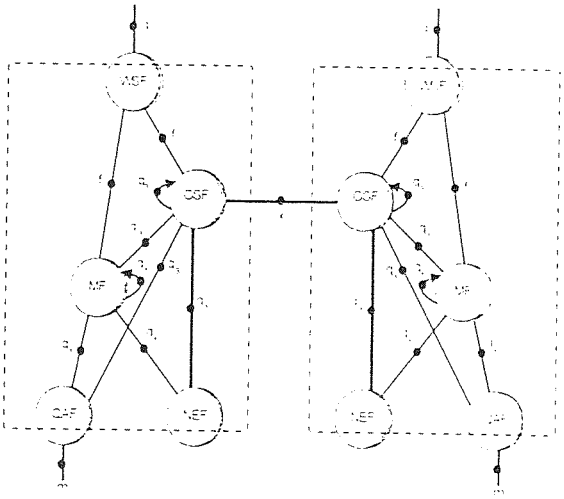


Fig. 1 - Diagrama de los Bloque Funcionales

La información intercambiada por los sistemas de gestión se define usando una técnica de especificación formal basada en el paradigma orientado a objeto. Se describe mediante un lenguaje semiformal llamado GDMO (*Guidelines for Definition of Managed Objects*).

Esto posibilita una homogeneidad de interfaz de gestión, que permite a una sola aplicación de gestión, el gestionar una amplia gama de equipos de forma similar. Usar un método de especificación de interfaz orientada a objetos es, por supuesto, independiente de si la implantación del elemento de red usa técnicas de diseño orientado a objeto ó no.

GDMO suministra un mecanismo normalizado para definir la sintaxis, semántica y aspectos de comportamiento de la información de gestión de una manera formal permitiendo así un común entendimiento de las capacidades de gestión limitando el espacio de interpretación. GDMO suministra plantillas de especificación formal para los objetos gestionados, que se usan para definir su derivación y todas las propiedades que caracterizan al objeto gestionado [5]. En un modelo de información se observa además que :

- no existe necesariamente una correspondencia "uno-para-uno" entre los objetos gestionados y los recursos reales, o elementos de red ;
- un elemento de red puede ser representado por uno o mas objetos gestionados. Cuando un recurso es representado por varios objetos gestionados, cada uno de ellos representa una visión distinta del recurso ;
- pueden existir objetos gestionados representando recursos lógicos del TMN ;

- si un elemento no es representado por un objeto gestionado, es invisible a los sistemas de gestión TMN ;
- un objeto gestionado puede proveer una visión abstracta de recursos que son representados por otros objetos gestionados ;

Un objeto gestionado esta definido por [5,6] :

- Atributo : representa la información que contiene un objeto gestionado. La lista de propiedades de un atributo define si un mensaje de un gestor puede leer (GET) ó escribir (SET) el valor del atributo.
- Notificación : es un mecanismo por el que un objeto gestionado informa autónomamente de los eventos relacionados con el recurso gestionado.
- Acción : es un mecanismo por el que un gestor pide a un objeto gestionado que realice una función de gestión dada.
- Comportamiento : abarca la definición del aspecto dinámico del objeto : bajo qué condición se envía una notificación, como reacciona, por ejemplo, cuando recibe una acción, condición para la transición de estado, etc. El comportamiento del objeto puede ser informal para ayudar al operador (humano) a interpretar la condición del recurso gestionado. Dicha definición se puede considerar como información de guía.
- Vinculación de nombres : dentro del sistema gestionado, los objetos gestionados tienen identidades únicas que se usan para la comunicación de gestión. La identidad única se conoce como nombre distinguido (DN - *distinguished name*). Este describe una posición del objeto gestionado en un árbol de nombres como una secuencia de nombres distinguidos relativos (RDN - *relative distinguished name*). Cada miembro de la secuencia representa en nivel en el árbol.

La sintaxis del atributo se define usando la notación de sintaxis abstracta número uno (ASN.1 - *Abstract Syntax Notation One*), que define de manera no ambigua el tipo de datos (p. ej., INTEGER, STRING, SEQUENCE OF, etc.) de la información y, por lo tanto, permite la exacta interpretación de datos entre gestor y agente.

2.3. Arquitectura Física TMN

La arquitectura física describe las interfaces que tienen que ser implementadas, junto con ejemplos de los componentes físicos que forman el TMN. Los bloques de la arquitectura física son :

- Sistema de Operación (OS - *Operation System*) ;
- Red de Comunicación de Datos (DCN - *Data Communication Network*) ;
- Dispositivo de Mediación (MD - *Mediation Device*) ;
- Elemento de Red (NE *Network Element*) ;
- Adaptador Q (QA - *Q Adapter*) ;
- Estaciones de Trabajo (WS - *Work Stations*) ;

La Fig. 2 muestra como se interconectan estos elementos.

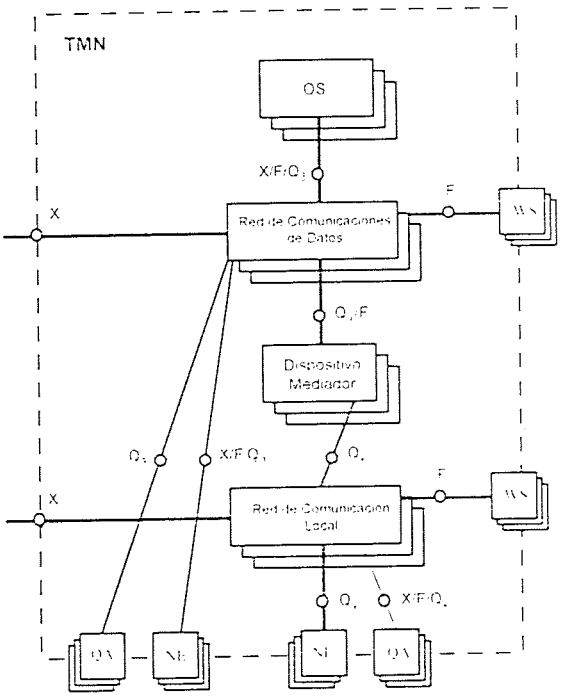


Fig. 2 - Arquitectura Física TMN

2.4. Paragdima Gestor- Agente

La gestión de redes de telecomunicaciones es una aplicación informática. Como la red de telecomunicaciones por si misma es un entorno distribuido, su gestión también lo es. Esto implica el intercambio de información de gestión entre procesos de gestión con la finalidad de supervisión y control de varios recursos físicos y lógicos de la red.

Los procesos informáticos envueltos en la gestión de redes de comunicaciones pueden asumir dos posibles papeles :

a) Gestor, parte de la aplicación distribuida que emite operaciones de gestión y recibe notificaciones ;
b) Agente, parte de la aplicación distribuida que gestiona los objetos asociados. El papel del agente es responder a las operaciones de gestión emitidas por el gestor y también suministrar al gestor una visión de estos objetos, emitiendo notificaciones que digan el comportamiento de estos objetos.

La Fig. 3 ilustra el funcionamiento entre gestor-agente.

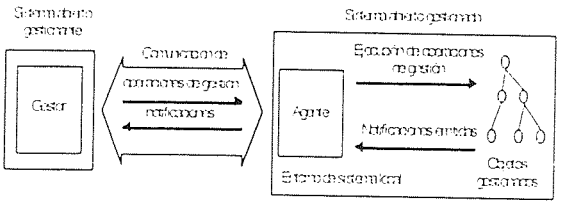


Fig. 3 - Agente - Gestor

Típicamente, existe una relación "muchos a muchos" entre gestores y agentes. También un agente puede rechazar una operación de gestión de un gestor por varias razones, por ejemplo en la inconsistencia del modelo de información o violación de aspectos de seguridad. Por lo tanto un gestor debe estar preparado para tratar peticiones negativas de un agente.

Todos los intercambios de información entre gestores y agentes obedecen a un conjunto consistente de operaciones de gestión y notificaciones, efectuándose siempre a través de Servicio y Protocolo de Información de Gestión Comun (CMIS/CMIP - *Common Management Information Service/Common Management Information Protocol*). Este protocolo es de nivel de aplicación en OSI.

2.4.1 Interfaces Estandares TMN

El objetivo de la especificación de las interfaces es asegurar la compatibilidad de los dispositivos interconectados. Esto requiere protocolos de comunicaciones compatibles y un método compatible de representación de datos para toda la información cursada en TMN.

Un conjunto mínimo de protocolos a ser utilizado en las interfaces TMN debe ser determinado (ver Figuras 1 y 2).

El interfaz Q es aplicable en el punto de referencia q. Para proveer flexibilidad de implementación la clase del interfaz Q es subdividida en dos subclases :

a) interfaz Qx aplicable en el punto de referencia qx ;
b) interfaz Q3 aplicable al punto de referencia q3.

Los interfaces Qx y Q3 se distinguen básicamente por las informaciones de gestión que transmiten. El interfaz Qx se caracteriza por aquella parte del modelo de información que es compartida entre el dispositivo de mediación (MD) y los elementos de red (NEs). Mientras que el interfaz Q3 se caracteriza por las partes del modelo de información compartida entre los sistemas de operación (OSs) y los elementos del TMN que realizan interfaz con estos OSs.

El interfaz Q3 implementa a nivel de aplicación las primitivas CMISE/CMIP. En la Fig. 5 se presenta un ejemplo de interfaz Q3 con todos los distintos protocolos en los diferentes niveles OSI.

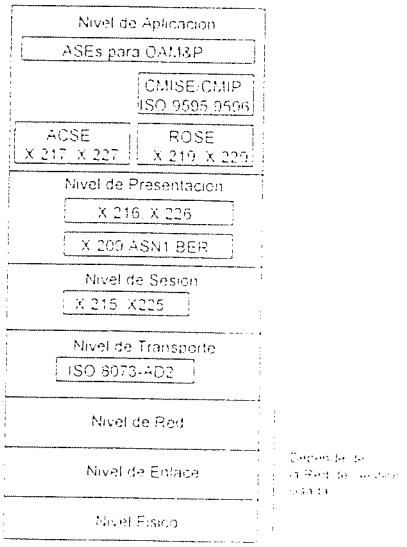


Fig. 5 - Ejemplo Interfaz Q3

El interfaz F es aplicable al punto de referencia f. Este interfaz conecta las Estaciones de Trabajo (WSs) al los Sistemas de Operaciones o a los Dispositivos de Mediación a través de la Red de Comunicaciones de Datos (DCN).

El interfaz X es aplicable en el punto de referencia x y es utilizado para interconectar dos TMNs o para interconectar un TMN con otros sistemas de gestión de red que poseen una interfaz no TMN.

2.5 Gestión de la Red en Niveles

Como se comentó en la arquitectura de información TMN, la gestión de red puede dividirse desde un punto de vista conceptual en varios niveles [7]. Cada nivel desempeña ciertas funciones y ofrece al nivel superior una visión abstracta de la red. La Fig. 6 ilustra conceptualmente la pirámide de niveles de gestión.

Pueden definirse interfaces de servicio entre los niveles similares a los empleados entre protocolos de comunicación. El nivel inferior ofrece ciertos elementos de servicios que pueden ser utilizados por el nivel superior inmediato.

Se describen cuatro niveles de gestión de red :

- Gestión de elemento de red - gestiona cada elemento de red sobre una base individual o de grupo, y soporta una abstracción de las funciones suministradas por la capa de elemento de red. La capa de gestión de elementos tiene una o varias OSF de elemento y o MF, que

tienen la responsabilidad individual, transmitida por la capa de gestión de red, de algunos subconjuntos de funciones de elementos de red. Como objetivo, se dará una visión de la capa de gestión de red independiente del vendedor.

- Gestión de red - tiene la responsabilidad de la gestión de una red soportada por la capa de gestión de elementos. En esta capa están situadas las funciones relativas a la gestión de una zona geográfica amplia. Es típico que haya una visibilidad completa de la totalidad de la red y, como objetivo, se suministrará a la capa de gestión de servicio una visión independiente de la tecnología.
- Gestión de servicios - tiene que ver con los aspectos contractuales de los servicios que se suministran a los clientes o que están disponibles para nuevos clientes potenciales, y es responsable de los mismos. Algunas de las funciones principales de esta capa son el tratamiento de los pedidos de servicio, las quejas y la facturación [8].
- Gestión de negocio - tiene la responsabilidad de la totalidad de la empresa. La capa de gestión de negocio abarca funcionalidades de dominio privado. La capa de gestión empresarial se incluye en la arquitectura TMN para facilitar la especificación de las capacidades que requiere de las otras capas de gestión.

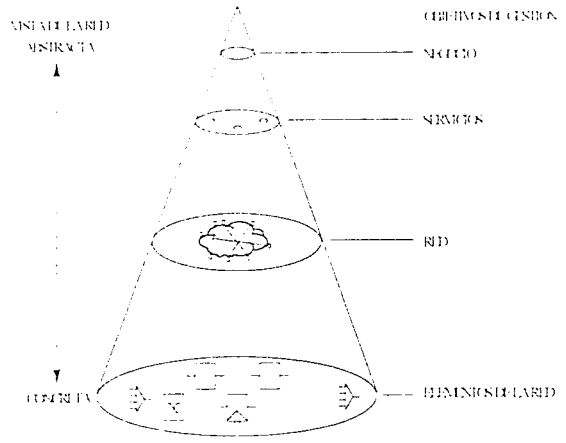


Fig. 6 - Visión de la Red en Niveles de Gestión

2.6 Funciones asociadas en TMN

El TMN ha sido concebida para soportar una gran diversidad de áreas de gestión que abarcan la planificación, instalación, operaciones, administración, mantenimiento y la puesta en servicio de redes de telecomunicaciones y la prestación de servicios.

ITU y ISO clasifica en cinco grandes áreas funcionales de gestión [3, 7]:.

- gestión de la calidad de funcionamiento - trata de las funciones necesarias para que un elemento de red coleccione, almacene, establezca umbrales y reporte los datos de calidad asociados con terminaciones de caminos de los elementos de red. Las recomendaciones G.784 y G.826 del ITU especifican las aplicaciones y parámetros opcionales relacionados con la gestión de calidad :
- gestión de fallos - responsable por el filtrado y gestión de alarmas resultante de varios acontecimientos. Proporciona las siguientes funciones como : detección, aislamiento y corrección de estados anormales en la red : vigilancia de alarmas : localización y chequeo de fallo :
- gestión de la configuración - proporciona un mecanismo para la gestión de elementos de red, o objetos gestionados, que están bajo el control del sistema de gestión. Proporciona a los operadores de red las facilidades para : soporte de las actividades de instalación de los elementos de red : control del estado de los elementos : cambio de configuración : inicialización de objetos activándolos y desactivándolos : aprovisionamiento de servicios y recursos bajo demanda :
- gestión de la contabilidad - ayuda en la preparación de facturas para los usuarios de la red, así como en el seguimiento de pago de las mismas. A veces también se incluye la gestión de inventario. Esta es la función que guarda la información de los elementos individuales manejados en la red, sus características, valores activos, etc :
- gestión de la seguridad - se encarga del control de acceso de usuario para proteger la red del acceso no autorizado a recursos y servicios. Se puede dividir en tres áreas : seguridad física, seguridad de acceso y seguridad de datos.

Estas funciones de gestión se encuentran mas detalladas en la recomendación M. 3400 del ITU [7].

3. Plataforma de Gestión de Red

Una plataforma de gestión de red es una colección de hardware y de módulos software normalizados, que soportan las diversas funciones de gestión. Lo importante de una plataforma de gestión sobre la cual se desarrollan las aplicaciones TMN es que debe ser modular, abierta, flexible y ampliable ya que sobre esta plataforma estará apoyado todo el sistema TMN [9].

La plataforma de gestión consta, básicamente de los siguientes componentes :

- Hardware :
- Software del sistema :
- Interfaz de usuario :
- Sistema de almacenamiento de datos :
- Conjunto de aplicaciones genéricas :

- Aplicación de gestión del sistema ;
- Interfaces de comunicación ;
- Entorno de desarrollo software .

4. Implementación Práctica

Este apartado trata de describir una implementación real de un sistema de gestión de una red de telecomunicaciones que cumple con la arquitectura TMN.

Este es el caso del SGRT (Sistema de Gestión de la Red de Telecomunicaciones) que desarrolla Unión Fenosa para la gestión de su red de telecomunicaciones.

En grandes rasgos, el SGRT considera al menos los siguientes elementos:

- la red a gestionar, que se compone de elementos de red multifabricantes.
- los agentes, que tiene una interfaz por un lado con el Gestor y por otro con los distintos elementos de red que gestiona.
- el gestor, a través de la cual el operador controla la red ;
- la plataforma de gestión, en la cual se ejecutan algunas de las aplicaciones de Gestión y o Agente, en este caso es el NetView 6000 TMN de IBM.

4.1 Elementos de Red

En una primera fase del sistema, la red a gestionar está formada por equipos Nokia PDH. Dentro de esta red existen varios tipos de equipos : multiplexores de varias jerarquías, equipos terminales de línea óptica, cross-conects y conmutadores de trama. Todos estos equipos disponen de un único protocolo de gestión propietario, por el cual el operador de red accede a cada elemento, independientemente, a través de un terminal de mano alfanumérico.

La información de gestión es distinta para cada tipo de equipo. Existen algunos dispositivos que disponen de distintas tarjetas (tarjetas de canales de voz y tarjetas de canales de datos a 64 kbit/s) para la configuración hardware final del equipo.

4.2 Agente

El Agente consta de dos partes :

1. El Adaptador-Q : Corresponde a la parte no TMN del agente. Es el mecanismo de comunicación con los recursos subyacentes ; el código que implementa el interfaz entre los equipos TMN y los recursos reales. Su papel es reflejar las capacidades de una interfaz de gestión para TMN. El adaptador-Q se compone de tres elementos :

- El interfaz físico : dependiente del suministrador de los Elementos de Red. En el caso de los equipos Nokia PDH es una salida ASCII V.11.
- El nivel de enlace de datos (*Data Link*)
- El driver del equipo : gestiona el enlace serie a nivel de Enlace de Datos y transfiere las tramas de bits al Agente TMN. Esta transferencia se realizar por medio de colas de mensajes. El driver del equipo incluye el Gestor de Interrupciones (*Interruption Handler*), que detecta las notificaciones eventos emitidas por los recursos y las transfiere al Agente TMN donde se convertirán en eventos-informes CMIP.

2. La parte TMN : Juega el papel de interfaz con las aplicaciones del gestión, ejecuta mandos recibidos desde el gestor y le devuelve información, mantiene la MIB (*Management Information Base*) y transfiere recibe datos hacia desde el Adaptador-Q . La parte TMN del Agente está compuesta por :

- el interfaz OSI : Incluye el interfaz CMIP del agente y la infraestructura de Comunicaciones.
- el interfaz del driver del equipo : Transfiere o recibe tramas de bits hacia desde el driver del equipo por medio de colas de mensajes.
- la MIB: Contiene las Clases de Objetos. La MIB es la imagen de la implementación real de los recursos que están siendo gestionados. La MIB puede ser persistente o no persistente. En esta fase del proyecto, donde sólo se implanta gestión a nivel de elemento de red, con un número limitado de Clases de Objetos y atributos, la MIB es no persistente.

La Fig. 7 ilustra la estructura de este Agente.

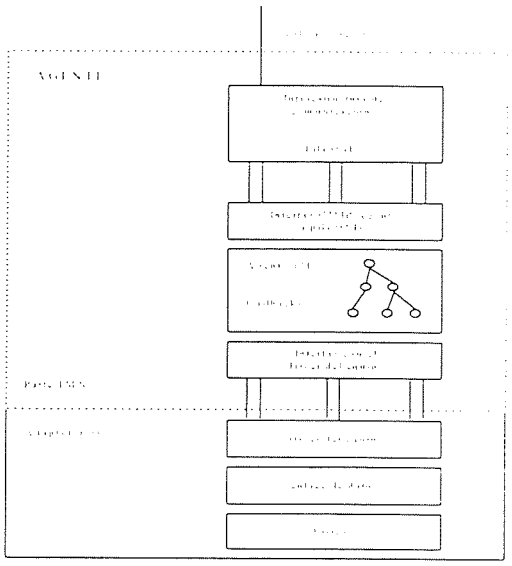


Fig. 7 - Estructura del Agente del SGRT

La Función Agente TMN se basa en :

1. Un modelo de información estándar, basado en la recomendación M. 3100 del ITU [6].
2. La Implementación de Información Específica : Se trata de una extensión específica a los equipos del modelo de información estándar para representar los equipos a gestionar [10].

Ambas partes se describen usando GDMO y ASN.1.

En la Fig. 8 se ilustra parte del modelo de información de utilizado a nivel de elemento de red por el SGRT, que se basa en los modelos genéricos del ITU y ETSI, respectivamente.

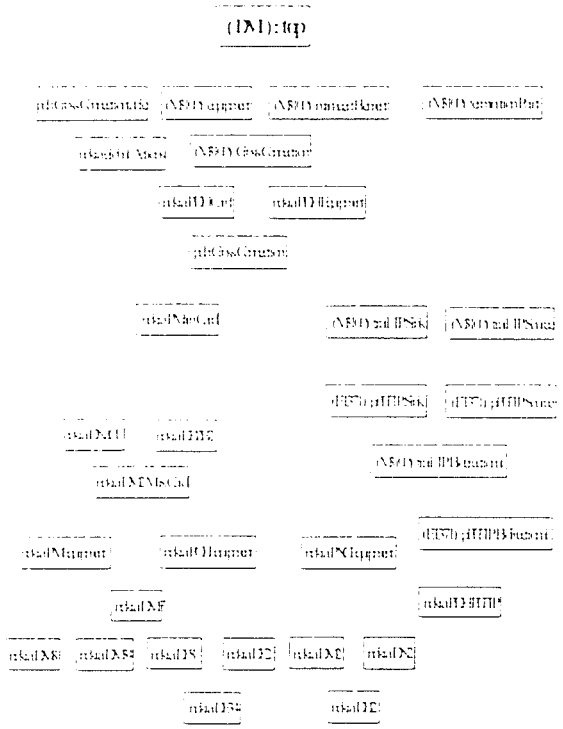


Fig. 8 - Ejemplo de Modelo de Información del SGRT

4.3 Gestor

Es la parte de una aplicación distribuida que envía las directivas de operación de gestión y recibe los eventos.

Es el gestor que inicia las operaciones de gestión y se mantiene la información del estado de los recursos gestionados a través de la comunicación CMIP con cada agente.

Actualmente es una aplicación gráfica que facilita enormemente los procesos de gestión de la red de telecomunicaciones.

4.3.1 Aplicaciones de Gestión de SGRT

Se distinguen tres tipos de aplicaciones :

1) Las Aplicaciones TMN genéricas son aplicaciones estándar que están disponibles con la Plataforma Gestora - NetView.

Las aplicaciones genéricas son :

1. Aplicación de Topología
2. Aplicación de Configuración
3. Aplicación de Fallos
4. Aplicación de Rendimiento
5. Aplicación de Registro (Log)

2) Las Aplicaciones TMN específicas son aplicaciones específicas del proyecto que se desarrollan por encima de la plataforma TMN para ajustarse a los requerimientos específicos del usuario. Se codifican en lenguaje C y C++, usando el TMN Workbench for AIX de IBM. Este es el caso de la aplicación de configuración específica - NECA (*Network Element Configuration Aplication*).

3) Las Aplicaciones no TMN son aplicaciones que se ajustan a los requerimientos del cliente y que no necesitan cumplir los estándares de la arquitectura TMN. Estas aplicaciones pueden intercambiar datos y control con las aplicaciones TMN, a través de los interfaces desarrollados.

5. Conclusiones

Una gestión avanzada de red de telecomunicaciones impone fuertes desafíos sólo superables desde soluciones estándares.

El trabajo desarrollado en el marco del proyecto SGRT de Unión Fenosa ha conducido a avances significativos en la gestión de la red de telecomunicaciones. A la vez que ha probado la validez de los estándares TMN y las herramientas de soporte existentes para abordar la implementación de este tipo de sistemas.

Los operadores disponen ahora de aplicaciones gráficas que permiten una gestión de los equipos de la red independientemente del fabricante, lo que supone un salto cualitativo importante.

Sin embargo, la principal aportación de este trabajo, no se queda ahí : sólo ahora cuando se aborda la integración de distintas tecnologías como SDH y PDH y el desarrollo de aplicaciones más avanzadas para gestionar la conectividad extremo a extremo de la red, se percibe como fundamental la existencia de una visión de los equipos unificada. Esta única visión independiente de la tecnología y fabricante, se materializa en el uso de modelos de información estándares en los agentes desarrollados.

6. Referencias

[1] CIGRE SC 35 WG 02 (1993), *Draft Report on Telecommunications Network Management in Power Utilities*

[2] ITU-T Recommendation M.3010 (1992), *Principles of a Telecommunicaction Management Network*.

[3] ITU-T Recommendation X.700 (1992), *Management framework for Open Systems Intercomnection (OSI) for ITU-T applications*.

[4] ITU-T Recommendation X.701 (1992), *Information technology – Open Systems Intercomnection – Systems management overview*.

[5] Hebrawi, Baha, *GDMO - Object Modelling & Definition for Network Management*. Ta - Tecnology Apprails - UK (1995).

[6] ITU-T Recommendation M.3100 (1995), *Generic network information model*.

[7] ITU-T Recommendation M.3400 (1992), *TMN management functions*.

[8] ITU-T Recommendation M.3200 (1992), *TMN management services: overview*.

[9] C. Guerrero, D. Sanchez, V. Carneiro, A. Viña, J. Coego, *Integrating Proprietary Managed PDH Networks using TMN-Based Platforms*". IEEE Enterprise Networking Miniconference (ENM) asociada a ICC 1997.

[10] C. Guerrero, D. Sanchez, V. Carneiro, "*Introducing TMN in legacy networks: A step towards integrated standard management systems*" 8th IFIP IEEE International Workshop for Distributed Systems Operations and Management (DSOM) 1997.

Control de Tráfico

Estimación de Pérdidas para el Control de Tráfico ATM mediante Enlaces Virtuales

I. Herrero, A. Díaz Estrella y F. Sandoval.
Departamento de Tecnología Electrónica
E.T.S.I. de Telecomunicación, Universidad de Málaga
Campus Universitario de Teatinos, 29071 Málaga
Tfno: (95) 2132874. Fax: (95) 2131447. E-mail: nhr@dte.uma.es

Abstract:

This paper addresses the problem of estimating low-order cell loss rates (CLR) of aggregated traffic in an ATM network using neural networks. Accurate CLR samples are needed to train a neural network (NN) which takes call-admission decisions in a Neural Call Admission Control. To achieve this accuracy on real-time a novel estimation method is proposed. This method relates CLR samples taken at virtual links with different and slower output rate, to the CLR at the real link, by means of an NN, thus solving accuracy and estimation-length problems of real-link estimates. Zero Loss Bandwidth Patterns are introduced as prior information necessary to allow the ANN to rightly interpolate the real CLR from the virtual samples.

1. Introducción

El Modo de Transferencia Asíncrono ("Asynchronous Transfer Mode", ATM) es el protocolo de comunicaciones en el que se basará la futura Red Digital de Servicios Integrados de Banda Ancha (RDSI-BA). La tecnología ATM está diseñada para soportar una amplia variedad de servicios, con muy diferentes características de tráfico y requerimientos de calidad de servicio ("Quality of Service", QoS). La flexibilidad en la gestión de ancho de banda, la capacidad de gestión de sus servicios de forma integrada, y el aprovechamiento de la multiplexación estadística son sus principales ventajas. Sin embargo, estas ventajas implican también la necesidad de complejos mecanismos de control de tráfico y congestión para cumplir con los contratos de tráfico establecidos, al tiempo que se emplean los recursos de la red de forma eficiente.

Por ello, el Control de Tráfico en las redes ATM es una de las áreas de investigación de mayor interés en el desarrollo de dichas redes. En el presente trabajo nos centraremos en un elemento del Control de Tráfico, el Control de Admisión de Conexiones (CAC), que actúa de forma preventiva, por lo que resulta de gran interés para los servicios en tiempo real [1].

La mayoría de las soluciones propuestas para el control de admisión, y en general para el control de tráfico ATM, tienen serios inconvenientes. Algunas son sencillas, pero incluyen aproximaciones y supuestos que son difíciles de justificar [2-3]. Otras se basan en complejos análisis matemáticos de modelos analíticos de tráfico que, por motivos computacionales, no podrían aplicarse para servicios en tiempo real y que además se centran en casos muy particulares [2] [4-7]. Estos son los motivos que nos han llevado a emplear Redes Neuronales Artificiales (RNAs), como elemento principal del control de admisión de conexión. Sus características de aprendizaje adaptativo, capacidad de generalización, alta velocidad de computación y tolerancia a fallos, las hacen especialmente aptas para este tipo de aplicaciones.

Por otra parte, la necesidad de adaptarse en tiempo real al tráfico que discurre por la red ATM, y el desconocimiento de la naturaleza estadística de algunos de los servicios de tráfico actuales -además de los servicios futuros- , aconsejan la obtención de la información de tráfico necesaria para la actuación del CAC mediante estimaciones en la propia red ATM. Son los denominados CAC basados en monitorización de tráfico [8]. Estos controles de admisión, tanto si se basan en métodos analíticos como en RNAs, necesitan obtener información muy precisa acerca de aspectos del agregado de tráfico tales como la tasa de pérdida de células ("Cell Loss Rate", CLR) y el retardo, los cuales a menudo resultan muy complicados de estimar en tiempo real, debido a las propias características de ATM. Así, han surgido también una serie de métodos que estudian cómo obtener estimaciones correctas de los parámetros de interés. En el presente documento, proponemos un método de estimación de la tasa de pérdida de células para el agregado de tráfico ATM, que emplearemos posteriormente en un Control Neuronal de Admisión de Conexiones (NCAC).

El esquema a seguir es el siguiente: en el apartado dos se realiza un rápido repaso a los controles de admisión en ATM, incidiendo en los basados en redes neuronales, que son brevemente comentados. El apartado tres se centra en el problema de la estimación de pérdidas en los CAC basados en monitorización del sistema; también se revisan algunas técnicas de reducción de varianza existentes, empleadas para intentar resolver este problema. Dentro de estas técnicas se presta especial atención a las basadas en subsistemas virtuales, centrándose finalmente en la técnica de estimación mediante enlaces. En el apartado cuatro se presentan las pruebas realizadas con esta última

técnica; en primer lugar unas pruebas preliminares con objeto de estudiar sus posibilidades de aplicación, partiendo de la versión original de Hiramatsu [9], para añadir después unas modificaciones necesarias para un correcto funcionamiento. A continuación, en el apartado cinco, se muestran los resultados obtenidos hasta la fecha en un CAC Neuronal con estimación de pérdidas mediante enlaces virtuales y en tiempo real. Las pruebas se realizaron para tráfico homogéneo y heterogéneo (dos clases de tráfico), en ambos casos empleando tráfico On-Off. Por último, en el apartado seis se indican las conclusiones correspondientes.

2. Control Neuronal de Admisión de Conexiones

El Control de Admisión de Conexiones tiene la misión de decidir sí se debe aceptar o rechazar una petición de conexión de una nueva llamada que accede al nodo ATM. Para aceptar una nueva llamada, el CAC debe asegurar que con dicha aceptación se sigue manteniendo la QoS de las llamadas ya establecidas, y se cumple la demandada por la nueva llamada. Además, se trata de maximizar el uso de recursos de la red, aprovechando, mediante la multiplexación estadística, la naturaleza rafagueante de muchos tipos de tráfico.

Se han desarrollado bastantes implementaciones de CAC basadas en métodos analíticos, como el cálculo del ancho de banda equivalente de cada llamada y del agregado [2][4], y estudio de la tasa de pérdidas a través de modelos de colas y de fuentes de tráfico [5-7]. Sin embargo, estos métodos presentan importantes inconvenientes para su aplicación práctica en tiempo real. Su excesiva complejidad implica tiempos de cálculo demasiado largos para un control en tiempo real, y las aproximaciones adoptadas para reducir estos tiempos de cálculo producen errores que limitan su utilidad.

En contraposición, los CAC Neuronales [1][9-12] utilizan una red neuronal para tomar la decisión de admitir o no la nueva llamada, aprovechando todas las ventajas propias de las RNAs:

- Aprendizaje Adaptativo: las RNAs pueden aprender de la observación de ejemplos tomados durante el funcionamiento de la red ATM. Por tanto no se necesita un conocimiento matemático exhaustivo del proceso de tráfico.
- Alta velocidad de computación, debido al paralelismo masivo en la implementación hardware de la RNA.
- Capacidad de Generalización, útil para situaciones en las que las observaciones son incompletas.
- Tolerancia a Fallos, por la naturaleza distribuida del proceso.

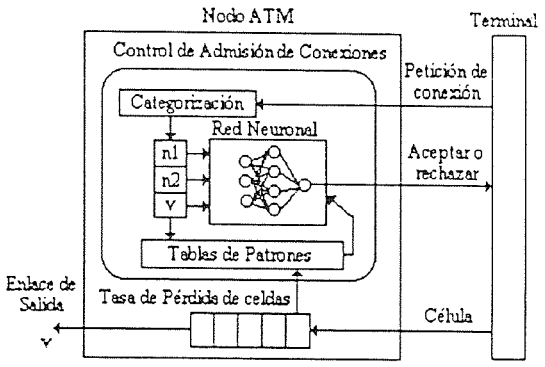


Figura 1. Control de Admisión Neuronal propuesto por Hiramatsu [9]

En la figura 1 se presenta un esquema de CAC Neuronal propuesto por Hiramatsu.

2.1 Redes Neuronales en un CAC

Una red neuronal es, generalmente, un elemento que permite mapear múltiples entradas con múltiples salidas, aprendiendo la relación -lineal o no lineal- entre entradas y salidas a partir de un conjunto de muestras de dicha relación. Una RNA consta de elementos de procesamiento simples llamados neuronas, agrupados de varias formas según la arquitectura de la RNA. Cada neurona tiene varias entradas y una salida cuyo valor viene dado por una función no lineal de sus entradas. Las conexiones entre neuronas se denominan pesos, y la relación entrada-salida de la RNA se modifica cambiando el conjunto de valores de pesos o vector de pesos.

El algoritmo mas ampliamente utilizado para ajustar el vector de pesos al conjunto de muestras de aprendizaje es el Backpropagation [13], que ajusta los pesos según el gradiente de la función error.

En la versión de NCAC presentada, se emplea una RNA con arquitectura de Perceptrón Multicapa Feedforward (PMF), en la que las neuronas se agrupan en capas -normalmente tres- recibiendo como entradas las salidas de la capa anterior, sin que exista ningún tipo de realimentación en cada neurona. Esta RNA tratará de aprender una función que relacione el número de conexiones activas en el nodo ATM con la tasa de pérdidas asociada al agregado de tráfico correspondiente. Por tanto, tiene tantas neuronas de entrada como clases de tráfico se consideren en el nodo ATM; dos neuronas ocultas, ya que al ser la función a aprender cuasi-lineal se consigue un mejor comportamiento en el aprendizaje; y una neurona de salida, que proporcionará la tasa de pérdida de células estimada por la RNA para una determinada combinación de llamadas activas.

En al apéndice I se presenta un resumen de los fundamentos matemáticos de la neurona, la arquitectura PMF, y el algoritmo de aprendizaje Backpropagation.

2.2 Funcionamiento del CAC Neuronal

Las tres operaciones básicas que se realizan en un CAC Neuronal son la monitorización de tráfico, el aprendizaje de la red neuronal y la toma de decisiones.

La monitorización de tráfico tiene por objeto obtener estimaciones de algún parámetro de tráfico de interés. En nuestro caso trataremos de obtener muestras de la tasa de pérdidas asociada al agregado de tráfico correspondiente a una combinación de conexiones activas. El par formado por un vector de estado del sistema -con el número de conexiones activas de cada clase de tráfico- junto con la tasa de pérdidas medida para dicho estado en un intervalo de tiempo, constituye un patrón de entrenamiento de la RNA. Los patrones se almacenan en una memoria para entrenar con ellos a la RNA del CAC Neuronal. Como se verá posteriormente, a menudo es muy complicado obtener muestras suficientemente precisas para conseguir un entrenamiento adecuado. Este trabajo se centra principalmente en cómo obtener estimaciones útiles para este fin.

El entrenamiento de la RNA se realiza de forma cíclica, escogiendo aleatoriamente un patrón de la memoria de almacenamiento. Para evitar perder información de zonas del espacio de entradas ya visitadas, cuando la memoria se llene, se suele dividir la memoria de patrones en ventanas, cada una de las cuales se corresponde con una zona determinada del espacio de entradas [14].

Por último, la toma de decisiones se realiza observando qué tasa de pérdidas asocia la RNA a la combinación de llamadas que resultaría si se admitiese la nueva llamada. Si la tasa predicha excede los parámetros de calidad de servicio, se rechazará la nueva llamada; en caso contrario, se aceptará.

Aunque esta es la estructura y funcionamiento general de un CAC Neuronal, se verá a continuación como este esquema no es suficiente para obtener un buen CAC en tiempo real, debido a las imprecisiones en las estimaciones de pérdidas empleadas para el entrenamiento de la RNA.

3. Estimación de la tasa de pérdidas en ATM: el Problema

La estimación de las pérdidas asociadas al tráfico de la red ATM es una actividad fundamental para el buen funcionamiento de un CAC Neuronal. Es imprescindible proporcionar a la RNA patrones de entrenamientos fiables y precisos, a partir de los cuales pueda generalizar el comportamiento del tráfico de forma correcta.

Un método para asegurar la precisión de estos patrones, consiste en trabajar en modo OFF-LINE [15] asegurando, mediante simulaciones de tráfico controladas, que las diversas combinaciones de llamadas permanecen estáticas el tiempo suficiente como para garantizar unas estimaciones precisas de las pérdidas asociadas. Una vez que la RNA se entrenase con estos patrones, el CAC estaría listo para entrar en funcionamiento. Sin embargo este método no es aplicable a situaciones reales, ya que exige un conocimiento previo de los tipos de tráfico que van a acceder al nodo ATM, conocimiento que a menudo no es posible. Por otro lado, esta forma de trabajo haría imposible la adaptación del CAC ante un tráfico cambiante, ya que la aparición de nuevas características de tráfico obligaría a parar el CAC, obtener nuevos patrones de entrenamiento de forma controlada y entrenar de nuevo a la RNA, antes de reanudar la operación del CAC.

Como una posible solución se propone la obtención de patrones de entrenamiento ON-LINE, de forma simultánea al funcionamiento del CAC en tiempo real. Esto permitiría una adaptación continua del CAC a las condiciones del tráfico en el enlace ATM, pudiendo manejar no sólo tráfico cambiante, sino con características desconocidas.

Desafortunadamente, este modo de operación también presenta inconvenientes importantes. Debido a la falta de control sobre la evolución del tráfico, las combinaciones de llamadas permanecen estáticas un tiempo indefinido, generalmente demasiado pequeño como para obtener estimaciones precisas de la tasa de pérdidas asociada [16]. Este problema se ve acentuado por los bajos órdenes de pérdidas con los que es necesario trabajar en las redes ATM, que pueden llegar hasta niveles de 10^{-10} e incluso 10^{-12}. Una tasa objetivo de pérdidas de 10^{-10} implica la pérdida de una célula cada 10.000 millones, con lo que para obtener una estimación precisa para este orden de pérdidas, el numero de células gestionadas por el nodo ATM en un intervalo de estimación ha de ser mayor de 10.000 millones. Así, para un enlace ATM de 150 Mbps con una carga de tráfico de 0'6, el sistema debería permanecer sin cambiar el número de conexiones activas un tiempo superior a 13 horas. Obviamente, esto no se corresponde con una situación real de tráfico en un nodo ATM.

La dinámica relativamente rápida del estado del sistema, de su número de conexiones activas, obliga a considerar intervalos de estimación con un tamaño muy inferior al necesario, con las imprecisiones en la estimación que esto conlleva. Por ello, tendremos que descartar las muestras cuyo tiempo de estimación sea inferior a un valor mínimo, para evitar entrenar a la RNA con una información excesivamente imprecisa, que impediría un correcto aprendizaje. Sin embargo, como se vio anteriormente, si se quiere obtener suficiente número de patrones de aprendizaje deberemos emplear un tiempo de estimación menor que el necesario, admitiendo las imprecisiones

asociadas, e intentado compensarlas de alguna forma. Estas imprecisiones se pueden compensar, en parte, por medio de la RNA [9]. Si por ejemplo, para la tasa objetivo mencionada de 10^{-10}, consideramos un intervalo de estimación más razonable, de un millón de células -equivaldría a unos 4´6s., en las condiciones antes mencionadas-, este intervalo permitiría estimar unas pérdidas de hasta un orden de 10^{-6}. Si se obtiene un número suficientemente grande de patrones para una misma combinación de entradas -unos 10.000- en alguno de ellos se habrán detectado pérdidas de orden 10^{-6} o menor, y en el resto pérdidas nulas, de forma que, considerando el conjunto de la muestra, el valor medio del mismo estará muy cerca del valor real. Entrenando a la RNA con todo el conjunto de patrones se consigue que aprenda este valor medio casi correcto, resolviendo el problema.

Sin embargo, es fácil comprobar que, si el número de posibles estados del sistema es muy grande, se necesitará una cantidad excesiva de memoria para almacenar todos los patrones necesarios.

De aquí la necesidad de encontrar nuevos métodos de estimación que permitan trabajar en tiempo real en una red ATM.

3.1. Técnicas de reducción de varianza

En los anteriores apartados se han visto las dificultades existentes para obtener estimaciones precisas de pérdidas en ATM y en tiempo real. Generalmente nos veremos obligados a trabajar con intervalos de estimación de un tamaño inferior a 100.000 células -unos 460 ms. para 150 Mbps de enlace y carga 0´6- , ya que la dinámica del sistema impediría obtener suficientes patrones para un tamaño mayor. Por este motivo, se debe aplicar alguna técnica de reducción de varianza en el conjunto de muestras obtenidas para compensar las imprecisiones y así aprender los valores correctos.

Algunas de las técnicas tradicionales de reducción de varianza, como el análisis MonteCarlo, no son aplicables al problema , al estar orientadas a la simulación, mientras que las estimaciones se deben realizar en tiempo real.

Por otra parte, existe otra metodología para la estimación de probabilidad de eventos raros, basada en la teoría de grandes desviaciones ("large deviations") [17]. Estas técnicas estudian las funciones de energía libre y entropía de los eventos raros -CLR para el caso que se estudia-, desarrollando unas funciones de gran desviación para la probabilidad de ocurrencia de dichos eventos. A través de un complejo estudio estadístico, se pueden derivar de estas funciones aproximaciones de los valores a estimar [18]. Los principales inconvenientes de estas técnicas son su excesiva complejidad matemática -es necesaria una gran base matemática para analizar las situaciones de tráfico a estudiar- y las aproximaciones y supuestos necesarios para compensar esta complejidad, que introducen imprecisiones en los cálculos.

Sin embargo, si que se pueden aprovechar las ideas fundamentales de otra técnica muy empleada, como es el Muestreo de Importancia ("Importance Sampling" [19]) para nuestros fines.

El Muestreo de Importancia es una técnica de reducción de varianza utilizada principalmente para acelerar simulaciones por computadora, en la que la estimación de las muestras se realiza en proporción a su importancia relativa en el resultado final. Se trata de buscar una nueva distribución del parámetro a estimar en la que la varianza de las estimaciones se minimice, y que se pueda relacionar con la distribución real. Este último aspecto, denominado ponderación o "biasing", es uno de los que presenta mayores problemas en las técnicas basadas en Muestreo de Importancia.

Algunas aplicaciones de esta técnica se basan en repeticiones, con diferentes semillas, de los intervalos de simulación en los que hay mayor probabilidad de aparición del evento raro a estimar - la pérdida de una célula-, para posteriormente estimar la probabilidad real de ocurrencia del evento en función de su número de apariciones en los intervalos de repetición [20-21]. Otros autores modifican en determinados momentos de la simulación las características de tráfico, incrementando la carga de tráfico durante los periodos de estimación, y posteriormente relacionando las muestras obtenidas en estos periodos con los resultados reales, en función de como se modificó esa carga [22].

Este tipo de métodos no son aplicables al problema de realizar estimaciones precisas de pérdidas en tiempo real, ya que actúan sobre las fuentes de tráfico, ya sea deteniéndolas para repetir intervalos o alterando su carga de tráfico. Se debe buscar simplemente observar el sistema, sin alterarlo en forma alguna.

Una forma de conseguir este objetivo consistiría en estudiar como se comportaría el tráfico ante unas condiciones diferentes del medio de transmisión, buscando provocar así un aumento de la tasa de pérdidas a observar, para posteriormente relacionar esta tasa de pérdidas con la correspondiente al sistema real. Este es el procedimiento a seguir en las técnicas basadas en Subsistemas Virtuales.

3.2. Estimación mediante Subsistemas Virtuales

Este tipo de técnicas se deben considerar como un subconjunto del Muestreo de Importancia, ya que también trabajan con situaciones diferentes de las reales, que facilitan la estimación del evento raro, y a partir de las cuales se pueden obtener los valores correspondientes a la situación real.

Como su propio nombre indica, estas técnicas estudian como evoluciona la tasa de pérdidas a estimar, ante unas mismas condiciones de tráfico cuando se modifica alguna característica del medio. Esta modificación del sistema debe permitir obtener más fácilmente muestras correctas del parámetro a estimar -tasa de pérdidas-, no debe implicar alteraciones importantes o complicadas en el sistema, y se debe poder relacionar las pérdidas obtenidas en las nuevas condiciones con el valor correspondiente al sistema real.

Las dos opciones principalmente utilizadas como subsistemas virtuales son las colas o "buffers" virtuales [16][23-25] y los enlaces virtuales [9]; en estos últimos se basa la técnica propuesta.

3.2.1 Colas Virtuales

En una red ATM, existe al menos una cola de servicio por cada salida en cada nodo de la red -también pueden existir sistemas multicola en caso de que se quieran tratar por separado las pérdidas de las distintas clases de tráfico admitidas-. Esta cola tiene por objeto absorber el efecto de la afluencia simultánea de células procedentes de las fuentes que acceden al nodo. A mayor tamaño de cola, mayor capacidad tendrá el nodo ATM para amortiguar la llegada de un mayor número de células; el tamaño máximo de la cola de servicio suele estar limitado por consideraciones de retardo de células.

Debido a los bajos ordenes de pérdidas con los que se trabaja en ATM, en la mayoría de los casos, las estimaciones en cola no lograrán medir pérdida alguna, siendo esta información poco útil a efectos de monitorización de la función de pérdidas.

La técnica de estimación de pérdidas mediante colas virtuales se basa en estudiar como evoluciona la tasa de pérdidas medida en una cola, cuando ésta reduce su tamaño [16].

A la cola de entrada real, de capacidad C, se le asocian varias colas virtuales C1, C2, C3,..., de tamaño menor pero con el mismo comportamiento en cuanto a tráfico entrante y saliente (Fig. 2). Al reducir su tamaño, su probabilidad de saturación aumenta, permitiéndonos detectar pérdidas en condiciones en las que no sería posible en la cola

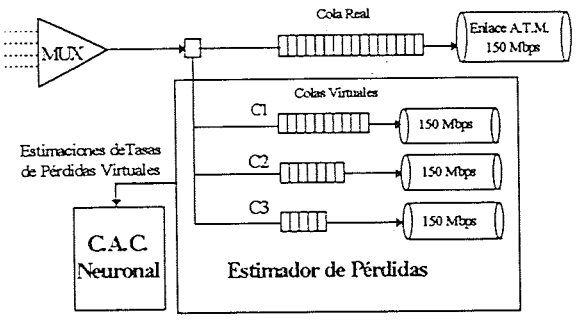


Figura 2. Estructura de Colas Virtuales para Estimación de Pérdidas

real, para un tiempo de estimación razonable. Por último, queda la cuestión de cómo relacionar estas estimaciones virtuales con la CLR buscada en la cola real.

En este sentido, varios autores han tratado de identificar un modelo analítico de función CLR vs. Tamaño de Cola -sería la función de "biasing" de las muestras-, en el que existirían unos parámetros a estimar mediante medidas en tiempo real en las colas virtuales [23-24]. En algunos casos se ha llegado a asumir un comportamiento lineal de la función de pérdidas bajo determinadas condiciones de tráfico [25]

Estas técnicas de estimación pueden dar buenos resultados para situaciones muy determinadas, pero tienen el inconveniente de asumir demasiadas suposiciones (por ejemplo, uso muy elevado de la multiplexación estadística, distribución normal del agregado de tráfico resultante, basarse en modelos de tráfico que no siempre son equivalentes al tráfico real,...) lo cual impide su generalización a todo tipo de tráfico; además, suelen necesitar una gran potencia de computación debido a su complejidad, dificultando su aplicación en tiempo real.

Para generalizar esta técnica de estimación a diversos tipos de tráfico, sin tener la necesidad de hacer un estudio matemático previo detallado, se propuso que el "biasing" lo realizase una RNA, a partir de las estimaciones en tiempo real en las colas virtuales [26].

La figura 3 muestra las evolución de la tasa de pérdidas a medida que aumenta el tamaño de la cola de servicio, para una clase de tráfico On-Off con 100 conexiones activas, tiempos en estados On y Off con distribución geométrica de media 10 ms y velocidad de pico de 2'5 Mbps. Se consideró un enlace ATM de 150 Mbps. Las estimaciones de tasas de pérdidas se obtuvieron mediante simulaciones controladas de suficiente duración -hasta 800 millones de células gestionadas-. En la figura 3, se distinguen dos zonas bien diferentes, una en la que predomina el "efecto célula" -pérdidas por afluencia simultánea de células- , con una pendiente muy alta, y otra en la que predomina el "efecto ráfaga" -pérdidas por simultaneidad de ráfagas de diferentes fuentes de tráfico- , en la que por el contrario, la pendiente es muy pequeña. La zona de "efecto célula" suele ser ignorada en la mayoría de los estudios, al reducirse a una porción inicial muy pequeña, correspondiente a tamaños de colas muy pequeños y que no se suelen emplear en casos reales -en la Fig. 3 finaliza en un tamaño de cola aproximado de 10-. Pero si considera solamente la zona de "efecto ráfaga", surge el problema de que no se obtiene mucho beneficio, en cuanto a la reducción de la varianza en las muestras al emplear colas virtuales. En el ejemplo, hay cuatro órdenes de magnitud de diferencia entre las

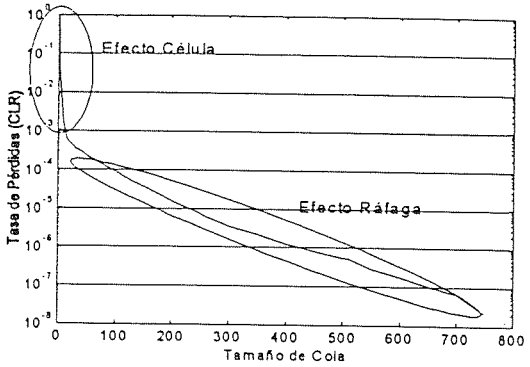


Figura 3. Función CLR vs. Tamaño de Cola

pérdidas en colas de tamaños 50 y 750 células, teniendo en cuenta que un tamaño de cola de 750 significa un retardo de célula, para 150 Mbps, de casi 2 ms., que puede ser demasiado alto para muchas aplicaciones. Además, si se entrena la RNA con muestras de colas de tamaños pequeños y cercanos -en la "zona ráfaga"- las muestras obtenidas tienen valores tan parecidos que no se consigue un buen entrenamiento de la RNA [26].

Estos problemas obligan a descartar el empleo de esta técnica con RNAs, para centrarse en el estudio de los Enlaces Virtuales.

3.2.2 Enlaces Virtuales

La técnica de estimación mediante enlaces virtuales, propuesta por Hiramatsu en [9], tiene un fundamento similar a las colas virtuales, pero en este caso lo que se estudia es la evolución de la tasa de pérdidas cuando disminuye el ancho de banda del enlace de salida del nodo ATM .

Para ello se asocian al nodo real una serie de elementos virtuales, -los enlaces virtuales, E1, E2, E3,..- constituidos, cada uno de ellos, por una cola del mismo tamaño que la cola real del nodo ATM pero con tasa de servicio de salida diferente y de menor velocidad que en el caso real (Fig. 4).

Así, para unas mismas condiciones de tráfico de entrada, estos enlaces virtuales se saturan con mayor facilidad que la cola real, permitiendo obtener una estimaciones mucho mas precisas con un menor intervalo de estimación, al trabajar con pérdidas de mayor orden de magnitud. Posteriormente se tratará de obtener el valor de pérdidas correspondiente al enlace ATM real, a partir de estas muestras virtuales, mediante una RNA.

Es necesario destacar la simplicidad de este esquema, que necesita únicamente de unos contadores adicionales y una lógica que determine el ritmo de salida de las células almacenadas en los enlaces virtuales.

En un CAC Neuronal que emplee esta técnica, la RNA tomará la decisión de aceptar o rechazar una nueva llamada relacionando el estado del sistema -el número de conexiones activas de cada clase de tráfico- y el ancho de banda del enlace virtual considerado con las tasas de pérdidas que se hubiesen estimado en dichas condiciones de tráfico y en dicho enlace. Para decidir la admisión de la nueva llamada - operación de "Recall" en la RNA - se introducirán como entradas, el futuro estado del sistema y el ancho de banda del enlace real, comparando la CLR predicha con la CLR objetivo.

En el siguiente apartado se verá cómo esta técnica, tal como se describe, no resulta suficiente para obtener un buen comportamiento del CAC Neuronal; son necesarias algunas modificaciones, como la inclusión de patrones adicionales de entrenamiento de la RNA, para mejorar dicho comportamiento.

4. Pruebas Preliminares

Previamente a la aplicación del método de estimación a un CAC en tiempo real, se han estudiado las características generales de la función CLR vs. Ancho de Banda, y si una RNA podría estimar correctamente la tasa de pérdidas correspondiente al enlace ATM real a 150 Mbps, a partir de muestras tomadas en enlaces de menor ancho de banda de salida. Para estas pruebas se considera una clase de tráfico On-Off [2] con 100 conexiones a 2'5 Mbps de velocidad de pico; tiempos en estados On y Off con distribución geométrica de igual media, 10 ms. El nodo ATM tiene un buffer de servicio con capacidad para 100 células.

4.1 Cálculo de la función CLR vs. Ancho de Banda

Mediante simulaciones controladas de gran duración -aproximadamente 100 millones de células gestionadas- se obtuvo la función CLR vs. Ancho de Banda de la figura 5 ("Función objetivo"), calculando la tasa de pérdidas asociada a distintos valores de ancho de banda del enlace de salida, desde 10 a 150 Mbps, en intervalos de 10 Mbps, para el tráfico descrito en el apartado cuatro. A diferencia de la función CLR vs. Tamaño de Cola, en esta función el decrecimiento con mayor pendiente se produce para valores altos de ancho de banda.

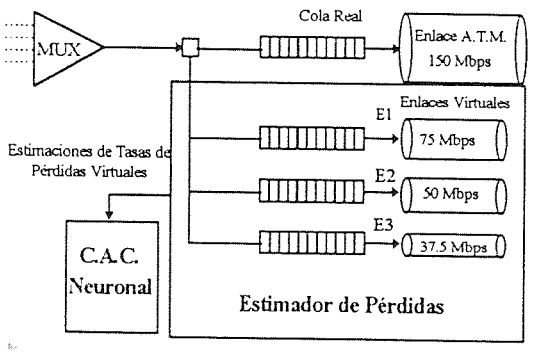


Figura 4. Estructura de Enlaces Virtuales para Estimación de Pérdidas

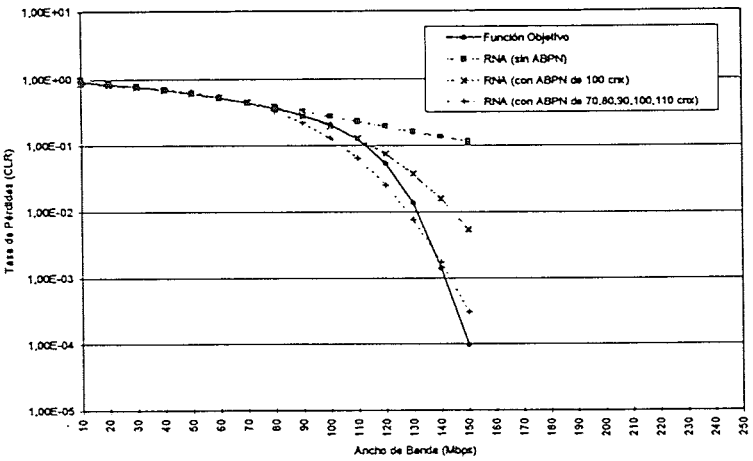


Figura 5. Curvas comparativas de la función CLR vs. Ancho de Banda aprendidas por la RNA en las distintas pruebas (100 conexiones)

4.2 Obtención de Muestras de la función objetivo

A continuación se estudia la capacidad de aprendizaje de la función por la RNA, mediante una serie de tests en modo OFF-LINE. Para ello, se obtuvieron, también mediante simulaciones controladas, 100 muestras de CLR, para las mismas condiciones de tráfico descritas antes, en enlaces virtuales con ancho de banda de E1=75, E2=50 y E3=37'5 Mbps -la mitad, tercera y cuarta parte del enlace real-; para obtener estas muestras se considera un intervalo de estimación de un tamaño mínimo razonable -unas 10.000 células, equivalente aproximadamente a 33'6 ms., para una carga de tráfico considerada de 0'83-. Se observó como la dispersión en el conjunto de medidas obtenidas en los enlaces virtuales era mucho menor que la que obtendríamos en las mismas condiciones de tráfico y medida para el enlace real (Fig. 6).

4.3 Aprendizaje de las Muestras. Definición del Ancho de Banda de Pérdidas Nulas

Una vez obtenidas las muestras de CLR, se estudió la capacidad de una RNA para aprender la función objetivo que relaciona CLR con ancho de banda. Estas muestras se emplearon para entrenar una RNA durante 1 millón de ciclos de aprendizaje.

La RNA tiene dos entradas, una para el número de conexiones activas -que permaneció fijada en 100 conexiones para estas pruebas-, y otra para el ancho de banda del enlace virtual en el que se tomó la estimación correspondiente. Al finalizar el entrenamiento, la RNA había aprendido la curva que se observa en la figura 5, con el nombre de "RNA sin ABPN". El aprendizaje de la RNA fue bueno en el área cuyos patrones de entrenamiento se le proporcionaron, obteniendo valores muy aproximados para anchos de banda del mismo o menor orden de pérdidas; sin embargo en la zona de interés, la correspondiente al ancho de banda real, la curva de la RNA tenía un comportamiento lineal, mientras que el comportamiento de la función real era exponencial decreciente.

Esto se debe a que una RNA con un número pequeño de neuronas ocultas tiende a extrapolar de la forma mas sencilla posible aquellas zonas de la función en las que carece de información. La solución a este problema pasaría por conseguir que la RNA interpolase el valor de CLR objetivo, en vez de extrapolarlo, para lo cual es necesario proporcionarle información de pérdidas correspondiente a anchos de banda mayores que el del enlace real.

Sin embargo, esto lleva de nuevo al problema original, de cómo obtener estimaciones precisas de

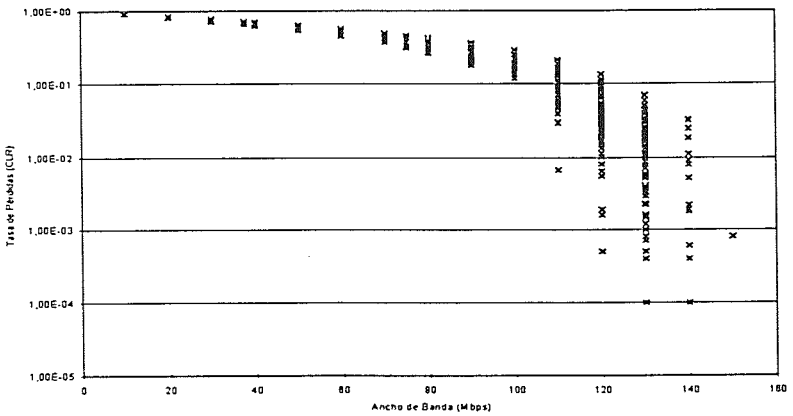


Figura 6. Evolución de la Dispersión de muestras para distintos anchos de banda del enlace

muy bajos órdenes de pérdidas con un tiempo razonable de estimación, ya que las pérdidas correspondientes a anchos de banda mayores que el objetivo son de ordenes aún menores.

Cómo solución para obtener esta información necesaria, se propuso el concepto de Ancho de Banda de Pérdidas Nulas (ABPN) [27]. Se define el ABPN como el ancho de banda necesario para el enlace de salida, en unas determinadas condiciones de tráfico, para asegurar pérdidas nulas en el nodo ATM, al menos a nivel de ráfaga. Aunque el cálculo del valor exacto de este ABPN podría resultar muy complicado, se puede conseguir un límite superior de forma muy sencilla, simplemente considerando que todas las conexiones activas trabajan siempre a su velocidad de pico. De esta forma, este límite superior vendría dado por la expresión:

$$ABPN(\vec{n}) = \sum_{k} n_i P_i \qquad (1)$$

donde $\vec{n}$ es el vector de estado del nodo ATM, considerando k clases de tráfico diferentes, y n_i y P_i el número de conexiones activas y la tasa de pico para la clase i, respectivamente.

Se repitió la prueba anterior, incluyendo el ABPN -que en este caso de 100 conexiones a 2'5 Mbps de velocidad de pico, vale 250 Mbps- con pérdidas nulas asociadas, en el conjunto de patrones de entrenamiento, obteniéndose una importante mejora en el aprendizaje de la función por parte de la RNA (Fig. 5, "RNA con ABPN de 100 conexiones"). Además, al incluir en el conjunto de patrones de entrenamiento muestras virtuales correspondientes a estados vecinos del objetivo -repetir el proceso del apartado 4.2, con combinaciones similares de conexiones- y los correspondientes ABPN de esos estados vecinos, la mejora en el aprendizaje consigue que la diferencia entre la función calculada por la RNA y la real sea menor de un orden de magnitud, con una forma muy parecida (Fig. 5, "RNA con ABPN de 70,80,90,100,110 y 120 conexiones").

Una vez vistas las posibilidades de la técnica de estimación propuesta, se debe demostrar su aplicación en un CAC Neuronal en tiempo real, para diferentes tipos de tráfico.

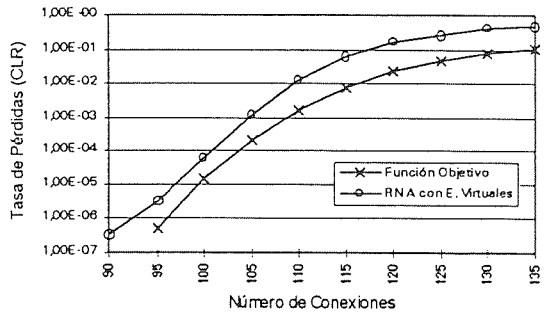


Figura 7. Comparativa entre las curvas de la función objetivo y la aprendida por la RNA (138 conexiones iniciales)

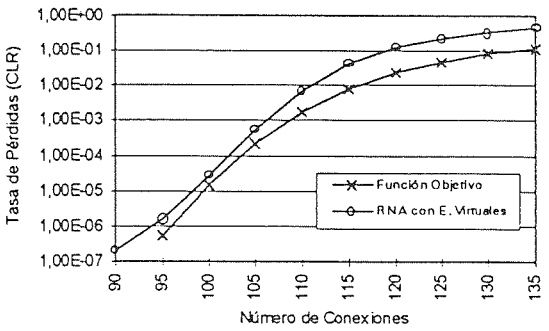


Figura 8. Comparativa entre las curvas de la función objetivo y la aprendida por la RNA (0 conexiones iniciales)

5. Simulaciones con un NCAC en tiempo real

5.1 Tráfico Homogéneo On-Off

Para las simulaciones experimentales con tráfico no controlado, se emplea, inicialmente, una sola clase de tráfico On-Off a 2'5 Mbps de velocidad de pico y tiempos medios en estados On y Off de 10 ms. El tiempo medio de actividad de las llamadas es de 10s., con distribución exponencial. Se considera una tasa de pérdidas objetivo de 10^{-6}.

Las simulaciones se realizaron programando un modelo de nodo ATM -su parte de CAC-, así como modelos de fuentes de tráfico en C++. Se considera un nodo ATM, con un buffer de servicio con capacidad para 200 células, y un ancho de banda de 150 Mbps. De nuevo se utilizan tres enlaces virtuales con anchos de banda de 75, 50 y 37'5 Mbps para obtener los patrones de entrenamiento de la RNA. El intervalo mínimo de estimación se fija en 10.000 células. Al almacenar en memoria un nuevo patrón - en realidad, uno para cada enlace virtual-, se calcula el ABPN correspondiente, y se almacena como patrón con pérdidas nulas. La RNA es entrenada periódicamente con todos los patrones almacenados.

Mediante simulaciones controladas, se evaluó la función CLR vs. número de conexiones, fijándose la frontera de admisión, para la tasa de pérdidas antes indicada, en 96 llamadas.

A continuación se realizaron pruebas con un número de conexiones activas iniciales, por encima y por debajo de la frontera de admisión (138 y 0 conexiones). La carga de tráfico se fijaba

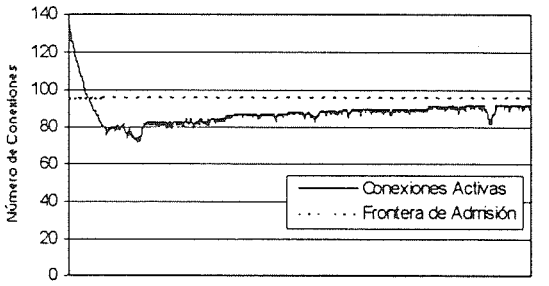


Figura 9. Evolución de las conexiones activas con 138 conexiones iniciales

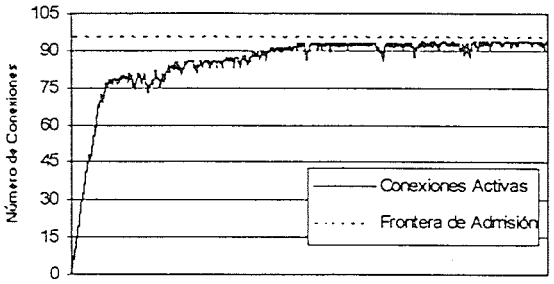


Figura 10. Evolución de las conexiones activas con 0 conexiones iniciales

inicialmente en 110 erlangs, cambiando a 70 al transcurrir un millón de slots de simulación , incrementándose posteriormente en 10 erlangs cada millón de slots, hasta llegar de nuevo a 110. La duración total de las simulaciones fue de 20 millones de slots , equivalente a unos 52 s.

Esta variación en la carga de tráfico tiene por objeto almacenar patrones de entrenamiento correspondientes a diferentes zonas de trabajo, lo cual mejora el entrenamiento de la RNA; si la RNA solo dispone de información alrededor de la frontera de admisión, efectuará correctamente sus chequeos de admisión, pero no identificará tan bien la función de pérdidas. Al finalizar las simulaciones, la frontera de admisión quedó fijada en 92 y 94 conexiones para las respectivas pruebas con 138 y 0 conexiones iniciales. En las figuras 7 y 8, se puede observar como las curvas aprendidas por la RNA son muy cercanas a la función objetivo. En las figuras 9 y 10, se puede ver la evolución de las conexiones activas para ambas pruebas.

Se realizaron pruebas similares en otras condiciones de tráfico, con tráfico homogéneo On-Off. Condiciones menos restrictivas -tasa de pérdidas de 10^{-4} y tamaño de cola 100- no suponen ningún problema, funcionando correctamente. Sin embargo, al repetir las pruebas con una tasa objetivo de 10^{-8}, fue necesario aumentar el tamaño mínimo del intervalo de estimación hasta 25.000 células, para que funcionase correctamente. Esto se debe a que las tasas de pérdidas de los enlaces virtuales son bajan de orden de magnitud , necesitando aumentar el tiempo de estimación para eliminar las imprecisiones en las muestras.

5.2 Tráfico Heterogéneo On-Off

A continuación se estudia la capacidad del método para aprender la frontera de admisión en el caso de fuentes heterogéneas On-Off.

Por una mayor sencillez, se consideran dos clases de tráfico con las mismas características: tasa de pico de 2'5 Mbps, tiempos medios On y Off de 10 ms. El tiempo medio de actividad de la llamada es de 10s. Aunque pueda parecer extraño usar dos clases idénticas, así se evita tener que calcular la tasa de pérdidas de todas las posibles combinaciones de llamadas, cada una de las cuales necesitaría una simulación de muy larga duración -para tráfico de tasa de pico del orden de 1 Mbps, habría que calcular un total de 10.000 combinaciones, para solo dos clases de tráfico- y basta con hacer el cálculo para una de las clases -del orden de 100 combinaciones-. Esto, por otra parte, no afecta para nada a la metodología de la técnica, que es lo que en realidad se quiere probar.

La tasa de pérdidas objetivo es 10^{-4} y el nodo ATM tiene una cola de servicio de capacidad para 100 células, y enlace de salida a 150 Mbps. La frontera de admisión está fijada por todas las combinaciones de llamadas en las que el número total sume 100.

Al tratar con tráfico heterogéneo, se necesita almacenar una diversidad lo mayor posible de patrones correspondientes a distintas zonas de trabajo del espacio de entradas, para beneficiar así la capacidad de generalización de la RNA. En [14] se demuestra la necesidad de dividir la memoria de patrones en subzonas, denominadas "ventanas" que almacenan patrones correspondientes a distintas áreas del espacio de entradas, evitando que el aprendizaje se vuelva local al sustituir patrones de zonas previamente visitadas cuando la memoria se llena.

Para almacenar la mayor variedad posible de patrones, se consideró una carga de tráfico variable para ambas clases. Se realizaron algunas pruebas con las características antes descritas, pero no se obtuvieron resultados muy correctos. La necesidad de almacenar un mayor número de patrones para cada intervalo de estimación válido -uno por cada enlace virtual mas otro adicional para el ABPN- limita la capacidad de almacenar una gran diversidad de patrones distintos, empeorando la capacidad de generalización de la RNA, especialmente para las zonas de extrapolación.

Ante esta situación, se planteó almacenar los patrones de ABPN en una zona distinta de memoria, y sólo uno para cada combinación de entradas, ya que el almacenamiento de varias muestras correspondientes a un mismo estado del

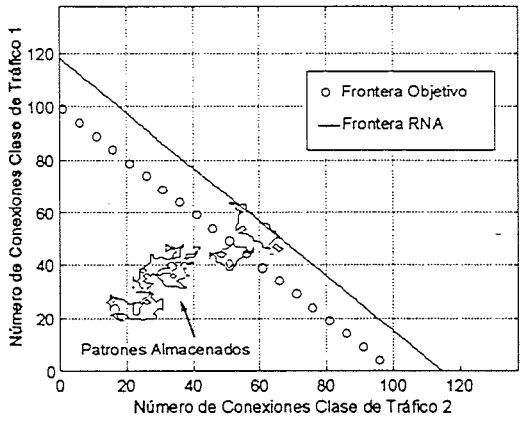


Figura 11. Aprendizaje de la frontera de admisión para 2 clases de tráfico, en malas condiciones de adquisición de patrones.

sistema, que es beneficioso para eliminar imprecisiones en las muestras de enlaces virtuales, no aporta ninguna información útil en el caso de los patrones de ABPN. Con este esquema de almacenamiento, se realizó una prueba en la que la evolución de la carga de tráfico para ambas clases era la siguiente: la clase 1 partía de 10 erlangs, aumentando en 10 cada 10 millones de slots de simulación; la clase 2 partía de 20 erlangs, aumentando en 20 cada 10 millones de slots. La simulación tuvo una duración total de 100 millones de slots. En estas condiciones, la frontera de admisión objetivo no fue aprendida correctamente por la RNA, principalmente debido a que los patrones almacenados se correspondían a zonas alejadas de dicha frontera, mientras que alrededor de ésta, no había apenas información. (Fig. 11). De esta forma, la RNA no fue capaz de generalizar la frontera de admisión al ser muy escasa la información disponible.

Sin embargo, se realizó otra prueba similar en la que la evolución de la carga de tráfico por parte de las dos clases tendía hacia combinaciones de conexiones cercanas a la frontera: la clase 1 partía de 110 erlangs, y asignaba una carga de 100, 85, 75, 60, 50, 40, 30, 20, y 10 erlangs en los instantes de simulación correspondientes a 5, 15 25, 35, 45, 55, 65, 75, y 85 millones de slots, respectivamente; en cuanto a la clase 2, su evolución de la carga de tráfico partía de 10 erlangs para incrementarse en 10 erlangs cada 10 millones de slots, hasta llegar a 100 erlangs. Así, al finalizar la simulación se almacenaron una gran diversidad de patrones en zonas cercanas a la frontera, permitiendo a la RNA generalizar los valores correspondientes a zonas no conocidas, y obteniéndose finalmente una frontera de admisión, aprendida por la RNA, muy cercana a la frontera real (Fig. 12). Se realizaron otras pruebas con condiciones de tráfico similares, cambiando número de conexiones iniciales, semilla, etc.., y en algunos casos se obtuvieron resultados correctos, pero en otros no. Finalmente se llegó a la conclusión de que la frontera de admisión aprendida por la RNA está muy influenciada por los patrones almacenados en memoria. Una buena distribución de patrones consigue casi siempre un aprendizaje correcto, pero no es posible asegurar que esto se cumpla. Además, el uso de una ventana adicional para patrones de ABPN también presentó inconvenientes, ya que a menudo también se llenaba esta ventana, eliminando patrones de ABPN, pero permaneciendo en memoria los patrones de enlaces virtuales correspondientes. Esto podría provocar problemas en el correcto aprendizaje de la RNA como se ha podido ver en el apartado cuatro.

Actualmente se está estudiando en un nuevo método de almacenamiento de patrones en memoria, qué, ocupando un menor espacio, asegure

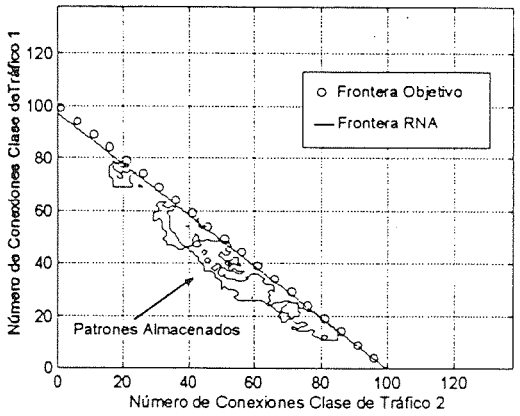


Figura 12. Aprendizaje de la frontera de admisión para 2 clases de tráfico, en buenas condiciones de adquisición de patrones.

la presencia simultánea de muestras de todos los enlaces virtuales y ABPN correspondientes a cada estado del sistema almacenado. También se está estudiando la forma de asegurar que en el peor de los casos de almacenamiento de información -es decir, que no se almacenen patrones de forma distribuida-, al menos se sigan cumpliendo los requerimientos de calidad de servicio, aunque se sacrifique ancho de banda.

6. Conclusiones

Para conseguir un correcto funcionamiento de un CAC Neuronal, se debe proporcionar a la RNA un número suficiente de patrones de entrenamiento, suficientemente precisos en su estimación. La estimación de pérdidas mediante enlaces virtuales se presenta como una técnica interesante para obtener patrones con estas características. Sin embargo, el uso de patrones ABPN es esencial para obtener mejoras importantes en la función de aprendizaje de tasa de pérdidas, como se pudo comprobar para el caso de tráfico homogéneo On-Off. Una aproximación mejorada del valor real de ABPN podría mejorar el método, pero no debe implicar cálculos demasiados complejos si queremos mantener simplicidad y velocidad. Se debe extender esta técnica a otros tipos de tráfico (vídeo, datos,...), así como a condiciones de tráfico heterogéneo, con vistas a desarrollar un futuro control de admisión de llamadas en un nodo ATM basado en este método.

Apéndice I

Modelo matemático de una neurona

En la figura 13 podemos ver el esquema de una neurona artificial.

La actividad interna de la neurona, u, viene determinada por una suma de productos de pesos por sus respectivas entradas.

$$Act.Interna = u = \sum_i x_i w_i \qquad \text{(A1)}$$

donde x_i y w_i son, respectivamente la señal presentada y el peso asociado a la conexión i.

La salida de la neurona viene dada por una función de transferencia de la actividad interna $\Psi(u)$. Existen varias funciones de transferencias posibles; la empleada en este caso es la sigmoide.

$$Sigmoide: y = \Psi(u) = \frac{1}{1+e^{-u}} \qquad \text{(A2)}$$

Perceptrón Multicapa FeedForward

En este tipo de topología las neuronas se agrupan en capas, conectándose con las de las capas adyacentes mediante enlaces unidireccionales, sin que existan realimentaciones (Fig. 14).

En la RNA la neurona j de la capa s hará el siguiente cálculo:

$$O_j^{[s]} = \Psi_j^{[s]}(u_j^{[s]}) = \Psi_j^{[s]}\left(\sum_{i=0}^{n_{s-1}} w_{ji}^{[s]} o_i^{[s-1]}\right) \qquad \text{(A3)}$$

y la ecuación total de la red quedará:

$$\vec{y} = \Psi(x) = \Psi^{[2]}\left(W^{[2]}\Psi^{[1]}\left(W^{[1]}\vec{x}\right)\right) \qquad \text{(A4)}$$

Algoritmo de Aprendizaje Backpropagation

Se basa en minimizar el error local *Ep* de los patrones de entrenamiento:

$$E_p = \frac{1}{2}\sum_{j=1}^{N_N^{[out]}} \left(d_{jp} - y_{jp}\right)^2 \qquad \text{(A5)}$$

con y_{jp} la salida estimada por la neurona j y d_{jp} la salida deseada, para el p-ésimo patrón de entrenamiento.

Cada patrón de entrenamiento consta de $N_N^{[in]}$ señales de entrada x_i y $N_N^{[out]}$ señales de salida dj.

El algoritmo de aprendizaje Backpropagation on-line se realiza en los siguientes pasos:

1. Se inicializan todos los pesos sinápticos $w_{ji}^{[s]}$ a valores aleatorios pequeños.

2. Se presenta un patrón de entrenamiento a la RNA y se calculan las salidas estimadas de todas las neuronas usando el valor actual de $w_{ji}^{[s]}$.

3. Se compara la salida deseada d_j con la estimada por la red y_j y se calculan los errores locales $e_j^{[s]}$ para todas las s capas:

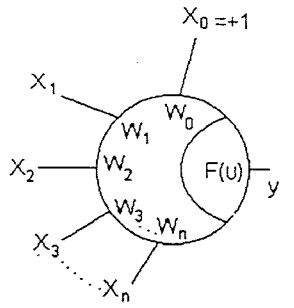


Figura 13.- Neurona Artificial

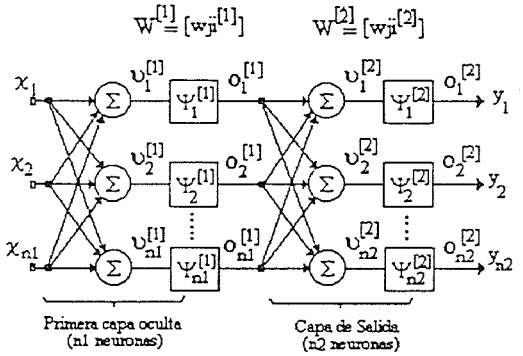


Figura 14. Arquitectura de un PMF de 3 capas

$$e_j^{[out]} = \frac{\partial \Psi_j^{[out]}}{\partial u_j^{[out]}}\left(d_j - y_j\right) \qquad \text{(A6)}$$

$$e_j^{[s]} = \frac{\partial \Psi_j^{[s]}}{\partial u_j^{[s]}} \sum_{i=1}^{N_N^{[s+1]}} e_i^{[s+1]} w_{ij}^{[s+1]} \qquad \text{(A7)}$$

4. Se ajustan los pesos sinápticos con la fórmula iterativa:

$$\Delta_p w_{ji}^{[s]} = \eta\, e_j^{[s]} x_i^{[s]} \quad (s = N_S, \ldots, 2, 1) \qquad \text{(A8)}$$

donde η es la tasa de aprendizaje.

5. Mientras el error *Ep* sea superior a cierto valor objetivo se vuelve al paso 2; en caso contrario se considera terminada la fase de aprendizaje.

Agradecimientos

Este trabajo ha sido parcialmente financiado por la Comisión Interministerial de Ciencia y Tecnología (CICYT), Proyecto Nº. TIC96-0743

Referencias

[1] Nordström, E., Carlström, J., Gällmo O., y Asplund, L. "Neural Networks for Adaptive Traffic Control in ATM Networks". *IEEE Communications Magazine*, **33**, 10, 43-49 (1995).

[2] Onvural, R.O. *Asynchronous Transfer Mode Networks*. Boston: Artech House (1994).

[3] Tse, P.W., y Zuckerman, M. "Connection Admission Control in ATM Networks". *Proc GLOBECOM'94*, **3**, 1790-1794 (1994).

[4] De Veciana, G., Kesidis, G., y Walrand, J. "Resource Management in Wide Area ATM Networks Using Effective Bandwidths". *IEEE Journal on Selected Areas in Communications*, **13**, 6, 1081-1089 (1995).

[5] Castelli, P., Cavallero, E., y Tonietta, A. "Policing and Call Admission problems in ATM Networks" en *Teletraffic and Data Traffic* (Jensen, A.,& Iverseen, J.E., Eds.), ITC-13, Elsevier Sciencie Publishers, 847-852 (1991).

[6] Park, H-S., Kwak, D-Y., Rhee, W-S., Jeon M-Y. y Kim J-K. "Global Traffic Control in ATM Networks". *IEICE Transactions on Communications*, **E78-B**, 4, 476-484 (1995).

[7] Gibbens, R.J., Kelly, F.P. y Key, P.B. "A Decision-Theorretic Approach to Call Admission Control in ATM Networks". *IEEE Journal on Selected Areas in Communications*, **13**, 6, 1101-1113 (1995).

[8] Díaz Estrella, A. "Control Neuronal Multiservicio de Admisión de Conexión para Tráfico ATM". *Tesis Doctoral*. Dpto. de Tecnología Electrónica, Universidad de Málaga. (1995).

[9] Hiramatsu, A. "Training Techniques for Neural Network Applications in ATM". *IEEE Communications Magazine*, **33**, 10, 58-67 (1995).

[10] Díaz Estrella, A., Casilari, E., Jurado, A. y Sandoval, F. "ATM Traffic Neural Control: Multiservice Call Admission and Policing Function" en *Applications of Neural Networks to Telecommunications 2*, (Alspector, J., Goodman, R.,& Brown, T.X., Eds.), Lawrence Elborn Associates Publishers, 104-111 (1995).

[11] Morris, R.J.T., y Samadi, B. "Neural Network for Control of Communications Systems". *IEEE Transactions on Neural Networks*, **5**, 4, 639-643 (1994).

[12] Tarraf, A.A., Habib, I.W., y Saadawi, T.N. "Intelligent Traffic Control for ATM Broadband Networks". *IEEE Communications Magazine*, **33**, 10, 76-82 (1995).

[13] NEUROCOM "Estudio de paradigmas de RNAs para Control Adaptativo". *Report Interno en Proyecto TEMA*. Dpto. Tecnología Electrónica, Universidad de Málaga (1994).

[14] Díaz Estrella, A., Jurado, A., y Sandoval, F. "New Training Pattern Selection Method for ATM Call Admission Neural Control". *Electronic Letters*, **330**, 7, 577-579 (1994).

[15] Tran-Gia, P., y Gropp, O. "Structure and Performance of Neural Nets in Broadband System Admission Control". en *Neural Networks in Telecommunications*, (Yuhas, B.,& Ansari, N., Eds.) Kluwer Academic Publishers, 128-141 (1994).

[16] Courcoubetis, C., Fouskas, G., Friesen, V., y Sartzetakis, S. "Real-Time Issues in Call Acceptance Management for ATM Networks". *Proc. 5th RACE TMN Conference*, Londres (1991).

[17] Roberts, J., Mocci, U., y Virtamo, J., Eds. *Broadband Network Teletraffic. Final Report of Action COST 242*. Berlin: Springer (1996)

[18] Duffield, N.G., Lewis, J.T., O'Conell, N., Russell, R., y Toomey, F. "Entropy of ATM Traffic Streams: A Tool for Estimating QoS Parameters". *IEEE Journal on Slected Areas in Communications*, **13**, 6, 981-990 (1995).

[19] Lewis, P.A.V., y Orav, E.J. *Simulation Methodology for Statisticians, Operation Analysts and Engineers*, 1, WadsworthBrooks/Cole Advanced Books& Software (1989).

[20] Villén-Altamirano, M., y Villén-Altamirano, J. "RESTART: A Method for acelerating Rare Events Simulation" en *Queuing Performance and Control in ATM*, (Cohen, J.W.,&Pack, C.D., Eds.), Elsevier Sciencie Publishers B.V., 71-76 (1991).

[21] Devetsiokis, M., y Townsed, J.K. " On the Efficient Simulation of Large Communication Networks Using Importance Sampling", *Proc. GLOBECOM'92*, **3**, 1455-1459 (1992).

[22] Helvik, B.E., y Heegard, P.E. "A Technique for Measuring Rare Cell Losses in ATM Systems". *Proc. 14th International Teletraffic Congress* (Labetoulle, J.& Roberts, J.W., Eds.), Elsevier Sciencie Publishers B.V. , 917-930 (1994).

[23] Courcoubertis, C., Fouskas, G., y Weber, R. "An On-line Estimation Procedure for Cell-Loss Probabilities in ATM Links". *Proc. 3rd IFIP Workshop on Performance Modelling and Evaluation of ATM Networks*, UK, (1995).

[24] Zhu, H., y Frost, V.S. "In-Service Monitoring for Cell-Loss Quality of Service Violations in ATM Networks". *IEEE/ACM Transactions on Networking*, 4, 2, 240-248 (1996).

[25] Petr, D.W., y Frost, V.S. "Cell Loss Quality of Service in an Integrated ATM Network". en http://www.ittc.ukans.edu/Products/Papers/ijcs_96.ps. Universidad de Kansas (1996).

[26] Herrero, I. "Estudio de las Colas Virtuales como Elemento de Estimación de Pérdidas en un Nodo ATM". *Informe Interno*. Dpto. Tecnología Electrónica, Universidad de Málaga (1996).

[27] Herrero, I., Díaz Estrella, A., y Sandoval, F. "Neural Call Admission Control Through Virtual Links Estimates". *Proc. IEEE ATM'97 Workshop*, 133-141, Lisboa (1997).

[28] Herrero, I. "Estimación de Pérdidas en un CAC Neuronal para la red MTA", *Proyecto Fin de Carrera*. Dpto. de Tecnología Electrónica, Universidad de Málaga. (1995).

Arquitectura neuronal difusa para el control de tráfico en redes ATM

JORGE CUSTODIO, YANNIS DIMITRIADIS, FRANCISCO J. DÍAZ-PERNAS, JUAN LÓPEZ CORONADO
DEPARTAMENTO DE LA TEORÍA DE LA SEÑAL
Y COMUNICACIONES E INGENIERÍA TELEMÁTICA
ESCUELA TÉCNICA SUPERIOR DE INGENIEROS DE TELECOMUNICACIÓN
UNIVERSIDAD DE VALLADOLID
REAL DE BURGOS S/N, 47011 VALLADOLID
Correo electrónico: yannis@tel.uva.es

Abstract:

In this paper, we propose a new neuro-fuzzy system called FasArt for traffic control in ATM networks. Recently, there has been a growing interest in developing communication networks that support services integration multimedia information, such as voice, data and video. The most prominent example of such networks is based on asynchronous transfer mode (ATM). In this sense different ATM services require different Quality of Service (QoS), that might be maintained during a connection. Then, adequate resources have to be reserved when the connection is established. On the other hand, an increase of bandwidth efficiency based on statistical multiplexing can seriously affect the corresponding QoS. Adaptive traffic control through neural networks can provide appropiate characteristic in order to obtain an optimal relation between QoS and bandwidth efficiency. In our case, we have developed a neuro-fuzzy system, FasArt, for traffic control, that combines an automatic generation of fuzzy rules from examples, on-line learning and compliance to the plasticity-stability dilemma. Initial experimental results show satisfactory performance of FasArt in the problem of Connection Admission Control (CAC). Two simulation enviroments were used and evaluated, namely the ATM network simulator by NIST and a generic public domain environment called Ptolemy. The latter has been proved to be more adequate for our problem since it contains a-priori defined elements, it is extensible and open, since it permits the user to add his own objects in C++. Finally, current work and open research subjects are discussed, with relation to the use of FasArt in general ATM traffic control.

1 Introducción

La transmisión de diversos tipos de información (voz, datos, vídeo, ...) necesitaban de distintos tipos de redes que diesen al usuario final una cierta calidad de servicio, a la vez que se pretendía maximizar la utilización de la red en beneficio del operador. Así se habla de una red dedicada para cada tipo de servicio. Con la Red Digital de Servicios Integrados de Banda Ancha (RDSI-BA) aparece el concepto de una única red que permita la transmisión de datos multimedia. ATM (Asyncronous Transfer Mode) o Modo de Transferencia Asíncrona es la técnica elegida para dar soporte a esta red única [1].

ATM se basa en la conmutación rápida de paquetes de longitud fija de 53 octetos denominados células o celdas, pero que es orientado a conexión, es decir, que se crea un circuíto virtual cada vez que se establece una conexión y todo el fiujo de células perteneciente a dicha conexión se encamina por el mismo circuíto.

Al ser las células de longitud fija, junto con el hecho de verse reducido el conjunto de funciones que se deben realizar en los nodos de conmutación, se pueden obtener unas grandes velocidades de procesamiento, y la flexibilidad que requiere la RDSI-BA.

Otra característica importante es que ATM permite el multiplexado estadístico de las diversas conexiones a fín de obtener un alto rendimiento de la red. Este tipo de multiplexado consiste en aprovechar la naturaleza estadística en la transmisión de las fuentes a la hora de reservar los recursos necesarios. De esta manera una fuente puede transmitir mientras otra puede estar en un estado de reposo. Sin embargo, el inconveniente que presenta esta técnica es que se pueden producir periodos en el que diversas fuentes transmitan a la vez, perdiéndose excesivas células en las colas de los conmutadores si no se han reservado los recursos adecuados al crear el circuíto virtual.

En la próxima sección se describe el problema del control de tráfico en una red ATM, y cuales son las acciones que realiza la red en orden de realizar este control. En la sección 3 se hace un repaso de soluciones propuestas encontradas en la literatura especialmente aquellas en la que se utilizan tanto redes neuronales como sistemas de lógica borrosa. Posteriormente presentamos el FasArt, la red neuro-

difusa que utilizamos para realizar el control. En la sección 5 presentamos la aplicación de la red neuro-difusa en una de las tareas de control, para finalizar con una sección de discusiones y otra de conclusiones.

2 Control de tráfico

2.1 Especificación del contrato de tráfico

Las diversas fuentes tienen distintas características y requisitos de calidad de servicio. Así por ejemplo, una fuente de voz necesita que las células lleguen a su destino en un tiempo máximo y que la varianza del intervalo entre células sea lo más pequeño posible de tal forma que el receptor pueda reproducir la señal de voz sin distorsiones. Sin embargo, la red puede permitirse perder alguna que otra célula sin que la calidad obtenida en el receptor disminuya significativamente. Por el contrario, la transferencia de un archivo entre dos ordenadores debe realizarse respetándose la integridad de la totalidad de los datos, mientras que el retardo en la transmisión, o la variabilidad en el mismo no es preocupante. Aparecen pues dos conceptos [1]:

Transparencia semántica : Determina la posibilidad de la red para transportar la información libre de error.

Transparencia temporal : Determina la capacidad de la red para transportar la información en un tiempo dado.

Una fuente puede especificar sus requisitos de calidad de servicio mediante una serie de parámetros QoS (*Quality of Service*), a saber:

- Porcentaje de células perdidas o CLR (Cell Loss Ratio).
- Retardo de transmisión de células o CTD (Cell Tranfer Delay).
- Variación en el retardo de transmisón o CDV (Cell Delay Variation).

Una descripción detallada de estos parámetros QoS se puede encontrar en [2] y [3].

Como se dijo anteriormente, ATM es orientado a conexión, de manera que es al inicio de la misma, cuando se reservan los recursos necesarios. Por eso, las fuentes además de especificar sus requisitos de calidad de servicio, tienen que indicar también cuál será su comportamiento. Para ello existen una serie de parámetros de tráfico que son:

- Tasa de pico de células o PCR (Peak Cell Rate).
- Tasa media de células o SCR (Sustainable Cell Rate).
- Tamaño máximo de ráfaga o MBS (Maximum Burst Size).
- Tasa mínima de células o MCR (Minimum Cell Rate).

Cada fuente debe especificar su comportamiento mediante una elección adecuada de valores para un conjunto de estos parámetros, en lo que se denomina el descriptor de tráfico de la fuente. Este descriptor, junto con unas tolerancias y una definición de conformidad que indica sin ambiguedades qué células son conformes a la conexión, forman lo que es el descriptor de tráfico para la conexión.

A la hora de establecer la conexión, el usuario y la red firman lo que se denomina el contrato de tráfico de la conexión, en el que se incluye el descriptor de tráfico de la conexión y los requisitos QoS. Mediante este contrato, la red promete ofrecer la calidad de servicio contratada, siempre y cuando el usuario no viole el contrato (por ejemplo, utilizando un ancho de banda mayor al especificado).

2.2 Funciones de control de tráfico

El control de tráfico y el control de congestión es el conjunto de acciones que toma la red para evitar la congestión o minimizar los efectos de ésta una vez que se haya producido [4]. Se puede hablar entonces de control preventivo y control reactivo. Dentro del control de tráfico destacar las funciones siguientes:

- Control de Admisión de Llamada o CAC (Connection Admission Control).
- Control de Parámetros de Uso o UPC (Usage Parameter Control).
- Descarte selectivo de células.
- Moldeado de tráfico.

Tanto en [4] como en [2] se pueden encontrar una descripción detallada de estas y otras funciones de control de tráfico y de congestión así como nuevas áreas de investigación.

La función de CAC se define como el conjunto de acciones que toma la red en el establecimiento de una conexión (o la renegociación de la misma) para determinar si dicha conexión debe progresar o si por el contrario debe ser rechazada. La nueva llamada debe ser aceptada sólamente si la red puede ofrecer la QoS contratada y no disminuye la calidad de servicio para el resto de conexiones previamente establecidas. Aparte de esta decisión, otras acciones que debe realizar el CAC es la de enrutamiento y la reserva de recursos.

El UPC es una función de vigilancia del tráfico generado por los usuarios, cuyo propósito principal es proteger a la red del uso erróneo o malintencionado de los recursos, que puede afectar seriamente a la QoS de otros usuarios. Esta función se lleva

a cabo detectando qué células violan el contrato de tráfico y tomando una acción apropiada sobre ellas que puede ser: descartar la célula o marcarla para su posterior descarte en algún nodo congestionado.

El descarte selectivo de células consiste en rechazar en los nodos congestionados las células que pertenezcan a conexiones no conformes con el contrato declarado y aquellas células que tengan menor prioridad (bit CLP=1 en la cabecera de la célula). Esto se hace para proteger lo máximo posible el flujo de células de mayor prioridad. Para conexiones conformes con la conexión, los objetivos de CLR se deben cumplir aún utilizando esta función.

El moldeado sirve para suavizar las características del tráfico generado por las fuentes, por ejemplo, reduciendo la tasa de pico durante las ráfagas, para así conseguir una mayor eficiencia de red. En muchas ocasiones esta función se realiza conjuntamente con el UPC.

3 Trabajo relacionado

La caracterización exacta del tráfico generado por las fuentes es crucial para el correcto funcionamiento de las funciones de control. Utilizar tan sólo los parámetros de tráfico no es suficiente para realizar esta caracterización. Sería mucho mejor disponer además de los momentos estadísticos de primer y segundo orden, y mejor aún, disponer del conocimiento de la función densidad de probabilidad. Sin embargo, esto es difícil ya que se requieren operaciones matemáticas muy complejas y por lo tanto grandes velocidades de computación, lo cual no es factible en un entorno donde las velocidades de transmisión ya han superado los cientos de megabits por segundo. Por otro lado, usar modelos simples (estadísticos o deterministas) pueden no recoger bien las características de correlación de un tráfico a ráfagas. Mas aún, algunos tipos de tráfico multimedia (por ejemplo, vídeo de tasa binaria variable) tienen características que no se conocen bien, y futuras aplicaciones tendrán otras caracteristicas que pueden ser distintas a las actuálmente conocidas.

La función CAC es una de las más importantes en el control de tráfico, ya que el objetivo final es maximizar el rendimiento de la red y así maximizar los beneficios del operador. Se puede usar el PCR para reservar los recursos necesarios a la hora de establecer una conexión, sin embargo, para fuentes que transmiten a ráfagas, se desperdiciaría un gran ancho de banda que podría ser usado por otras conexiones con semejantes características. También, se podría usar la tasa media de la fuente para reservar los recursos. En este caso se reservan menos recursos para cada llamada, y se permiten más conexiones activas en un periodo de tiempo, pero la QoS disminuiría drásticamente en los intervalos donde al menos dos fuentes transmitieran a la tasa de pico. Se puede usar el concepto de ancho de banda equivalente [5], pero este método sólo es exacto a medida que el tamaño de la cola se aproxima a infinito y el CLR se aproxima a cero. Además se ha demostrado que es muy ineficiente comparado con otros métodos que usan redes neuronales (p.e. en [6]).

Otra aproximación, es utizar las redes neuronales como medio de estimar las características complejas de las fuentes por medio de la observación de los patrones de tráfico. Un modelo de control de tráfico basado en redes neuronales no requiere un conocimiento exacto del entorno, por lo que es robusto. Además es flexible, ya que puede aprender nuevos patrones de tráfico de las aplicaciones futuras. Finalmente debido a su alto grado de paralelismo, una implementación hardware sería lo suficientemente rápida como para operar en un nodo ATM.

Desde el comienzo de la década de los noventa, la aplicación de redes neuronales en el control de tráfico ATM ha sido un punto de creciente interés [7].

En [8] se propone usar un perceptrón multicapa (MLP) para hacer corresponder un vector de estado de la cola del conmutador con la calidad de servicio estimada por un modelo matemático llamado Modelo de flujo de fluidos heterogéneos para fuentes on/off. Este modelo matemático estima la probabilidad de pérdida de células en la cola resolviendo una serie de ecuaciones diferenciales de primer orden. La complejidad de este problema incrementa geométricamente a medida que aumenta el número de clases de fuentes, haciendo que una implementación en tiempo real sea imposible. Utilizando el perceptrón multicapa, se aprende mediante el algoritmo *backpropagation* la relación existente entre el vector de estado de la cola, y la probabilidad de pérdida de células estimada por el modelo matemático. Esta aproximación permitiría realizar el control en tiempo real. Sin embargo, esta solución presenta varios inconvenientes:

1. El modelo propuesto requiere que el aprendizaje sea fuera de línea (*off-line*), por lo que si aparece un nuevo tipo de fuente con otras características, la red tendría que ser entrenada de nuevo en su totalidad. Esto es debido al problema conocido como dilema estabilidad-plasticidad, consistente en cómo hacer que la red aprenda patrones nuevos, sin olvidar lo anteriormente aprendido. Este problema está presente en todas las redes que utilizan el *backpropagation* como algoritmo de aprendizaje.

2. El modelo matemático sirve para tamaños de cola lo suficientemente grandes y para fuentes on/off. Si no se cumplen estos requisitos, el modelo no estimará bien la probabilidad de pérdida de células, por lo que el control no será eficiente.

Otro esquema de control es el propuesto por Hiramatsu en [9]. El modelo está basado en la clasificación de las llamadas en una de las clases definidas. Todas las fuentes pertenecientes a la misma clase tienen las mismas características de tráfico (como el PCR, o el SCR). La red neuronal aprende mediante *backpropagation* la relación entre un vector cuyas componentes son el número de llamadas activas para cada clase, y el CLR observado en la cola de salida del nodo ATM. De esta manera se aprende indirectamente cual es el número máximo de llamadas que permiten obtener un porcentaje de pérdidas de células inferior a un objetivo dado. Sin embargo, este esquema presenta unos inconvenientes:

1. Como se vió anteriormente el algoritmo de aprendizaje *backpropagation* no permite un control adaptativo frente a nuevas situaciones de tráfico, ya que nuevos patrones de tráfico requieren un entrenamiento fuera de linea junto con los patrones anteriormente observados. La solución propuesta por Hiramatsu es usar una tabla de patrones donde se almacenan el vector de entrada a la red neuronal (número de conexiones activas) junto con el CLR observado para esa situación de tráfico. De manera que el aprendizaje se puede realizar en linea presentandole a la red de forma periódica patrones almacenados en esta tabla. Esta solución también es utilizada por A. D. Estrella y col. en [10].

2. La red neuronal no aprende cuál es el límite para el número de llamadas n_{lim}, hasta que se ha sobrepasado este límite. Esto es debido a que mientras el número de llamadas es inferior a n_{lim}, el CLR observado es próximo a cero, y únicamente cuando se sobrepasa el límite, el CLR crece rápidamente, y la red comienza a aprender la curva. La solución propuesta es lo que Hiramatsu denomina el método de colas de salida virtuales, en el que se emplean colas de menor capacidad para aprender una superficie que mediante una extrapolación permite calcular cuál es n_{lim} antes de llegar a la congestión.

Otra aproximación para realizar el CAC es la que se muestra en [6]. En ella se presenta un esquema de control que es independiente del tipo de fuentes y del número de clases permitiendo un control de tráfico para fuentes muy heterogéneas. Esto es debido a que se utiliza un proceso de cuenta como entrada a la red neuronal. Las entradas son el número de células que se han contado en intervalos geométricamente crecientes, y la salida de la red es el CLR. El problema que tiene este método es que el aprendizaje se debe realizar fuera de linea, por lo que tampoco es adaptativo.

También se han aplicado redes neuronales para la función UPC del control de tráfico. En [11] se usan dos redes neuronales *backpropagation* para predecir el comportamiento de la fuente en instantes futuros. La red aprende cuál será el número de células que llegarán en un periodo de tiempo posterior a partir del número de células que llegan en periodos anteriores, y por tanto, de una manera indirecta, la función densidad de probabilidad del tráfico generado por la fuente. La primera red se entrena con patrones de tráfico generado por fuentes que son conformes con los parámetros de tráfico declarados. La segunda red se entrena con los patrones de tráfico que se obtienen de un entorno real, donde existen comportamientos correctos y violaciones del contrato. La diferencia entre las salidas de ambas redes se toma como una señal de error que si supera un umbral fijado, descarta las las células siguientes y por lo tanto no entrarán en la red. Este esquema presenta presenta los siguientes inconvenientes:

1. El umbral que determina qué células no cumplen con el contrato es fijo y no está claro qué valor se debe elegir para realizar el control.

2. La acción que se realiza sobre las células que se han identificadas como no conformes al contrato de tráfico es exclusivamente rechazarlas, cuando un control que permitiese marcarlas sería mas eficiente.

3. Aunque se promete que la segunda red puede seguir aprendiendo, una clase de fuente distinta de aquellas con las que se ha entrenado, puede destrozar el aprendizaje anterior, por lo que este esquema tampoco es adaptativo.

La lógica difusa, por otra parte, ha experimentado un gran desarrollo en estos últimos años, y su utilización en problemas de control ha resultado muy satisfactoria, especialmente en problemas en los que es difícil, si no imposible, encontrar un modelo matemático que describa el sistema, o en problemas donde no se conocen las reglas de inferencia que gobiernan el control o donde las entradas de control sean imprecisas o tengan un grado de incertidumbre.

En [12] se describe un esquema de control donde se utiliza la lógica difusa para implementar un CAC difuso y un controlador de congestión difuso. El controlador de congestión usa tres variables linguísticas para determinar el grado de congestión del nodo ATM. Estas entradas son el tamaño de la cola, la variación del tamaño de la cola y el CLR observado en la cola. El controlador de congestión determina qué acción se debe realizar sobre las fuentes. Las acciones son los conjuntos difusos "incrementar mucho", "incrementar un poco", "no cambiar", "decrementar un poco" y "decrementar mucho". En principio los conjuntos difusos y las reglas se determinan según una aproximación intuitiva, pero se

pueden optimizar mediante algoritmos genéticos. El controlador CAC se implementa siguiendo la misma filosofía. Las entradas para este controlador son el CLR, la salida del controlador de congestión, y otra variable difusa que indica cual sería el ancho de banda necesitado por la nueva conexión. La salida del controlador de CAC es una decisión difusa de aceptar o rechazar la nueva llamada.

Para realizar la función UPC también se han propuesto varias alternativas usando lógica difusa. En [13] se propone una versión difusa del *Leaky Bucket* que es un algoritmo muy conocido para identificar qué células son conformes con el tráfico y cuales no. En este algoritmo, se generan testigos a la tasa de pico declarada por la fuente. Las células que obtengan uno de estos testigos pueden entrar en la red, mientras que las que no obtengan testigos son descartadas. Las células con testigo, se transmiten a la red con una tasa igual a la media declarada por la fuente. El tamaño de una ráfaga se limita con el número máximo de testigos que se pueden generar. La versión difusa de este algoritmo, permite generar además otra clase de testigos de manera que la células que los obtengan son marcadas bajando la prioridad. El número generado de estos testigos depende del estado de congestión del nodo. Las entradas al controlador son la QoS observada en el buffer y el número de testigos existente. La salida es el número de testigos a generar.

En [14] se propone otro controlador difuso para realizar la función del UPC. El modelo propuesto se basa en un sistema de ventanas. En una ventana se mide el número de células que llegan durante la duración de esa ventana. Si este número es menor que el valor esperado se incrementa el tamaño de la ventana, proporcionando así un crédito a la fuente, ya que su comportamiento no viola el contrato firmado. Si por el contrario, el número de células que se generan durante el periodo que dura la ventana es superior al esperado, el tamaño de la misma se disminuye, ya que la fuente puede estar violando el contrato, por lo que se requiere una vigilancia más intensiva. Las entradas del controlador son el número de células generadas por la fuente desde el inicio de la conexión, el número de células en la ventana, y el tamaño de la ventana. Todas estas entradas son variables difusas. La salida del controlador es el incremento (o decremento) del tamaño de la ventana.

4 La red neurodifusa FasArt

FasArt [15],[16] está basado en la arquitectura Fuzzy ARTMAP [17] y pretende manteniendo las principales características de este modelo solventar ciertos aspectos del mismo que no consideramos satisfactorios. Fuzzy ARTMAP es una red cuya principal aplicación es la de clasificación mediante aprendizaje supervisado. En la presente sección nos planteamos las limitaciones de Fuzzy ARTMAP para así llegar a nuestra propuesta que intenta mantener las principales características del Fuzzy ARTMAP, especialmente el concepto de máxima generalización con mínimo error de predicción y el principio de estabilidad-plasticidad. Como veremos, FasArt se puede interpretar mediante dos puntos de vista: o bien como una red neuronal o como un sistema lógico difuso.

4.1 Fuzzy ARTMAP

La arquitectura Fuzzy ARTMAP está compuesta por dos módulos Fuzzy ART (Fuzzy ART-a y Fuzzy ART-b) relacionados entre sí por un mapa Inter ART. Los módulos Fuzzy ART son sistemas autoorganizativos, en los que patrones semejantes se clasifican bajo una misma categoría o *cluster*. La semejanza entre un patrón y el prototipo de la categoría viene dado por el operador difuso AND, de manera que si esta semejanza supera un umbral dado por un factor de vigilancia ρ, el patrón se clasifica bajo la misma categoría que el prototipo. Si no existe ninguna categoría bajo la cual se pueda englobar el patrón dado, se crea una nueva. El mapa Inter ART relaciona las categorías de un móludo con las del otro.

El funcionamiento del sistema tiene dos fases:

- Fase supervisada, en la cual se presenta la entrada al Fuzzy ART- a, y la salida deseada al Fuzzy ART-b, y se deja que la red aprenda el emparejamiento entre categorías.

- Fase no supervisada, en la que a la red A se le presenta el vector de entrada, y esperamos obtener una salida en la red B coherente con el aprendizaje realizado.

Como se puede ver, la utilización de Fuzzy ARTMAP como identificador de funciones está limitada por la idea de la clasificación. Una forma rápida de ver esta limitación, es considerar el ejemplo en el que se quiere aproximar una función $f(x)$ contínua, utilizando esta arquitectura. La salida de la red tendrá el aspecto de la salida de un cuantificador donde el número de niveles será igual al número de clases. Si se requiere una buena aproximación, se necesitará una gran cantidad de nodos, aumentando la complejidad de la red. Si por el contrario, es el tamaño de la red lo que está limitado, la función estimada tendrá pocos niveles, aumentando el tamaño del salto. Además el caracter difuso o discreto *crisp* de las categorías no está definido al no estar definida una función de pertenencia para cada una de las categorías. Esta característica sería deseable tanto para la identificación de funciones, como para la clasificación de patrones. En la identificación de funciones, la salida se suavizaría reduciendo la altura de los escalones en el ejemplo anterior. Para la clasificación de patrones, también puede ser útil

tener información acerca del grado de confianza con la que un patrón ha sido clasificado bajo una clase.

Podemos decir por tanto, que la arquitectura Fuzzy ARTMAP no es propiamente una arquitectura difusa, ya que a pesar de utilizar operaciones asociadas con la lógica difusa, no se definen las categorías como conjuntos difusos, y no se da una función de pertenencia asociada a cada uno de estos posibles conjuntos. Además no se dispone de ningún mecanismo que nos dé el grado de acoplamiento entre las clases aprendidas durante la fase de aprendizaje y las entradas de la fase de test. Estas razones son las que nos motivan a desarrollar el modelo FasArt, basado en la arquitectura vista, manteniendo sus principales características, pero incluyendo estas nuevas apropiedades.

4.2 FasArt

A la hora de diseñar el FasArt (*Fuzzy adaptive system ART based*) buscamos una doble interpretación como sistema neuronal y como sistema difuso. Esto nos permite construir un sistema basado en reglas de conjuntos difusos, en el que no necesitamos de un experto que nos diga cuales son estos conjuntos, ya que éstos se generan automáticamente. Como en Fuzzy ARTMAP, mantenemos los dos clasificadores ART, que en nuestro caso generan autoorganizativamente conjuntos difusos. El mapa Inter ART lo podemos ver como una memoria asociativa que almacena las reglas que relacionan los conjuntos antecedentes (obtenidos del ART-a) y los consecuentes (del ART-b). Además se mantiene el mecanismo de RESET de la arquitectura ARTMAP con lo que se mantiene las características de máxima generalización y mínimo error de predicción. El dilema estabilidad-plasticidad también queda resuelto con ARTMAP, ya que nos permite aprender patrones nuevos, creando nuevas categorías con lo que no se pierde el aprendizaje anterior. También el paradigma ART permite reducir el tiempo de aprendizaje al realizarlo por "trozos" y no necesitar todo el conjunto de patrones de entrenamiento tal como puede suceder con el *backpropagation*, ya que el ajuste de los pesos se hace tan solo para las categorías implicadas. Todo esto hace de FasArt una arquitectura que elimina el concepto de "caja negra" de las redes neuronales ya que permite un seguimiento de las reglas y los conjuntos aprendidos, permitiendo un aprendizaje incremental de manera que podemos usarlo *on-line* en sistemas de control en tiempo real, tanto para clasificación como para identificación de funciones ya sean de carácter contínuo o discreto.

FasArt está basado en una nueva forma de calcular la activación de las unidades de la capa competitiva de cada modulo ART. Esta nueva función tiene la forma de la función de pertenencia del producto cartesiano de conjuntos difusos definidos en el espacio de las diferentes componentes del vector de entrada:

$$T_k = \eta R_k = \prod_{i=1}^{M} \eta R_{ki}(I_i)$$

donde T_k es el valor de activación de la unidad k de la capa competitiva. M es la dimensión del vector de entrada I y R_k representa el conjunto asociado a la categoría k. La función de pertenencia ηR_k tiene la forma siguiente:

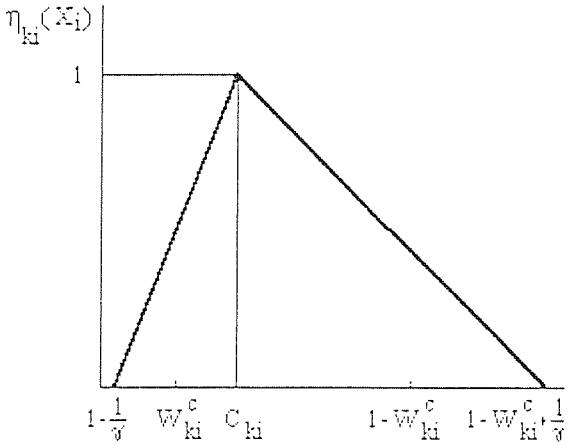


Figura 1: Función de activación-pertenencia para cada nodo

La función de pertenencia que hemos definido viene dada por los pesos asociados a cada nodo y por un nuevo vector de pesos asociado a cada unidad $C_k = (c_{ki})$, y por un parámetro de diseño γ que nos determina el caracter difuso o discreto del conjunto difuso. Valores $\gamma \to 0$ hacen que los conjuntos sean más difusos, mientras que si $\gamma \to \infty$ los conjuntos serán más discretos, reduciéndose por tanto la generalización.

El objetivo principal del mapa Inter ART es reflejar las relaciones entre las categorías de ambos módulos Fuzzy ART. Un valor del peso del mapa Inter ART $w^{ab}_{ij} = 1$, nos indica una regla del tipo

$$\text{IF } I^a \text{ IS } R_i \text{ THEN } Y^b \text{ IS } R_j$$

En el modelo Fuzzy ARTMAP el mecanismo de Reset Inter ART hace que se desechen las categorías del Fuzzy ART-a, manteniendose la clasificación del módulo Fuzzy ART-b. Como vemos, el módulo Fuzzy ART-b lo estamos usando como un registro donde se almacenan las descripciones de las diferentes clases. Sin embargo, este mecanismo para problemas de identificación de funciones no se comporta bien. En el modelo FasArt, cuando se detecta una situación de contradicción entre el aprendizaje realizado hasta el momento y la nueva entrada supervisada, el Reset Inter ART eleva los parámetros de vigilancia para cada módulo Fuzzy ART.

Una vez realizada la fase de aprendizaje, en la fase de test nos podemos plantear diversos modos de funcionamiento según el punto de vista que consideremos. Si utilizamos el FasArt como una arquitectura neuronal del tipo ARTMAP, la salida será la

activación de un único nodo de la capa F2 cuyos pesos asociados nos determina la clase de salida. Sin embargo, según este punto de vista estamos desperdiciando la información adicional que nos proporciona el FasArt sobre el grado de pertenencia a un conjunto difuso. Podemos por tanto, calcular un vector de salida

$$Y^b = (y_1^b, \ldots, y_{N^b}^b)$$

con $y_j^b = max_i(T_i^a w_{ij}^{ab})$.

Este vector, nos da el grado de pertenencia con la que la entrada se clasifica bajo cada categoría.

Finalmente, podemos usar el FasArt como un sistema lógico difuso. Podemos calcular un valor de salida con un método de defuzificación utilizando el vector definido anteriormente como la evaluación de los antecedentes de las reglas.

Tenemos que destacar varias propiedades del FasArt que pueden ser aprovechables en nuestra tarea de control de tráfico ATM:

- Debido al carácter del aprendizaje basado en emparejamiento y el consiguiente cumplimiento del dilema plasticidad-estabilidad, FasArt es capaz de seguir aprendiendo durante la fase de test [16]. Así, se pueden ir aprendiendo situaciones no aparecidas en la fase *off-line* de aprendizaje, situación muy importante en nuestro problema. Esta propiedad contrasta con la incapacidad de las redes basadas en el algoritmo *Backpropagation*, que buscan un mínimo global del error.

- Su definición como sistema lógico difuso permite su implantación en arquitectura hardware difusas sencillas y rápidas. Esta misma formulación garantiza teóricamente la capacidad de FasArt de aproximar cualquier función.

5 Trabajo experimental

Para comprobar cuál es el comportamiento del FasArt en el control de tráfico tenemos que recurrir a la simulación. Consideramos tres opciones:

1. Utilizar un simulador de ATM de los existentes, en cuyo caso tenemos que tener acceso al código fuente del simulador si queremos incorporar nuestro controlador a la simulación. El problema es que no existen apenas simuladores comerciales, ni de libre distribución que nos permitan modificar el código, por lo que la elección se vió reducida al simulador desarrollado por el NIST [18], que aunque completo, está diseñado para analizar el rendimiento de redes y a la hora de realizar el control de tráfico presenta muchos inconvenientes como se verá en la sección 5.1.

2. Utilizar una herramienta de desarrollo de software que nos permita crear simulaciones a medida. Un ejemplo de este tipo de herramientas es el Ptolemy [19] que nos permite crear objetos en C++. En esta herramienta se incluyen diversos dominios de simulación entre los que existe uno de eventos discretos en el que los objetos consumen y generan eventos que pasan a otros objetos, lo cual hace de Ptolemy una herramienta adecuada para nuestros propósitos. El inconveniente de esta opción es el alto coste de desarrollo software, ya que aunque se incluyen algunos objetos que se pueden utilizar (colas, relojes, muestreadores, etc.) se tienen que modelar los distintos tipos de fuentes, conmutadores, etc. En la seccion 5.2 veremos el trabajo que actualmente estamos realizando con esta herramienta.

3. Crear una aplicación completa propietaria que cumpla con nuestros requisitos, pero esto tiene el inconveniente de la baja portabilidad, reutilización del código y el tiempo de desarrollo software.

5.1 Simulador ATM del NIST

Inicialmente probamos el simulador de ATM desarrollado por el NIST. Esta es una herramienta realizada en lenguaje C y X-Window que puede correr bajo plataformas UNIX. El programa ofrece una interfaz gráfica que permite crear distintas topologías de red interconectando los diversos elementos de red [20]. Los elementos que se pueden utilizar son:

- Fuentes CBR, VBR, ABR.
- Terminadores de red.
- Enlaces.
- Conmutadores.

Todos estos elementos tienen una gran cantidad de parámetros que se pueden determinar al inicio de la simulación, por lo que se tiene un gran grado de libertad a la hora de hacer las simulaciones.

Sin embargo, a medida que avanzamos en la utilización del simulador y la incorporación de un mecanismo de control, descubrimos una serie de limitaciones:

- No existe ningún mecanismo para realizar peticiones de conexión.
- Las rutas entre las distintas fuentes origen y las destino se deben especificar antes de realizar la simulación, por lo que el simulador no puede realizar enrutamiento.
- Una fuente sólo puede generar una llamada durante la simulación.

Tabla 1: Valores de los parámetros para las distintas clases de fuentes

Fuente	Clase 1	Clase 2	Clase 3
Tasa binaria ($Mb/seg.$)	45-60	150-200	500-700
Duración media de ráfaga ($\mu seg.$)	20-35	40-75	65-100
Interv. medio entre ráfagas ($\mu seg.$)	100-200	300-500	600-800
Mbits a transmitir	0.1-0.25	0.25-0.4	0.4-0.5
Interv. medio entre llamadas ($\mu seg.$)	70-120	60-170	70- 130

- El simulador no realiza medidas sobre la semántica temporal.
- No es posible especificar parámetros de calidad de servicio.

Se han realizado una serie de modificaciones en el código fuente que nos permiten que cuando una fuente termine pueda volver a transmitir transcurrido un intervalo de tiempo aleatorio según una distribución exponencial. Además se ha incluido un mecanismo de CAC [21].

El esquema de simulación propuesto consta de fuentes VBR de tres clases cuyos parámetros se pueden ver en la tabla 1.

El conjunto de las fuentes están conectadas a un conmutador que tiene una capacidad de salida de $160Mb/seg$. En Fig. 2 se muestra cuál es el diagrama de bloques para el esquema de control propuesto.

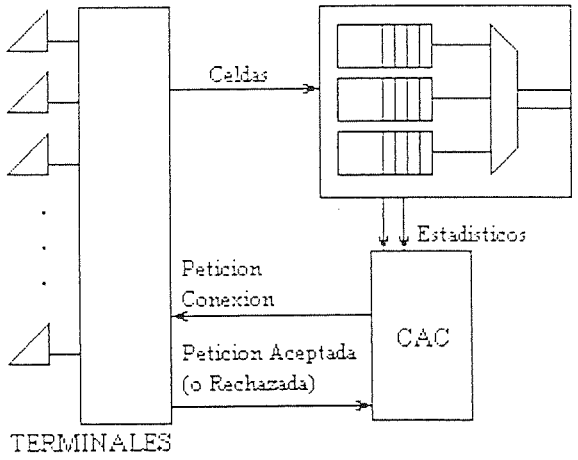


Figura 2: Diagrama de control

El FasArt realiza la función principal de control que es estimar la calidad de servicio. Como el simulador del NIST no permite especificar los parámetros de calidad de servicio, se ha utilizado el tamaño de la cola para indicar esta calidad, de manera que si controlamos este parámetro podemos tener controladas las situaciones en las que se produce un desbordamiento de la cola. Para entrenar el FasArt utilizamos como entrada un vector $\bar{b}$

$$\bar{b} = (b_1(t), b_1(t-1), b_2(t), b_2(t-1), b_3(t), b_3(t-1))$$

donde $b_i(t)$ es la tasa media de llegada (normalizada por la tasa máxima) para la cola de la clase de fuente i en el intervalo de monitorización t. El intervalo de monitorización elegido en nuestros experimentos fue de $100\mu seg$. Como entrada supervisada se utiliza un valor que nos indica si se ha sobrepasado el 80% de la capacidad de la cola en el intervalo de monitorización actual o no. Este valor es 1 si se ha sobrepasado y 0 en caso contrario.

El controlador consta de tres fases:

- Durante la fase de observación se obtienen las tasas de llegada para cada una de las colas en el intervalo de monitorización.
- El entrenamiento consiste en modificar los pesos de la red mediante aprendizaje supervisado. Este entrenamiento se tuvo que realizar off-line debido a que sólo se pueden obtener los estadísticos de la red a través de un fichero, en el que el simulador del NIST vuelca todos los valores de las variables que se pueden observar.
- La fase de control se activa cada vez que se recibe una petición de conexión, en cuyo caso se suma la tasa media de la fuente a la tasa media observada en el intervalo de monitorización actual. Si la red predice que el tamaño de la cola va a superar el 80%, entonces la llamada no se acepta.

En nuestros experimentos, el periodo de aprendizaje consta de $25000\mu segs$. En Fig. 3 se muestra cual es la ocupacion de la cola del conmutador durante $12000\mu seg$. si no se realiza el control. En este periodo será cuando se realice el test. Durante este tiempo, se produjeron 41 peticiones de conexión. Como se puede ver, sin el control existen 7 zonas en las que el tamaño de la cola alcanza el valor máximo. A efectos de control, el primer intervalo donde se produce la congestión, no se tendrá en cuenta debido a cuestiones de inicialización del simulador. Sin contar pues este periodo, se identificaron 7 llamadas como no aceptables.

En Fig. 4 se puede ver cómo aplicando el control se rechazaron correctamente estas 7 llamadas, eliminando el desbordamiento de la cola y por lo tanto la pérdida de células, mientras que la utilización de la línea es plena tal y como muestra el hecho de que la cola casi nunca se vacía.

Una vez vistos estos resultados, y profundizando más en el estudio del control de tráfico, nos decidimos a experimentar con otra herramienta que nos permitiese desarrollar un entorno de simulación que estuviese más orientado al control que el que nos permitía este simulador.

5.2 Entorno de simulación Ptolemy

Ptolemy proporciona una infraestructura de software (X, Tcl/Tk) que corre bajo diferentes plataformas UNIX, sobre la cual se pueden construir diversos entornos de simulación. La infraestructura o núcleo consta de diversos objetos de una familia de definiciones de clases en C++, sobre las que se construyen los diferentes dominios [22]. En concreto el dominio que se adapta más a nuestros propósitos es el de eventos discretos, sobre el cual nosotros podemos crear nuestros propios objetos, o utilizar los que ya vienen en la librería de objetos para este dominio. Además, los dominios pueden interactuar entre ellos debido a la tecnología orientada a objetos que permite que cada dominio funcione sin tener conocimiento alguno de las características de los otros. Esto permite por ejemplo, transmitir células con información de vídeo por una red ATM usando el dominio de eventos discretos y realizar un procesado de la imagen en el terminal destino usando el dominio de procesado de la señal.

El trabajo actual consiste en ver qué capacidad tiene la red FasArt para predecir la congestión en un nodo. Para ello utilizamos fuentes de un mismo tipo pero con características cambiantes de una petición de conexión a otra. Las fuentes están modeladas

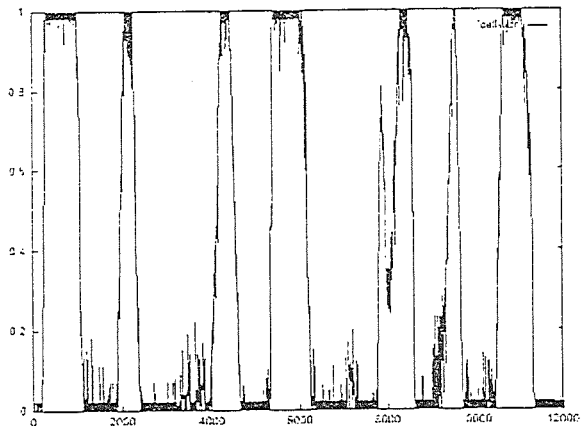


Figura 3: Evolución del tamaño de la cola del conmutador sin control

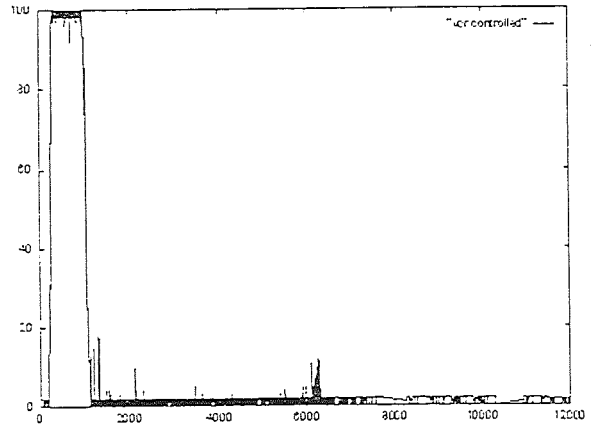


Figura 4: Evolución del tamaño de la cola del conmutador con control

Tabla 2: Parámetros de las fuentes

Parámetro	Valor
T_p (seg.)	0.010 – 0.016
T_{OFF} (seg.)	0.650 – 0.750
N (células)	22 – 32
t_s (seg.)	9
t_a (seg.)	15

como un proceso interrumpido de Poisson. En este modelo existen dos estados ON y OFF. El tiempo de estancia en el estado OFF está caracterizado por una variable aleatoria con una distribución exponencial y mientras se está en este estado no se generan células. Cuando se está en el estado ON se producen células a la tasa de pico según un proceso de Bernoulli. Tres parámetros caracterizan este tipo de fuentes:

1. Tiempo de encapsulación (T_p). Es el tiempo que se tarda en generar una célula. La inversa de este tiempo es pues, la tasa binaria de pico.
2. Tiempo medio de estancia en el estado OFF (T_{OFF}).
3. Número medio de células en el estado ON (N).

Estos tres parámetros se dejan variar de una petición de conexión a otra según una variable aleatoria uniforme, con lo que el proceso de llegada a la cola del conmutador no tiene valores tan discretos como sucedía en el ejemplo anterior, donde los valores eran fijados al comienzo de la simulación y se mantenían a lo largo de toda ella.

Una fuente cuando finaliza la llamada, vuelve a solicitar una nueva conexión transcurrido un tiempo aleatorio t_s. Si la llamada se acepta, transmite durante un tiempo también aleatorio t_a. Ambos tiempos siguen una distribución exponencial.

Los parámetros elegidos para las fuentes se recogen en la tabla 2.

Tenemos un conjunto de 8 fuentes conectadas a una cola de tamaño 100, y cuya tasa de servicio se eligió para que el número de llamadas máximo que no producen congestión sean 6 o 7 dependiendo de los parámetros de la fuente. El valor concreto es de 0.00675 células/segundo.

Para medir la congestión lo que se hace es monitorizar la tasa de pérdidas en la cola. Si esta tasa es superior a una tasa objetivo, la entrada de supervisión es 1, mientras que en los intervalos donde la tasa sea inferior a la objetivo, la entrada al FasArt será 0. El patrón de entrada que queremos que aprenda la red, consiste en un vector de 2 componentes que al igual que en el experimento anterior representan el número de células que entran en la cola en el intervalo actual y en el intervalo anterior. Con esto lo que se trata es que la red aprenda

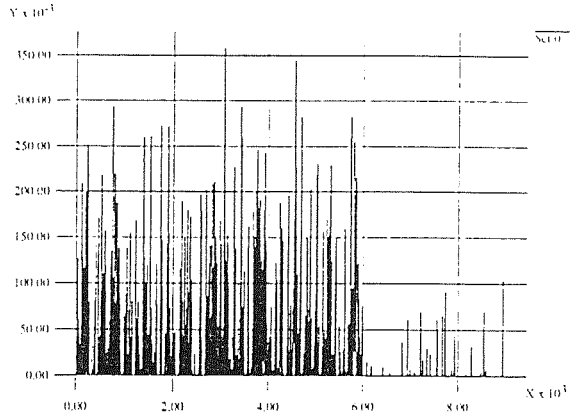


Figura 5: Tasa de pérdidas en la cola (normalizada por 400)

qué situaciones son las que llevan a la congestión y cuáles no.

El funcionamiento del esquema de control propuesto es el siguiente: Durante la fase de aprendizaje todas las peticiones de conexión se aceptan, por lo que se presentan situaciones de congestión. Durante la fase de test, cada vez que se produce una petición de conexión, al FasArt se le presenta el vector con la tasa media calculada en el intervalo de monitorización anterior, junto con la suma de la tasa media declarada por la fuente y la tasa media observada para el instante actual. La salida del FasArt es un número entre 0 y 1 que nos indica el grado en el que se va a producir congestión. Valores próximos a 0 indican que no se producirá congestión, mientras que valores cercanos a 1 predicen congestión. Si el valor de salida de la red es inferior a un cierto umbral, la llamada se acepta y en caso contrario se rechaza. Otra posibilidad es que la red no haya observado el patrón de entrada en la fase de aprendizaje o ninguno que se le parezca, por lo que la salida será indeterminada, en cuyo caso la estrategia de control es rechazar la petición de conexión.

En nuestras simulaciones utilizamos un intervalo de monitorización de 3 segundos y una tasa de pérdida objetivo de 0 células/segundo con lo que se pretende que la red sólo acepte aquellas llamadas que no produzcan congestión. El periodo durante el cual la red aprende es de 6000 segundos y la fase de test consta de 3000 segundos. Los parámetros elegidos para la red neuronal fueron $\rho = 0.95$ y $\gamma = 10.0$. En Fig. 5 se muestra cual es la tasa de pérdidas en una simulación.

6 Discusión

En este trabajo, el sistema neuro-difuso FasArt se ha usado como un identificador, dentro de un esquema general de control. Por tanto, presenta los problemas típicos de los sistemas basados en aprendizaje de funciones a partir de ejemplos. En este sentido, es muy importante tener en cuenta el grado de aprendizaje del espacio de variación de la función a identificar. Igualmente se presenta el problema de encontrar un compromiso entre capacidad de generalización y la precisión de la predicción durante la fase de test. Por eso, además de la definición de los vectores de entrada y salida, tenemos que decidir qué valores de ρ y γ elegimos a la hora de diseñar la red neuronal, así como qué patrones se presentan a la red.

En nuestros experimentos vimos que si aumentabamos la dimensión del vector de entrada, al crecer el espacio de llegada era más difícil que dos patrones fuesen similares por lo que, o bien se hacen los conjuntos más difusos para que los patrones sean semejantes, o bien aumentamos el número de patrones de aprendizaje haciendo que el periodo de aprendizaje sea mayor. Si hacemos que los conjuntos tengan un caracter más borroso, estamos difuminando la frontera entre las llamadas que deben ser aceptadas, de las llamadas que tienen que ser rechazadas, por lo que pueden darse situaciones de congestión en la cola del conmutador. Por otra parte, si mantenemos los valores de ρ y γ, necesitamos mayor número de muestras, ya que si no el número de patrones no clasificados será alto.

Como vemos la elección de estos parámetros es un gran problema, ya que de ellos depende que la red proporcione un resultado bueno, o que no se comporte lo suficientemente bien. Se necesita pués de una sintonización de estos parámetros, bien por el método de prueba y error, o bien mediante una técnica basada en algoritmos genéticos.

Una vez que los parámetros sean sintonizables automáticamente, se puede pensar en incrementar el número de fuentes así como incluir nuevos tipos de fuentes, para así conseguir un tráfico altamente heterogéneo y comprobar el funcionamiento del controlador en un entorno parecido al que sería en la vida real.

Además de esta comprobación inicial de la capacidad de FasArt para realizar tareas de CAC en redes ATM, nuestro trabajo actual de investigación trata los siguientes aspectos:

- Mejora del control efectuado, a través del aprendizaje incremental durante la fase de test.
- Estudio del carácter difuso de FasArt en nuestra aplicación. Especialmente se estudian las implicaciones de la generación automática de los conjuntos difusos y de la base de conocimiento difuso, en comparación con otros controladores difusos propuestos en la literatura. La fusión de las reglas difusas generadas en el FasArt permitirá su adecuada implantación en hardware difuso.
- Análisis de las distintas situaciones de fuentes y enlaces, así como de la posibilidad de medir el QoS a través de células.

- Finalmente, se está trabajando en la aplicación de esquemas de control basados en FasArt para realizar otras funciones de control de tráfico con mayor énfasis en el UPC [7].

7 Conclusiones

En este artículo hemos hecho una breve introducción a la problemática del control de tráfico en redes ATM, indicando cuáles son las principales funciones de control sobre las que más se está investigando. Se ha repasado la utilización de redes neuronales artificiales y arquitecturas de lógica difusa que se pueden encontrar en la literatura, viendo las ventajas que ofrecen y puntualizando en ciertos inconvenientes, especialmente resaltando la necesidad de sistemas de control que puedan ser entrenados en línea y que sean adaptativos a las características cambiantes del tráfico.

Presentamos la red neuro-difusa FasArt y la arquitectura Fuzzy Artmap sobre la que se basa. Hemos visto cuales son las ventajas del FasArt: mantiene las características de máxima generalización y mínimo error de predicción, resuelve el dilema estabilidad-plasticidad, permite aprendizaje en línea siendo adaptativo, ofrece grados de pertenencia de los patrones de entrada en problemas de clasificación, permite obtener reglas difusas y se puede utilizar como identificador de funciones.

Para ver el comportamiento del FasArt como controlador en redes ATM se han propuesto dos esquemas para realizar la función de control de la congestión CAC. En ambos casos se ha recurrido a la simulación. En el primer ejemplo usamos el simulador de ATM desarrollado por el NIST para ver si la red puede aprender las características de tráfico de las fuentes. Vimos que para un caso sencillo, la red aprende los patrones de tráfico, de manera que si se realiza el test con las mismas fuentes, se identifican las llamadas que producen congestión.

Movidos por las limitaciones de este simulador, y pensando en profundizar en este área de control, utilizamos una herramienta de desarrollo llamada Ptolemy que nos permite simular redes de comunicación. Con esta herramienta presentamos otro esquema de control para realizar el CAC, en el que la red FasArt predice situaciones en las que al aceptar una nueva llamada habrá congestión. Hemos visto también los inconvenientes de tener unos parámetros sintonizables, y como se ve en los resultados, no siempre se consiguen ajustar bien.

Finalmente hemos visto posible trabajo actual que incluye aprendizaje incremental, el uso de algoritmos genéticos para encontrar los valores de los parámetros que den un funcionamiento óptimo del controlador y la aplicación del FasArt a otras funciones de control de tráfico.

Agradecimientos

Queremos expresar nuestro agradecimiento a los miembros del Grupo de Ciencias Cognitivas Aplicadas y Visión Artificial: R. García, J. M. Cano, M. Arauzo.

Este trabajo está parcialmente apoyado por el proyecto europeo Brite/Psycho BRE-2-CT94-0976 PSYCHO.

Referencias

[1] De Prycker, M. *Asynchronous Transfer Mode: Solution for Broadband ISDN*. London: Ellis Horwood Limited (1993).

[2] ATM Forum. *Traffic Management Specification Version 4.0*. ATM Forum/af-tm-0056.000 (1996).

[3] UIT-T Rec. I.356. *Calidad de Transferencia de Células en la Capa de Modo de Transferencia Asíncrono de la Red Digital de Servicios Integrados de Banda Ancha*. Geneve, Suiza (1993).

[4] ITU-T Rec. I.371. *Traffic Control and Congestion Control in B-ISDN*. Perth, Australia (1995).

[5] Guerin, R., Ahmadi, H. y Naghshineh, H. "Equivalent Capacity and Its Aplication to Bandwidth Allocation in High-Speed Networks", *IEEE Journal on Selected Areas in Communications*, 9, 7, 968-981 (1991).

[6] Ogier, R., Plotkin, N. T. y Khan, I. "Neural Network Methods with Traffic Descriptor Compression for Call Admission Control", *Proceedings of the IEEE Infocom* (1996).

[7] Hiramatsu, A. "ATM Communications Network Control by Neural Networks", *IEEE Transactions on Neural Networks*, 1, 1, 122-130 (1990).

[8] Nordström, E., Gällmo, O., Gustafsson, M. y Asplund L. "Statistical Preprocessing for Service Quality Estimation in a Broadband Network", *Proceedings of the World Congress on Neural Networks WCNN'93*, (1993).

[9] Hiramatsu, A. "ATM call admission control using a neural network trained with a virtual output buffer method", *Proceedings of the IEEE International Conference on Neural Networks '94*, 3611-3616 (1994).

[10] Estrella, A. D., Casilari, E., Jurado, A. y Sandoval, F. "ATM Traffic Neural Control: Multiservice Call Admission and Policing Function", *Proceedings of International Workshop on Artificial Neural Networks to Telecommunication (IWANNT)*, 104-111 (1995).

[11] Tarraf, A., Habib, I. W. y Saadawi, T. N. "A Novel Neural Network Traffic Enforcement Mechanism for ATM Networks", *IEEE Journal on Selected Areas in Communications*, **12**, 6, 1088-1095 (1994).

[12] Cheng, R. y Chang, C. "Design of a Fuzzy Traffic Controller for ATM Networks", *IEEE/ACM Transactions on Networking*, 4, 3, 460-469 (1996).

[13] Ndousse, T. D. "Fuzzy Neural Control of Voice Cells in ATM Networks", *IEEE Journal on Selected Areas in Communications*, **12**, 9, 1488-1494 (1994).

[14] Catania, V., Ficili, G., Palazzo, S. y Panno, D. "A Comparative Analysis of Fuzzy Versus Conventional Policing Mechanism for ATM Networks", *IEEE/ACM Transactions on Networking*, 4, 3, 449-459 (1996).

[15] Cano, J. M., Dimitriadis, Y. A., Arauzo, M. y Coronado, J. L. "FasArt: A new neuro-fuzzy architecture for incremental learning in system identification," *13th World Congress of IFAC*, F, 133-138 (1996).

[16] Cano, J. M. *Modelos Neuro-difusos para identificación y control*, Tesis doctoral, Universidad de Valladolid, (1997).

[17] Carpenter, G. A., Grossberg, S., Markuzon, M., Rosen, D. B. y Reynolds, J. H. "Fuzzy ARTMAP: A Neural Network Architecture for Incremental Supervised Learning of Analog Multidimensional Maps". *IEEE Transactions on Neural Networks*, 3, 5, 698-713 (1992).

[18] "NIST ATM Network Simulator", Disponible en URL. *http://isdn.ncsl.nist.gov/misc/hsnt/prj_atm-sim.html*

[19] "Ptolemy Project", Disponible en URL. *http://ptolemy.eecs.berkeley.edu/*

[20] Golmie, N., Koening, A. y Su D. *The NIST ATM Network Simulator, Operation and Programing, v. 1.0*, (1995).

[21] Universidad de Valladolid, "Generalization and Extrapolation for Control Models", Technical Report, BRITE Project BRE-2-CT94-0976 PSYCHO, (1996)

[22] *The Almagest. Ptolemy 0.6 User's Manual*, Universidad de California, (1996).

Funciones UPC para tráfico VBR basadas en técnicas de inteligencia artificial

C. García Berdonés, F. D. Trujillo, A. Calisti, E. Casilari, A. Díaz Estrella y F. Sandoval
Dpto. Tecnología Electrónica. E.T.S.I. Telecomunicación.
Universidad de Málaga, Campus de Teatinos, s/n, 29071 Málaga.
Correo Electrónico: berdones@dte.uma.es

Abstract:
An important element of the control in ATM Networks is the Usage Parameter Control function (UPC). The UPC procedures take the necessary actions to enforce the compliance of an ATM connection to a negociated traffic contract. The vigilance of the parameters like sustainable cell rate and this vigilance influence in QOS parameters are an open issue. A UPC combining fuzzy and neural techniques is proposed. Its quality is measured in terms of users and impact over the QOS . In both senses, good results has been reached with a simple algorithm.

1. Introducción

La Red Digital de Servicios Integrados de Banda Ancha (RDSI-BA) debe soportar servicios multimedia (voz, vídeo y datos), con posibilidad de diversos tipos de conexiones (multipunto, punto a punto, permanentes, conmutadas, etc..) y en configuraciones unidireccionales o bidireccionales. Debe pues dotarse de una estructura capaz de soportar las actuales aplicaciones de comunicaciones y las que previsiblemente puedan aparecer en un futuro. Tiene además que, procurando una gestión óptima de sus recursos, garantizar a los usuarios un grado de satisfacción con el uso de la Red, o sea asegurar el mantenimiento de la Calidad de Servicio (QOS, *Quality of Service*) contratada por el usuario con la compañía que ofrezca los servicios RDSI-BA.

Las características del tráfico que debe soportar la red serán muy variadas y cada clase de servicio tendrá parámetros QOS en general distintos, dependiendo de la propia naturaleza del servicio. La carga de la red dependerá en cada instante no sólo del número de conexiones en curso, sino también de los tipos de conexiones. Además, debe soportar altas velocidades de transferencia para posibilitar aplicaciones en tiempo real.

El Modo de Transferencia Asíncrono (ATM, *Asynchronous Transfer Mode*) ha sido el seleccionado para la RDSI-BA. Los tres puntos básicos que definen como trabaja una Red ATM son: Conmutación rápida de pequeños paquetes de longitud fija (células), servicios orientados a conexión (definición de canales y trayectos virtuales) y uso de técnicas de multiplexación estadística para mejorar la eficiencia en la compartición de recursos (ancho de banda y *buffers*) por parte de los usuarios.

Gracias a la multiplexación estadística, cuando se utilizan fuentes VBR (*Variable Bit Rate*), los enlaces pueden admitir más conexiones de las que permitirían si consideraran que las fuentes transmiten a su velocidad de pico. Este mejor aprovechamiento de recursos tiene un coste, la probabilidad de pérdida de células (CLR, *Cell Loss Rate*) no es nula. La utilización de *buffers* puede reducir esta CLR pero no eliminarla. Así que el único recurso es o bien descartar información, y los controles de flujo tienen problemas debido a la alta velocidad que se maneja en ATM, o bien limitar el número de conexiones establecidas mediante controles preventivos.

La recomendación I.371 de la ITU-TS [1] propone un control de admisión de conexión (CAC, *Connection Admission Control*) cuya función será aceptar una nueva solicitud de conexión de usuario si la red dispone de suficientes recursos para ofrecer la QOS requerida y mantener las QOS de las conexiones previamente establecidas. Para ello, el usuario negociará con la red los recursos que se utilizarán y que quedarán reflejados en lo que se denomina contrato de tráfico. Este contrato, básicamente, describe las características del tráfico y la QOS solicitada.

Por otro lado, una vez establecida la conexión, habrá que proteger a la red de desviaciones (intencionadas o no) de los parámetros de tráfico y QOS negociados que puedan afectar adversamente a la QOS de las otras conexiones establecidas. Para ello, la recomendación I.371 establece un control de vigilancia denominado UPC (*Usage Parameter Control*).

Para un funcionamiento eficiente, los UPCs deben ser *a*) muy simples, ya que habrá uno por cada canal virtual, *b)* rápidos, para trabajar en tiempo real, y *c*) capaces de capturar las características estadísticas del tráfico a vigilar. La vigilancia de la velocidad de pico es imprescindible y, afortunadamente, hay algoritmos muy sencillos que la pueden llevar a cabo [2]. Sin embargo, la vigilancia de fuentes VBR requiere el concurso de más parámetros, siendo el más inmediato y significativo la velocidad media que afecta de directamente al asignamiento de recursos. Por

desgracia, las características de tráfico y requerimientos de QOS de los múltiples servicios que transporta la red ATM son muy variadas lo que complica enormemente el desarrollo de UPCs de velocidad media. Los controles muy complejos que detectan con precisión ligeras desviaciones del tráfico contratado son caros y relativamente lentos, mientras que controles simples, económicos y rápidos tienden a sacrificar prestaciones.

Algunas técnicas de inteligencia artificial, como las basadas en lógica difusa (LD), sistemas expertos (SEs) y redes neuronales artificiales (RNAs), pueden ayudar a mejorar los controles propuestos hasta fecha, mediante un buen equilibrio entre prestaciones, velocidad y coste. El estudio de estos novedosos métodos es el objetivo de este trabajo.

En el apartado 2 se presentan las especificaciones funcionales de un UPC ideal, así como una breve revisión de las diversas implementaciones prácticas propuestas hasta la fecha. En el apartado 3, se muestran las características básicas de las técnicas de inteligencia artificial, en particular de la RNAs, y las ventajas que pueden aportar en este terreno. En el apartado 4, se describe un UPC basado en lógica difusa y se propone una implementación neuronal del mismo. En el apartado 5 se comparan las prestaciones de los controles estudiados (difuso y neuronal) con controles convencionales (*leaky bucket*), tanto desde el punto de vista de la fuente (transparencia y tiempo de reacción) como desde el punto de vista de la red (CLR). Finalmente, en el apartado 6 se extraen algunas conclusiones de los resultados obtenidos y se proponen futuras líneas de trabajo.

2. Control de Vigilancia en ATM.

2.1 Situación del Control de Vigilancia.

En la Figura 1 se puede ver la situación del control de Vigilancia en el nodo de acceso a la red ATM. Cada fuente de tráfico es seguida por un UPC, que "filtrará" las células no conformes, realizándose la multiplexación sin la sobrecarga que la presencia de éstas supone. Es claro así que conseguir una alta ganancia estadística estará íntimamente ligado con un buen funcionamiento de UPC.

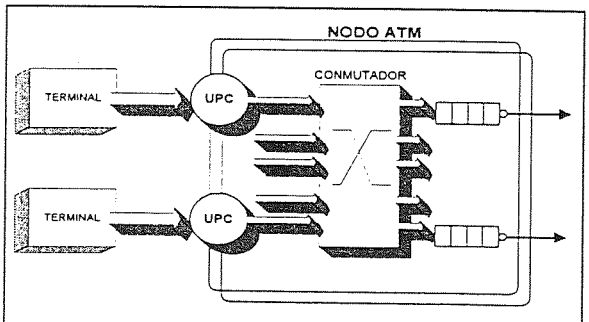


Figura 1. Situación del control de vigilancia

2.2 Parámetros sujetos a control de Vigilancia.

Los parámetros de tráfico que pueden ser objeto de control son los incluidos en el denominado descriptor de tráfico fuente [1]. El valor de estos parámetros será negociado durante la fase de establecimiento de conexión junto con la QOS solicitada.

La recomendación I.371 establece pocos parámetros a vigilar. Distingue entre características cuantitativas (velocidad de pico V_P, velocidad media V_M, etc..) y cualitativas (videoconferencia, voz, etc..). Considera que la vigilancia de la V_P es obligada, mientras que la de V_M, es propia de fuentes VBR y su eficiencia depende de la ganancia estadística que se quiera obtener. Los compromisos entre la complejidad del UPC, el caso de tráfico más desfavorable y la optimización de recursos de la red quedan a discreción de los operadores de red.

La utilidad de la vigilancia es evitar que se produzca un deterioro en la QOS de las conexiones en curso. Así, se deberían escoger los parámetros de tráfico que mejor describieran el efecto que su violación tiene en los recursos de red. En general, se deberían vigilar la función de densidad de probabilidad de la velocidad de la fuente de tráfico y estadísticos de orden superior, pero esto puede ser extremadamente complejo y llevar a no actuar en tiempo real como es requerido.

La mayor parte de las implementaciones de UPC sólo vigilan V_M y V_P. Mientras V_P es fácilmente monitorizable por cualquiera de las técnicas propuestas hasta la fecha, de las que más adelante se hará una revisión, es la vigilancia de V_M la que aún no está plenamente resuelta. La velocidad media a la que ha transmitido la fuente es evidente que no se conocerá hasta que no finalice la conexión, estimarla con la conexión en curso implica o estudiar la velocidad media en ventanas temporales largas, que aumentará la exactitud de la estimación pero también el tiempo de respuesta del control, o en ventanas cortas, que invertirá la problemática.

2.3 Medida de la calidad de UPC.

La ITU-TS [1] propone los siguientes dos parámetros para la medida de calidad de los UPC:

- Tiempo de respuesta: tiempo necesario para detectar una no conformidad con el contrato de tráfico.
- Transparencia: exactitud con la cual el UPC inicia las acciones de control adecuadas en una conexión no conforme y evita acciones de control inadecuadas en una conexión conforme (las denominadas falsas alarmas).

La transparencia se cuantifica definiendo la siguiente variable:

$$\gamma = \frac{\text{N° de células marcadas por UPC}}{\text{N° de células emitidas}} \quad (1)$$

$$\left.\begin{array}{l}\gamma_P = \gamma(\text{ UPC real }) \\ \gamma_M = \gamma(\text{ UPC ideal })\end{array}\right\} \text{ Transparencia} \equiv \gamma_M - \gamma_P$$

Estos dos parámetros son los habitualmente usados para cuantificar la calidad de un control de Vigilancia, y en cierta forma representan las acciones que dicho control está tomando de cara al usuario: cuándo y cuántas células de su conexión pueden ser marcadas.

La razón del control de Vigilancia es disminuir el impacto que en la QOS de las conexiones en curso tiene la violación del contrato de tráfico por parte de alguna o algunas de dichas conexiones. Así que se debe también medir dicho impacto para cuantificar de forma más completa la calidad de una UPC. Una posible medida es observar como afecta el inclumplimiento del contrato de tráfico en los parámetros de QOS de las colas de salida del nodo ATM.

Por último, uno de los requisitos más importantes a exigir a una implementación de UPC es la economía de costes, ya que debe existir un control de vigilancia por cada llamada.

2.4 Implementaciones de UPC.

Las primeras implementaciones de UPCs para la vigilancia de la velocidad media se basaron en el mecanismo *leaky bucket* (LB) y en técnicas de ventanas [2][3].

El LB es el mecanismo más sencillo y también el más popular. Consiste en un contador que se incrementa cada vez que la fuente genera una célula y se decrementa a velocidad constante ($V_L \approx V_M$) siempre que el valor del contador sea mayor que 0. Si la velocidad instantánea de la fuente excede la velocidad de decremento, el contador empieza a crecer. Cuando el contador alcanza cierto valor umbral N, se considera que la fuente ha incumplido el contrato de tráfico y se descartan o marcan todas las células entrantes hasta que el contador reduzca su valor por debajo del umbral. La elección de N es un valor de compromiso, para evitar que se descarten células conformes, N debería ser alto, sin embargo cuánto mayor sea N mayor será el tiempo de reacción. La vigilancia de V_M exige valores excesivamente altos de N; una forma de solventar esto es aumentando el valor de V_L a $V_L = C \times V_M$, donde C se denomina factor de sobredimensionamiento y tiene un valor ligeramente superior a la unidad.

Los mecanismos basados en ventanas observan, básicamente, cuántas células llegan en una ventana temporal, descartando las que exceden cierto valor umbral. Por ejemplo, el mecanismo JW (*Jumping Window*) consiste en un contador que se incrementa cada vez que llega una célula y se pone a cero cada T *slots* (T sería el tamaño de la ventana). El control decide que se viola el contrato cuando el contador supera cierto valor umbral N. Para evitar que se descarten células conformes, T debería ser grande, sin embargo, cuánto mayor sea T mayor será el tiempo de reacción. Tiene, por tanto, problemas similares a los detectados en el LB. Se han propuesto mejoras tales como el EWMA *(Exponentially Weighted Moving Average)* [3], similar al JW con la diferencia de que N es variable para cada ventana. En este caso, el valor umbral N_i para la ventana i está afectado por el número de células aceptadas en ventanas precedentes; esto permite una mayor tolerancia a fluctuaciones estadísticas, reduciendo el descarte de células conformes aunque, de nuevo, aumentando el tiempo de reacción.

En resumen, la vigilancia de la velocidad media no se resuelve de forma eficiente con los métodos anteriores, ya que es muy difícil conjugar los distintos requerimientos de un UPC ideal: Una mejora de la transparencia (mediante un mayor tiempo de observación o un algoritmo más complejo) trae consigo un aumento del tiempo de reacción y un aumento de complejidad en el hardware.

Parte de las limitaciones de estos mecanismos radica en su simplicidad, ya que vigilan sólo un parámetro de tráfico y tienen parámetros de funcionamiento fijos; de aquí, que se hayan propuesto múltiples mejoras, entre las que destacan *arrays* de LBs, donde se vigila la velocidad media a distintas escalas de tiempo[4][5] y mecanismos que vigilan momentos de segundo orden [6] ,o incluso la función densidad de probabilidad (fdp) de la velocidad de la fuente [7][11], lo que permite obtener, sin duda, controles mucho más estrictos. Sin embargo, todas estas mejoras complican la implementación, es más, en algunos casos, el cálculo de estos estadísticos requiere un gran esfuerzo computacional y puede llegar a hacer inviable el control en tiempo real.

Para mejorar las prestaciones sin complicar demasiado la implementación se han estudiado algoritmos basados en técnicas de inteligencia artificial usando LD [8][9][10], RNAs [11][12][13][14] y SEs [15]. [10] propone un JW donde N_i es una función de lógica borrosa de 18 reglas que dependen de la historia de V_M desde el comienzo de la conexión. Su comportamiento dinámico permite combinar pequeños tiempos de respuesta y baja probabilidad de falsa alarma, manteniendo todavía una implementación sencilla. [12] entrena RNAs que predicen la velocidad del tráfico a vigilar, de manera que una comparación con el tráfico real puede determinar la presencia de

violaciones. Su inconveniente principal es que las características de las fuentes afectan directamente al hardware implicado que, en algunos casos, podría ser grande y costoso.

Por otro lado, si se quiere obtener ganancia estadística es obligado vigilar la fuente sobre diferentes niveles de carga y/o de escalas de tiempo (esto se reflejaría en el contrato) o imponer un formato conocido a la fuente de tráfico [16]. [17] propone otra solución, basada en una compleja función UPC que controle la QOS en el multiplexor en lugar de los parámetros de tráfico. [18] confirma que las medidas estadísticas estándar que se utilizan pueden no ser apropiadas, ya que no tienen porqué estar midiendo las características del proceso más importantes para tanto predecir el efecto de la fuente sobre los recursos en la red como las prestaciones que la fuente experimentará.

En resumen, los UPCs para velocidad media y que tengan en cuenta la ganancia estadística siguen siendo un tema abierto que requiere controles más eficientes, y por tanto más complejos, sin complicar demasiado la implementación.

3. Técnicas de Control Inteligente.

El control inteligente es adaptativo, tolerante a fallos y capaz de funcionar en ambientes de incertidumbre y poca seguridad donde se pueden producir situaciones inesperadas. Los controladores inteligentes usan normalmente modelos experimentales que determinan las respuestas del sistema a modelar en función de diversos estímulos aplicados. Hay tres aproximaciones básicas [19]: SE, LD y RNAs. Los SEs están más enfocados a la resolución de problemas complejos, mientras que las técnicas de LD y RNAs están mostrando su utilidad en problemas de control. La lógica difusa necesita que un experto conozca el sistema a modelar o controlar y sea capaz de describirlo mediante expresiones lógicas del tipo condicional IF-THEN-ELSE. A partir de esta información es posible diseñar un hardware que implemente la función deseada. La RNA es complementaria y se usa cuando no se dispone de reglas fiables para describir el comportamiento del sistema, pero sí de muchos (y relevantes) datos que permitan a la RNA abstraer su funcionamiento, también es muy útil en problemas multidimensionales, debido a su arquitectura masivamente paralela y cuando se requiere alta velocidad y reducido tiempo de desarrollo. En este trabajo nos centraremos en el estudio de las RNAs.

Control neuronal

Las RNAs son sistemas de computación formados por un conjunto de unidades de proceso (neuronas artificiales) conectadas entre sí, donde la fuerza de conexión entre dos neuronas se denomina peso sináptico. En general, las RNAs se caracterizan por aprender, mediante ejemplos, el comportamiento de un sistema determinado; y este aprendizaje se materializa ajustando los pesos sinápticos mediante algún algoritmo adaptativo, denominado algoritmo de aprendizaje.

Las RNAs más usadas para control son las que pueden aproximar funciones [20]. Varios autores [21][22] han demostrado que cualquier función continua definida sobre un dominio compacto puede ser aproximada con tanta seguridad como sea necesaria por medio de una RNA *feedforward* con una capa oculta. La RNA más popular es el Perceptrón multicapa *feedforward* (PMF) con aprendizaje *backpropagation* [23].

4. UPC Neuronal.

4.1 Precedentes.

Nuestra propuesta de UPC se basa en el controlador difuso propuesto por [10], cuya arquitectura se puede ver en la Figura 2.

Los variables de entrada de este controlador difuso son tres:

- A_{Oi}: número medio de células llegadas por ventana desde el comienzo de la conexión.
- A_i: número de células que han llegado en la última ventana.
- N_i: número de células permitido en la última ventana.

La variable de salida, ΔN_i , permitirá conocer el valor umbral para la siguiente ventana en la forma:

$$N_{i+1} = N_i + \Delta N_i \qquad (2)$$

Como se puede ver el control se basa en la evaluación de tres parámetros: comportamiento global de la fuente desde el comienzo de la conexión hasta el instante en que se realiza el control (A_{oi}), comportamiento local del tráfico en la ventana en curso (A_i), comportamiento del algoritmo en la ventana anterior (N_i) .

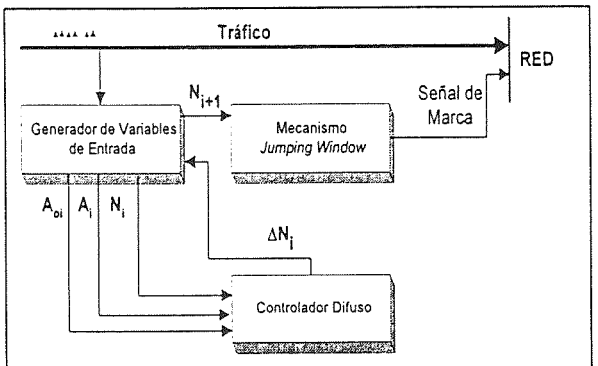


Figura 2. Estructura de UPC con control Borroso

La forma en que el control otorga o quita créditos (aumenta o disminuye el umbral N_i) en función de estos parámetros es materializado en una serie de reglas que serán la base del control difuso.

4.2 Implementación Neuronal.

La implementación neuronal sigue la filosofía de trabajo de [10] con tres variantes fundamentales:*a*) se modifican las reglas de lógica difusa, *b*) se sustituye el *hardware* de lógica difusa por una RNA y *c*) se reduce el número de parámetros de entrada a dos (A_{oi} y A_i). En la Figura 3 se puede ver un diagrama de bloques de este control.

4.2.1 Reglas de Control.

Cuantificaremos los distintos estados en que puede estar el sistema en función del valor esperado de células por ventana ($M = T \times V_M$). También en función de M definiremos los créditos para cada ventana y el valor umbral N_{INIC} para la primera ventana.

Las reglas establecidas para el control aparecen en la Tabla I. Como se puede ver, ante un mismo comportamiento "puntual" (en la ventana actual) de la fuente, la acción a tomar será diferente en función de la historia del tráfico a lo largo de la conexión. Por ejemplo, se penaliza en diverso grado si la conexión viene transmitiendo justo con la velocidad media y el comportamiento en la ventana no es conforme (reglas 7, 8, 9).

La elección de los valores de N_{INIC} e ΔN_i es el resultado del compromiso entre la consecución de un buen tiempo de respuesta y una baja probabilidad de falsa alarma. Nuestra elección ha sido mapear los valores de ΔN_i en el intervalo [7*M*/16, -7*M*/16] con un valor N_{INIC} de 4*M*.

El siguiente paso es hacer que la RNA aprenda los puntos característicos definidos en la Tabla I, de forma que en funcionamiento pueda interpolar para cualesquiera otros puntos que se presenten en el sistema. Para trabajar con la RNA se necesitan valores numéricos concretos por lo que se simulará un tráfico determinado que permitirá cuantificar *M*.

4.2.2 Tráfico.

Las características de tráfico usadas para el

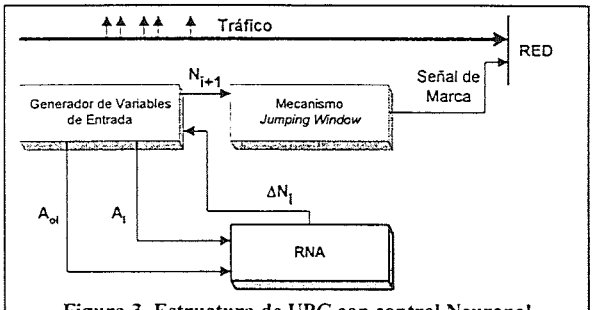


Figura 3. Estructura de UPC con control Neuronal

	A_{oi}	A_i	ΔN_i
1	0	M	6/16 M
2	0	2M	0/16 M
3	M	0	7/16 M
4	M	M/2	6/16 M
5	M	3/4M	4/16 M
6	M	M	0/16 M
7	M	5/4M	-4/16 M
8	M	3/2M	-6/16 M
9	M	2M	-7/16 M
10	2M	0	7/16 M
11	2M	M	0/16 M
12	2M	2M	-7/16 M

Tabla I. Reglas de Control

entrenamiento, y posteriormente para las diversas pruebas han sido las siguientes:

- Fuente ON-OFF con distribución exponencial de los tiempos ON y OFF.
- T_a (Tiempo entre células consecutivas en Ton) = 34.8 μs
- Media *Ton* =174 μs
- Media *Toff*=243 μs

Con estas características el tráfico generado tiene una V_P de 12.2Mbps y una V_M de 5.04Mbps.

Se ha escogido un tiempo de ventana para el algoritmo de Jumping Window de $T = 2.67$ ms, que en media contendrá 6 periodos ON-OFF. Esto supone un valor de M de 32 células.

4.2.3 Topología de la RNA. Entrenamiento.

La Figura 4 muestra la arquitectura de la RNA utilizada. Es un PMF con dos neuronas de entrada, dos ocultas y una de salida. En la fase de entrenamiento, se utilizaron como patrones de aprendizaje , las reglas de la tabla I. El algoritmo de aprendizaje fue el *backpropagation.*

El valor máximo que pueden llegar a alcanzar las variables A_{oi} y A_i será el utilizado para normalizar las entradas de la red. Este valor es el máximo número de células que pueden llegar en una ventana, T/T_a. Los valores de salida serán normalizados en el intervalo [0.1, 0.8]. Así, los parámetros de la RNA a_{oi}, a_i y dN_{i+1}, corresponden a los A_{oi}, A_i y ΔN_{i+1} normalizados.

5. Simulación y resultados.

Se ha realizado la simulación en dos escenarios distintos. En el primero se pretende evaluar la calidad de la UPC propuesta en los términos que la ITU establece, transparencia y tiempo de respuesta, obteniendo los resultados respecto a una sola fuente cuyos parámetros de tráfico incumplen en diversa medida el contrato.

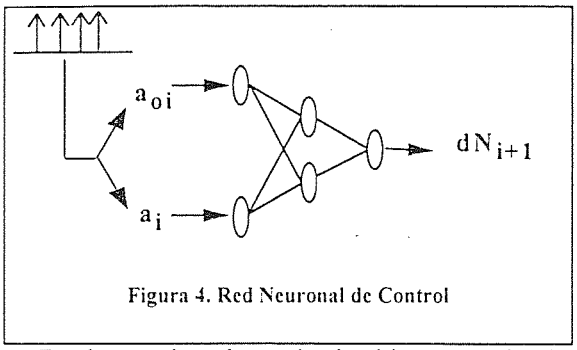


Figura 4. Red Neuronal de Control

En el segundo se ha evaluado el impacto sobre la QOS en un entorno que integra a varias fuentes que son no conformes en distinto grado y cantidad.

Ambos escenarios se describirán a continuación junto con los resultados obtenidos con el control de vigilancia propuesto.

5.1 Resultados para una fuente.

Escenario y condiciones de simulación

Fuente: La descrita en el apartado 4.2.2

Parámetro de No Conformidad:

Definido σ como

$$\sigma = \frac{\text{Velocidad media real}}{\text{Velocidad media contratada}} \qquad (3)$$

las variaciones en este parámetro permitirán medir la eficiencia de nuestro controlador. Concretamente se hace variar sobre el tráfico conforme la duración del periodo ON, hecho que modifica la velocidad media de la fuente pero no su velocidad pico.

El comportamiento ideal que debe seguir un mecanismo de vigilancia viene dado por la siguiente expresión:

$$\gamma_M = \begin{cases} \dfrac{\sigma-1}{\sigma} & \sigma \geq 1 \\ 0 & \sigma < 1 \end{cases} \qquad (4)$$

Mecanismos de vigilancia.

Se han implementado:

1. El algoritmo *leaky bucket*, eligiendo un valor para el límite del contador, *N*, de 300 células y un factor de sobredimensionamiento *C* de valor 1.1.
2. El algoritmo difuso propuesto en [10], tal y como allí se describe, con la única variante de realizar la defuzificación con el algoritmo de Centro de Máximo.
3. El algoritmo neuronal propuesto en este artículo tal y como se ha descrito.

Resultados

En la Figura 5 se muestra el comportamiento de un UPC ideal junto con las tres implementaciones comentadas. Como se puede ver el ajuste a la curva ideal es realizado de forma óptima por la UPC neuronal.

Respecto al tiempo de respuesta, en las Figuras 6 y 7 se puede observar como el control de vigilancia neuronal resulta ser comparable e incluso mejor en el caso de no conformidad leve que el resto de mecanismos.

5.2 Resultados para varias fuentes.

Se pretende, por un lado, comprobar en qué medida puede ser necesario un algoritmo que efectúe correctamente la función de vigilancia y por otro comparar la efectividad del *leaky bucket* con la del control neuronal desarrollado.

Escenario y condiciones de simulación

Fuentes: Se han multiplexado veinte fuentes ON-OFF de las características descritas en el apartado 4.2.2. (Ver Figura 8)

Para llevar a cabo la simulación, se han ido aumentando progresivamente el número de fuentes no conformes y disminuyendo el de fuentes conformes. En las fuentes no conformes la velocidad media es de 7.5 Mbps (σ = 1.5) .

Mecanismos de vigilancia.

Se han implementado:

1. El algoritmo *leaky bucket* descrito en el apartado 5.1.
2. El control neuronal descrito en el apartado 5.1.

La opción usada en estos mecanismos de vigilancia ha sido la de descartar células.

Variables de salida:

Se pretende obtener la CLR en una cola, tanto en el caso de existir vigilancia como cuando ésta no exista. Por lo tanto, se ha diseñado una cola de un tamaño de 100 células con un tiempo entre servicios determinista de valor 2.7 μs. (la velocidad del enlace ATM es de 155 Mbps).

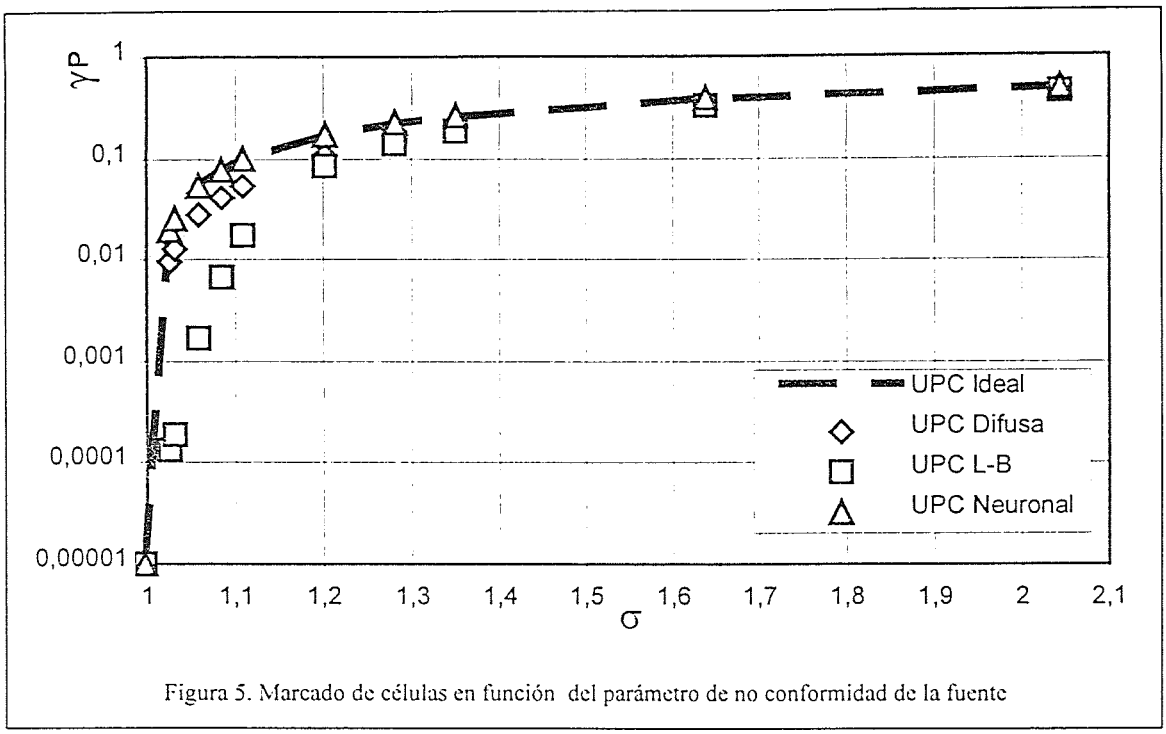


Figura 5. Marcado de células en función del parámetro de no conformidad de la fuente

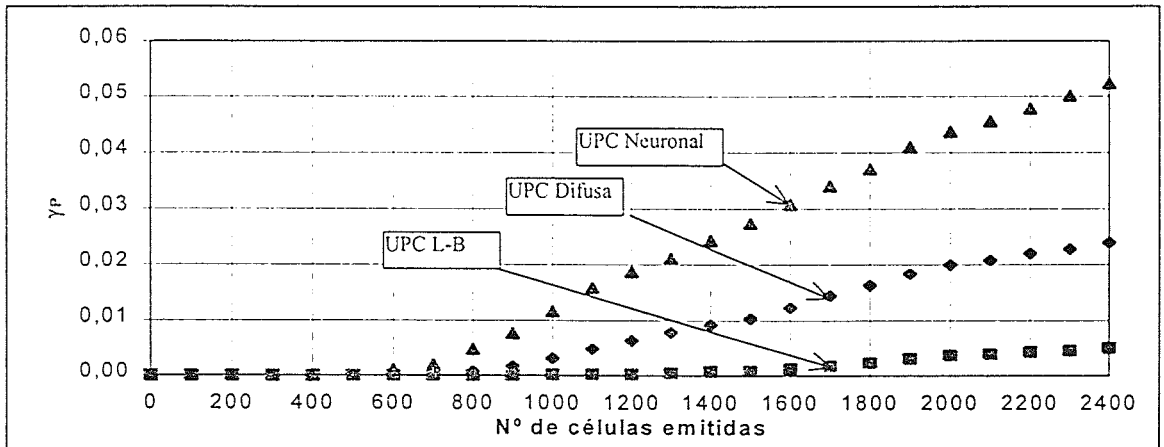


Figura 6. Tiempo de reacción. Marcado de células en función del N° de células emitidas para $\sigma = 1.09$

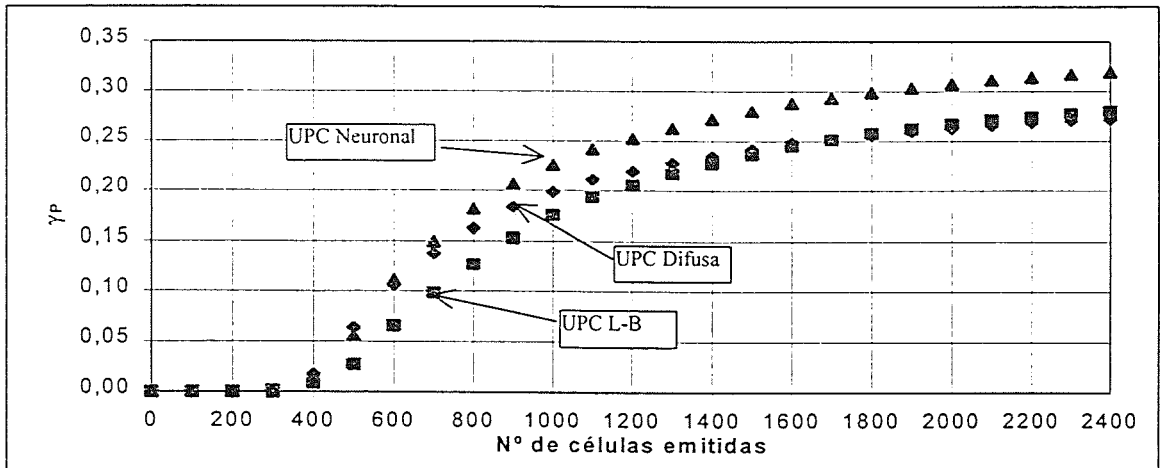


Figura 7. Tiempo de respuesta. Marcado de células en función del N° de células emitidas para $\sigma = 1.64$

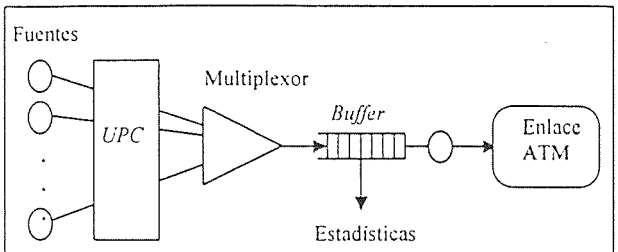


Figura 8 Escenario de simulación.

Resultados

Los resultados obtenidos se muestran en la Figura 9. En esta gráfica se puede observa que, mientras la mejora que se ha conseguido con el *leaky bucket* apenas es significativa, sí se ha conseguido reducir bastante más la probabilidad de pérdidas al usar un control neuronal.

La mejora aumenta conforme crece el número de fuentes no conformes, puesto que el control neuronal tiene un comportamiento más cercano al ideal y, por lo tanto, descarta un mayor número de células comparado con el *leaky bucket*. Este hecho lleva consigo la correspondiente mejora en la CLR, debido a que el mecanismo descarta un mayor número de células no conformes y el *buffer* colocado a la salida del nodo se sobrecarga menos.

6. Conclusiones. Líneas Futuras.

Se ha propuesto un mecanismo para el Control de Vigilancia que en estas primeras estimaciones resulta ser una alternativa mejorada a diversas técnicas propuestas en la literatura. Es un control simple, tal y como marcan los requisitos para una UPC, ya que se implementa en base a una RNA de tan sólo tres elementos de proceso (dos neuronas ocultas y una de salida).

Se ha comparado la calidad del mecanismo, no sólo de cara al marcado de células sino también de cara al fin último de una UPC: conseguir mantener la QOS del resto de usuarios que comparten el medio cuando alguna conexión viola el contrato de tráfico. Resultando ser eficaz en este sentido el mecanismo aquí propuesto.

Se ha constatado en el transcurso de este trabajo que el mayor problema de la arquitectura propuesta, JW con control neuronal de umbral, no es tanto este control, sino los propios parámetros del JW. Se han propuesto reglas eficientes, pero sin embargo sencillas, que han conducido al uso de una RNA simple como controladora; sin embargo los parámetros del JW, tamaño de la ventana y del umbral de descarte, se han escogido en este trabajo de manera heurística, comprobando básicamente dos hechos: por un lado la elección de los parámetros debe ser el resultado de un compromiso entre ajuste correcto de probabilidad de descarte y tiempo corto de reacción, y por otro, la elección óptima va ligada al tipo de tráfico que se desee vigilar, que desde este punto de vista no queda en absoluto definido por velocidad media y de pico.

Así, futuros trabajos deben incidir en estudiar como se puede definir un UPC para diversas clases de tráfico, de forma que los parámetros de este mecanismo de control quede unívocamente definidos por las características contractuales del tráfico. Esto con independencia de que, por supuesto, dichas características sean correctamente vigiladas por la UPC.

De otra parte, pensamos que la eficiencia de los mecanismos de Vigilancia en la gestión de la QOS deben ser siempre un parámetro a considerar para la correcta cualificación de dichos mecanismos, como lo ha sido en este artículo. Se debe incluso considerar la posibilidad de diseñar futuros UPCs que incorporen como entrada alguno o algunos de

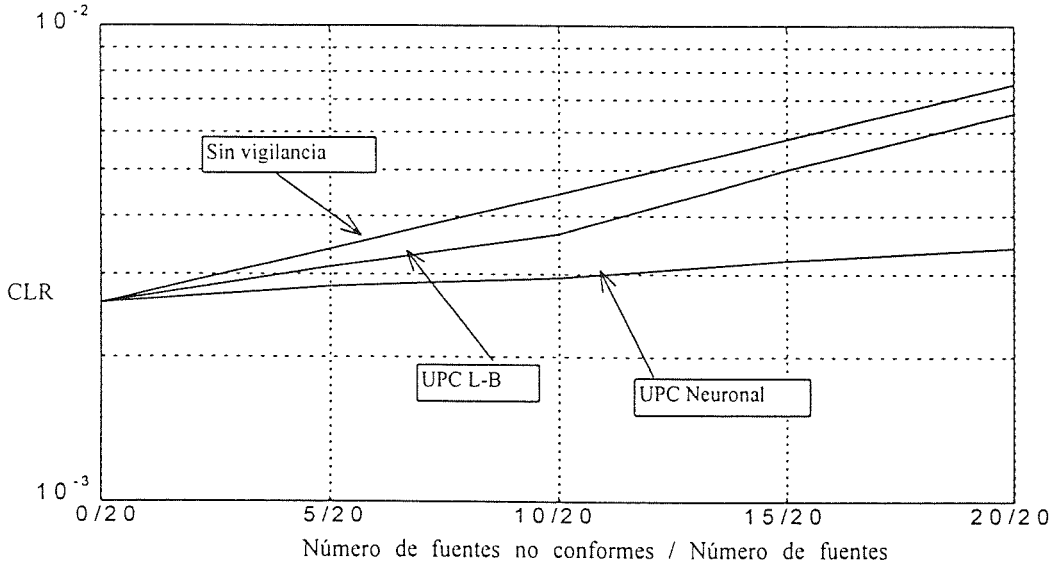


Figura 9. Efecto del Control de Vigilancia sobre la QOS

los parámetros del estado de los *buffers* de salida del nodo, con lo que previsiblemente se aumente su eficiencia.

Agradecimientos

Este proyecto ha sido parcialmente financiado por la Comisión Interministerial de Ciencia y Tecnología (CICYT), Proyecto No. TIC96-0743.

Referencias

[1] ITU-T Recommendation I.371-Traffic control and congestion control en B-ISDN, Genova (1996).

[2] Onvural, R.O. *Asynchronous Transfer Mode Networks*, Artech House, Boston (1994).

[3] Rathgeb, E., "Modeling and performance comparison of policing mechanism for ATM networks", *IEEE Journal on Selected Areas in Communications*, **9**, 3, 325-334 (1991).

[4] Örs, T., "An investigation of the leaky bucket for preventive control in ATM networks", *Tesis doctoral*, Universidad de Surrey (1994).

[5] Hemmer, H. y Huth, P. "Evaluation of policing functions in ATM networks", en *Queueing, performance and control in ATM*. Cohen J. y Pack C. (eds.). Elsevier Science Publishers, 111-116 (1991).

[6] Hluchyj M. G., y Yin, N. "A second-Order Leaky Bucket Algorithm to Guarantee QOS in ATM Networks", *Proceedings of GLOBECOM'96*, London, **2**, 1090-1096 (1996).

[7] Prycker, M. *Asynchronous transfer mode. Solution for broadband ISDN*, 3ª ed., Prentice Hall International, Estados Unidos (1995).

[8] Ndousse, T.D. "Fuzzy Neural Control of Voice Cells in ATM Networks", *IEEE Journal on Selected Areas in Communications*, **12**, 9, 1488-1494 (1994).

[9] Scheffer, M. y Kunicki, J., "Fuzzy adaptive traffic enforcement for ATM networks", *Proceedings of MELECOM'96*, 1047-1050 (1996).

[10] Catania, V., Ficili, G., Palazzo, S. y Panno, D. "A comparative analysis of fuzzy versus conventional policing mechanism for ATM networks", *IEEE/ACM Transactions on Networking*, **4**, 3, 449-459 (1996).

[11] Tarraf, A., Habib, Y. y Saadawi, T. "Neural networks for ATM multimedia traffic prediction", en *Applications of neural networks to telecommunications*. Alspector, J., Goodman, R. y Brown T. (eds.), Lawrence Erlbaum Associates Publishers, 85-91 (1993).

[12] Tarraf, A., Habib, Y. y Saadawi, T., "A novel neural network traffic enforcement mechanism for ATM networks", *IEEE Journal on Selected Areas in Communications*, **12**, 6, 1088-1096 (1994).

[13] Díaz Estrella, A., Casilari, E., Jurado, A. y Sandoval, F. "ATM traffic Neural Control: Multiservice Call Admission and Policing Function" en *Applications of Neural Networks toTelecommunications 2*, Alspector, J., Goodman, R. and Brown, T.X (eds.), Lawrence Erlbaum Associates, Publishers, 104-111 (1995).

[14] Onyiagha, G., Krasniqi, X. y Clarkson, T., "Adaptive access control of ATM traffic using neural networks", *Proceedings of GLOBECOM'96*, **1**, 201-205 (1996).

[15] Tsui, K. y Azvine B., "ATM traffic policing using a classifier system", en *Applications of Neural Networks toTelecommunications 3, Alspector, J., Goodman, R. and Brown, T.X (eds.)*. Lawrence Erlbaum Associates, Publishers, 97-106 (1997).

[16] Baiocchi, A., Bléfari-Melazzi, N., Cuaomo, F. y Listanti, M. "Achieving Statistical Gain in ATM Networks with the Same Complexity as Peak Allocation Strategy", *Proceedings of INFOCOM'94*, Toronto, Canadá, **1**, 3374-382 (1994).

[17] Borgonovo, F. and Fratta, L. "Policing Procedures: Implications, Definitions and Proposals", en *Teletraffic and Data traffic*, Jensen A. and Iversen V.B. (eds.), Elsevier Science Publishers, 859-866 (1991).

[18] Lucantoni, D.M., Neuts, M.F. y Reibman, A.R. "Methods for Performance Evaluation of VBR Video Traffic Models", *IEEE/ACM Transactions on Networking*, **2**, 2, 176-180 (1994).

[19] White, D.A. y Sofge, D.A. *Handbook of Intelligent Control. Neural, Fuzzy and Adaptive Approaches*, Van Nostrand Reinhold (1992).

[20] Harris, C.J., Moore, C.G. and Brown, M. *Intelligent Control: Aspects of Fuzzy Logic and Neural Nets*, World Scientific Publishing, (1993).

[21] Hornik, K., Stinchcomb, M. y White, H. "Multilayer Feedforward Networks are Universal Approximators", *Neural Networks*, **2**, 359-366 (1989).

[22] Funahashi, K. "On the approximate realization of continuous mappings by neural networks", *Neural Networks*, **2**, 183-192 (1989).

[23] Rumelhart, D.E., Hinton G. E. y Williams R. J. "Learning internal representations by error propagation", en *Parallel Distributed Proccesing: Exploration in the Microstructure of Cognition*, Rumelhart, D.E. and McClelland, J.L (eds.), MIT Press, 318-362 (1986).

Técnicas para el Control de Congestión en la Clase de Servicio ABR

Jorge Martínez, José Ramón Vidal y Luis Guijarro
Departamento de Comunicaciones, E.T.S.I.T.
Universidad Politécnica de Valencia
Camino de Vera s/n, 46071-Valencia
jmartinez@upvnet.upv.es

Abstract:

The search for efficient switch algorithms for the ABR class of service is a topic of continuous interest. This paper presents a new algorithm called CAPAC (Congestion Avoidance with Proportional Adaptive Control) that improves the performance of a previously proposed algorithm called CAPC. The results of a simulation study that evaluates its dynamic performance is also provided, both with greedy sources and with bursty sources.

1. Introducción.

El Modo de Transferencia Asíncrono (ATM) ha sido adoptado como la técnica de multiplexación y conmutación que soporta la Red Digital de Servicios Integrados de Banda Ancha (RDSI-BA). Uno de los objetivos con el que se han diseñado las redes ATM ha sido el de hacer posible el soporte de aplicaciones que requieren una calidad de servicio (QoS) garantizada cuantitativamente. La red permite garantizar el QoS solicitado por las aplicaciones realizando un tratamiento diferenciado de las células de las diferentes conexiones.

Tanto el ATM Forum en [For96] como la UIT-T en [I.371] definen un conjunto de clases de servicio que pueden ser utilizadas por las aplicaciones según sus requerimientos de QoS. Estas clases de servicio pueden clasificarse, de acuerdo con la forma en que garantizan el QoS de las aplicaciones, en tres grandes grupos: i) las que ofrecen garantías cuantitativas; ii) las que ofrecen garantías cualitativas; iii) las que no ofrecen garantías.

Utilizando la nomenclatura del ATM Forum, el primer grupo esta compuesto por las clases de servicio CBR (*Constant Bit Rate*), VBR-rt (*Variable Bit Rate-real time*) y VBR-nrt (*Variable Bit Rate-non real time*); el segundo grupo por la clase de servicio ABR (*Available Bit Rate*); y el tercer grupo estaría compuesto por la clase de servicio UBR (*Unspecified Bit Rate*).

Para que la red pueda garantizar un QoS a las aplicaciones, es preceptivo que se firme el denominado "contrato de servicio". Para que éste pueda ser firmado, la aplicación tiene que facilitar a la red el QoS deseado y los parámetros que describen estadísticamente el perfil de tráfico que ésta generará durante el tiempo de conexión. Con esta información la red pone en funcionamiento sus mecanismos de CAC (*Call Admission Control*) e intenta encontrar una ruta con los recursos libres suficientes para garantizar el QoS de la nueva conexión sin poner en peligro el QoS de las conexiones ya establecidas. El mecanismo de CAC es por tanto el que realiza las funciones denominadas de Control de Congestión Preventivo.

La clase de servicio ABR, pensada en principio para el soporte de tráfico de datos, permite ofrecer un QoS garantizado únicamente de forma cualitativa, puesto que las aplicaciones que la utilizan se reparten el ancho de banda disponible, es decir, aquel que no está siendo utilizado por las conexiones que han solicitado garantías cuantitativas. Para ello se ha estandarizado un mecanismo de realimentación en bucle cerrado que permite a la red enviar señales a las fuentes para que éstas adapten su tasa de emisión al estado de la red en cada instante. Este mecanismo de realimentación permite implementar el denominado Control de Congestión Reactivo.

La clase de servicio UBR, también pensada para el soporte de servicios de datos, no ofrece ningún mecanismo de control de congestión. Para evitar que las prestaciones desciendan por debajo de límites razonables debido al proceso de segmentación*, se han propuesto mecanismos de descarte selectivo en los conmutadores. Por ejemplo, los descritos por Romanow y Floyd en [Rom95].

Este artículo se centra en el estudio del mecanismo de control de congestión estandarizado por el ATM Forum en [For96] para la clase de servicio ABR. Aunque se ha estandarizado la forma en la que la fuente† debe responder a las señales de la red y el mecanismo de realimentación entre

---

* Cuando un mensaje se segmenta en células, la pérdida de una sola célula hace inservible el mensaje en destino. Además, el resto de células del mismo mensaje malgastan recursos de red, puesto que serán descartadas de cualquier forma por la máquina de reensamblado del de la capa AAL.

† En este caso se ha preferido utilizar el término fuente en vez de aplicación puesto que los mecanismos de control de flujo a los que nos referimos son propios de la capa ATM y no son percibidos directamente por la aplicación en muchos casos. Ésta pudiera tener una percepción indirecta a través de señales de *backpressure* enviadas por la capa ATM como consecuencia del vaciado lento de la cola existente entre ésta capa y la capa superiores. Por tanto, por fuente entenderemos a partir de ahora un ente generador de células localizado en la capa ATM.

conmutadores y fuentes, se han dejado sin estandarizar los algoritmos que utilizan los conmutadores para estimar la porción de ancho de banda equitativa (PABE) que le corresponde a cada conexión ABR.

En particular, se presenta un nuevo algoritmo de conmutador ABR denominado CAPAC (*Congestion Avoidance with Proportinal Adaptive Control*), basado en una propuesta anterior hecha por Barnhart en [Bar94] denominada CAPC. El algoritmo propuesto por Barnhart define un conjunto de constantes que deben ser optimizadas para cada configuración de red y para cada subconjunto de fuentes, lo cual limita considerablemente su aplicación práctica. El algoritmo CAPAC en cambio permite que el valor de dichas constantes se adapten de forma dinámica a las condiciones de la red, consiguiendo mejorar substancialmente las prestaciones del algoritmo propuesto por Barnhart.

Para evaluar las prestaciones del algoritmo propuesto, se hace uso de la metodología de evaluación propuesta en [Mar96], que permite medir las prestaciones de los algoritmos de conmutador en régimen dinámico mediante la introducción de cambios de ancho de banda tipo escalón.

El resto del artículo se estructura como sigue. En la sección 2 se presentan los elementos que componen la clase de servicio ABR y se describen con mayor detalle el comportamiento estandarizado de las fuentes y de los conmutadores. En la sección 3 se define el modelo de simulación utilizado para obtener los resultados que se presentan en el resto del artículo. Los resultados ha sido obtenidos mediante utilizando el programa BONeS. En la sección 4 se clasifican los algoritmos de conmutador ATM atendiendo principalmente a dos criterios: las necesidades de almacenamiento y la complejidad de computación. En la sección 5 se presentan los aspectos básicos del algoritmo CAPC y se enumeran sus limitaciones. En la sección 6 se resumen los resultados del algoritmo CAPC+ propuesto anteriormente en [Mar96]. En la sección 7 se describe el nuevo algoritmo denominado CAPAC y se presentan los resultados de su evaluación, tanto con fuentes persistentes como con fuentes tipo ráfaga. Finalmente en la sección 8 se apuntan los aspectos de los algoritmos ABR que a priori pueden tener un mayor impacto sobre las prestaciones de protocolos de capa superior, en particular TCP. Finalmente, en la sección 9 se presentan las conclusiones.

2. Elementos que Componen la Clase de Servicio ABR.

Los elementos que componen la clase de servicio ABR son: i) los terminales fuente y destino; ii) los conmutadores; iii) los mecanismos de realimentación; iv) las fuentes y destinos virtuales; y v) los mecanismos UPC/NPC. Una descripción detallada de éstos ya ha sido realizada por Bonomi y Fendick en [Bon95] y no se repetirá aquí. En cambio, sí es conveniente describir brevemente el funcionamiento, tanto del protocolo de fuente/destino que se ejecuta en los terminales, como de las acciones que realizan los conmutadores.

El protocolo fuente/destino tiene dos funciones principales: i) insertar unas células especiales, denominadas RM (*Resource Management*), en el flujo de datos de la conexión para hacer llegar a la fuente las señales de realimentación desde los conmutadores; ii) adaptar la tasa de transmisión de acuerdo a la señal de realimentación recibida. Una descripción exhaustiva del protocolo fuente/destino ha sido realizada por Jain *et al.* en [Jai96a].

El formato completo de las células RM ha sido definido en [For96]. En este artículo destacaremos sólo algunos campos:

- DIR (1 bit). Indica si la célula RM es *forward* (FRM) o *backward* (BRM). Las células RM emitidas por la fuente se denominan FRM. Cuando éstas llegan al destino, éste les cambia el bit DIR y las devuelve hacia la fuente como células BRM.
- CI (*Congestion Indication*)(1 bit). Activado por un conmutador cuando experimenta congestión.
- NI (*No Increase*)(1 bit). Activado por un conmutador para evitar que una fuente continúe aumentando su tasa.
- ER (*Explicit Rate*). Campo que la fuente suele incializar a PCR (*Peak Cell Rate*). Las fuentes declaran el PCR deseado en la fase de establecimiento de la conexión. Los algoritmos ABR de conmutador comparan el valor que han estimado para la PABE con el campo ER de las células RM que reciben y lo sobre-escriben con la primera si ésta es inferior. Los algoritmos ABR de conmutador realizan este proceso normalmente sólo sobre las células BRM, aunque algunos también lo realizan sobre las células FRM.

El comportamiento de las fuentes viene definido por un conjunto de parámetros que se definen en tiempo de establecimiento (por ejemplo,

PCR) y un conjunto de variables. De éstas últimas destacaremos ACR (*Allowed Cell Rate*), que mantiene la tasa de células a la que la fuente tiene derecho a transmitir en un momento dado. Cuando la fuente recibe una célula BRM, compara ER con ACR, y si éste último es superior actualiza ACR=ER. Si ER fuese superior y el campo NI no estuviese activado, la fuente podría aumentar su tasa de emisión según el proceso aditivo:

$$ACR=ACR+RIF*PCR$$

donde RIF (*Rate Increase Factor*) es un parámetro que se define en tiempo de establecimiento. En ABR se crea por tanto un mecanismo de control de flujo en bucle cerrado basado en la cooperación entre fuentes y conmutadores.

3. Evaluación de Prestaciones. Modelo de Simulación.

Algunos de los objetivos de diseño de los algoritmos de conmutador ABR son: i) utilización eficiente del ancho de banda disponible; ii) minimización de las pérdidas de células por desbordamiento de las colas en los conmutadores; iii) reparto equitativo del ancho de banda disponible, es decir, un reparto lo más próximo posible al que se obtendría aplicando el criterio de equidad elegido por el operador*; iv) convergencia rápida a una situación estable al cambiar el ancho de banda disponible. Cuanto más rápida sea la convergencia a la nueva situación, mayor será la utilización del enlace y menores serán las pérdidas.

Los parámetros de mérito que se han elegido para evaluar las prestaciones de los algoritmos son: i) ACR instantáneo de las fuentes; ii) ocupación instantánea de las colas en los conmutadores; iii) utilización instantánea de los enlaces.

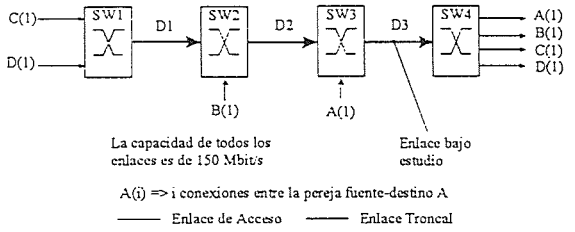


Fig. 1. Configuración *Parking Lot.*

Para evaluar cuantitativamente en qué medida un algoritmo de conmutador ABR cumple los objetivos anteriormente descritos, se recurre habitualmente a la evaluación mediante simulación por eventos discretos. El ATM Forum ha propuesto un conjunto de configuraciones sobre las que es deseable realizar las simulaciones, pues, en cierta medida, representan escenarios reales de diferente complejidad. Las configuraciones que se han elegido en este trabajo son las denominadas *Parking Lot* y *Generic Fairness Configuration 1* (GFC1); esta última propuesta por Simcoe en [Sim94]. Aunque el interés inicial de estas configuraciones fue evaluar el grado de equidad en el reparto de ancho de banda en régimen permanente, aquí se modificarán ligeramente para evaluar su comportamiento en régimen transitorio.

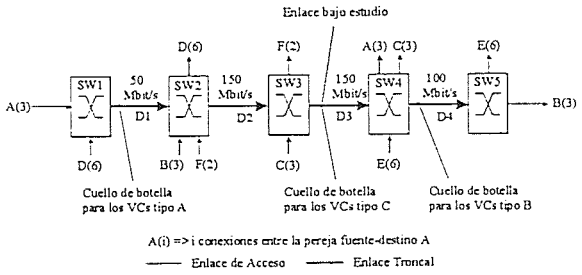


Fig. 2. Configuración GFC1.

Las configuraciones se describen el las Fig. 1 y Fig. 2 . En ellas se ha supuesto distancias MAN, es decir, la longitud de los enlaces troncales es de 50 Km., mientras que la de los enlaces de acceso es de 0.2 Km. Se utilizarán fuentes persistentes, excepto cuando se indique lo contrario, pues éstas son más adecuadas para estudiar el comportamiento dinámico de los algoritmos. Para evaluar cómo el algoritmo se adapta, a medida que se van estableciendo nuevas conexiones que compiten por el mismo ancho de banda, se fuerza a que las fuentes comiencen a transmitir en instantes diferentes de tiempo. Así, en la configuración *Parking Lot,* la fuente A comienza en el instante 0 ms, las fuentes B y C lo hacen el instante 20 ms. y la fuente D lo hace en el instante 30 ms. En la configuración GFC1 todas las fuentes comienzan en el instante inicial.

En este trabajo se ha supuesto una arquitectura de conmutador con colas localizadas en los puertos de salida. La disciplina de servicio en las colas se ha supuesto FIFO. En las redes evaluadas sólo se han establecido conexiones ABR. La existencia de tráfico "de fondo" que requiere una calidad garantizada se ha simulado con cambios tipo escalón en el ancho de banda disponible para las conexiones ABR. En este caso, se eligen los enlaces etiquetados como "enlace bajo estudio" para introducirles los cambios de ancho de banda disponible. La amplitud de los cambios de ancho de banda son: de 150 a 75 Mbit/s en t1 y de 75 a 150 Mbit/s en t2. Los valores de t1 y t2 valen: i) en la configuración *Parking Lot,* 50 y 100 ms.; ii) en la configuración GFC1, 30 y 60 ms.

* Diferentes criterios de equidad son posibles, ver por ejemplo en [Yin94]. En este trabajo se ha elegido el criterio Max-Min [Ber92].

PCR	354 céls/ms	Nrm	32	Crm	1000
MCR	PCR/1000	Mrm	2	CDF	0
ICR	PCR/100	Trm	10 ms	TCR	10 céls/s
TBE	-	ADTF	10 s	RIF	1/16
FRTT	-			RDF	1/16

Tabla 1. Definición de los parámetros de fuente.

Además, los parámetros de mérito ocupación instantánea de la cola y utilización instantánea del enlace, se miden sólo en el "enlaces bajo estudio" de cada configuración.

Los parámetros de fuente que se utilizarán en las simulaciones se definen en la Tabla 1 . Como se observa, se ha preferido desactivar el mecanismo *use-it-or-loose-it* eligiendo un valor de ADTF suficientemente elevado. También se ha preferido desactivar el mecanismo de reducción del ACR en caso de rotura de enlace haciendo Crm=1000 y CDF=0. El PCR seleccionado corresponde a una capacidad de 150 Mbit/s, es decir un STM-1 al que se ha descontado la capacidad dedicada a la cabecera de trama.

4. Clasificación de los Algoritmos ABR.

En la literatura se han propuesto diferentes criterios para clasificar los algoritmos de conmutador ABR. Un posible criterio de clasificación es el de su complejidad de implementación, bien sea en cuanto a necesidades de almacenamiento o bien en cuanto a sus necesidades de potencia de cálculo.

Algunos algoritmos requieren almacenar y actualizar el valor de diferentes variables que describen el estado de cada conexión, además de un conjunto reducido de variables globales. En este caso, diremos que su complejidad es O(n) (de orden n), siendo n el número de conexiones ABR establecidas. Otros en cambio sólo requieren almacenar un conjunto reducido de variables globales, en este caso diremos que son de complejidad O(1). Algoritmos con requerimientos de almacenamiento O(1) son por ejemplo el EPRCA, propuesto por Roberts en [Rob94], y el CAPC, mientras que el ERICA, propuesto por Jain *et al.* en [Jai95], es O(n).

En cuanto a las necesidades de potencia de cálculo, éstas son muy variadas. La mayoría de algoritmos requieren realizar algún tipo de operaciones cuando reciben células BRM. Estas operaciones pueden ser tan sencillas como una comparación y la escritura de un registro; de complejidad media, como la obtención del resultado de una fórmula que incluye multiplicaciones, divisiones, sumas y comparaciones; y de complejidad elevada, como la de realizar operaciones con uno de los registros almacenados por conexión para el conjunto de todas las conexiones. En los dos primeros casos se habla de complejidad computacional O(1), mientras que el último caso se habla de complejidad computacional O(n). Además, en algunos algoritmos es necesario también realizar algún tipo de operaciones cuando se reciben células FRM, que normalmente son de orden O(1). Aunque la mayoría de algoritmos requieren una complejidad computacional O(1), el propuesto por Charny, Clark y Jain en [Cha94], requiere una complejidad O(n).

Otro criterio de clasificación interesante es según la técnica utilizada para la detección de la congestión. Los primeros algoritmos detectaban la congestión cuando el número de células en la cola superaba un umbral, por ejemplo el EPRCA. Una mejora posterior del EPRCA, propuesta por Siu y Tzeng en [Siu94] detectaba la congestión cuando la derivada de la función Q(t) era positiva, siendo Q(t) la ocupación instantánea de la cola. En general, estos algoritmos ofrecen un comportamiento oscilatorio, que se manifiesta con una ocupación de las colas en régimen permanente elevado, lo cual limita la rapidez con la que los algoritmos pueden converger y encarece el coste de los conmutadores puesto que deben ser dimensionados con colas de mayor tamaño.

Una segunda generación de algoritmos, detectaba la congestión mediante el denominado Factor de Carga (*Load Factor*, LF), es decir la relación entre la tasa agregado a la entrada de la cola (*Input Rate*, IR) y el ancho de banda disponible en el enlace correspondiente. De forma que se detecta congestión cuando LF>1. Esta técnica es empleada, por ejemplo, por los algoritmos CAPC y ERICA. Para estimar la tasa agregado de entrada, el tiempo se divide en intervalos de medida. Un intervalo de medida finaliza cuando se recibe un número dado de células (N), o bien cuando vence un temporizador (T). En ambos casos la tasa de entrada se estima como la relación entre el número de células recibida y duración de la ventana de observación. Un valor común para N es 100 células, mientras que para T es 1 ms.

Los algoritmos que utilizan LF como indicador de congestión, incorporan además la técnica denominada *congestion avoidance*. Esta técnica consiste en utilizar para el cálculo de LF, el denominado *Target Rate* (TR), en vez del valor del ancho de banda disponible. El TR se obtiene multiplicando el valor real del ancho de banda disponible por un coeficiente denominado factor de utilización. Puesto que el factor de utilización suele estar entre 0.85 y 0.95, se consigue una ocupación

media de las colas prácticamente nula en régimen permanente.

5. El Algoritmo CAPC.

Como su nombre indica, el algoritmo CAPC incorpora la técnica de evitación de la congestión, al calcular el factor de carga como LF=IR/TR . Al finalizar cada intervalo de medida, el algoritmo estima la Porción de Ancho de Banda Equitativa (PABE) que le corresponde a cada conexión en función del último valor estimado para la PABE y de unas constantes de convergencia según la siguiente expresión:

$$PABE_{n+1} = ERX * PABE_n$$

La forma en que se computa ERX es diferente según el puerto esté o no congestionado:

cuando $LF_{n+1} \leq 1$,

$$ERX = \min [ERU, 1+(1-LF_{n+1})Rup]$$

cuando $LF_{n+1} > 1$,

$$ERX = \max [ERF, 1+(LF_{n+1} -1)Rdn]$$

Como puede observarse, ERU y ERF representan respectivamente los valores máximo y mínimo de ERX. Además el valor de ERX se adapta linealmente según LF por medio de dos constantes: Rup y Rdn. Los valores de estas constantes deben ser elegidos cuidadosamente para evitar oscilaciones. Barnhart en [Bar94] recomienda los siguientes valores: ERU=1.5, ERF=0.5, Rup=0.1 y Rdn=0.8. En la Fig. 3 se representa el valor de ERX en función de IR.

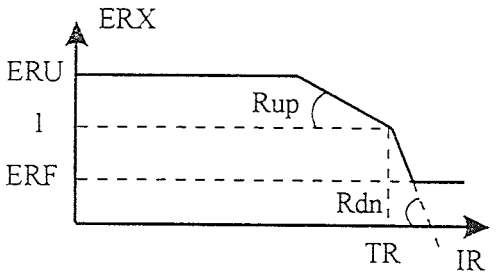


Fig. 3. ERX en función de IR

El algoritmo sólo procesa las células BRM, de forma que si el campo ER de éstas es mayor a la PABE, ER se sobre-escribe con el valor de ésta. El algoritmo observa también la longitud de la cola, de forma que mientras ésta esté por encima de un umbral Qthres=20, se activará el bit NI de las células BRM de todas las conexiones, hasta que esta situación, considerada de congestión severa, haya desaparecido.

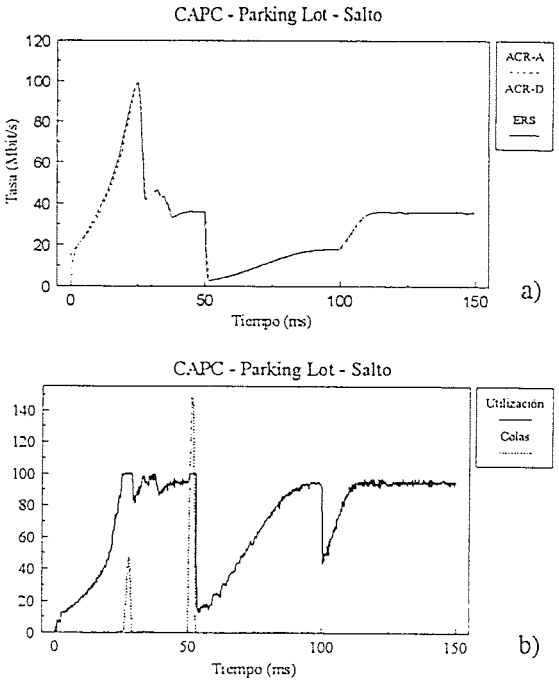


Fig. 4. a) ACRs instantáneos para el algoritmo CAPC; b) Utilización instantánea y Q(t) para el algoritmo CAPC.

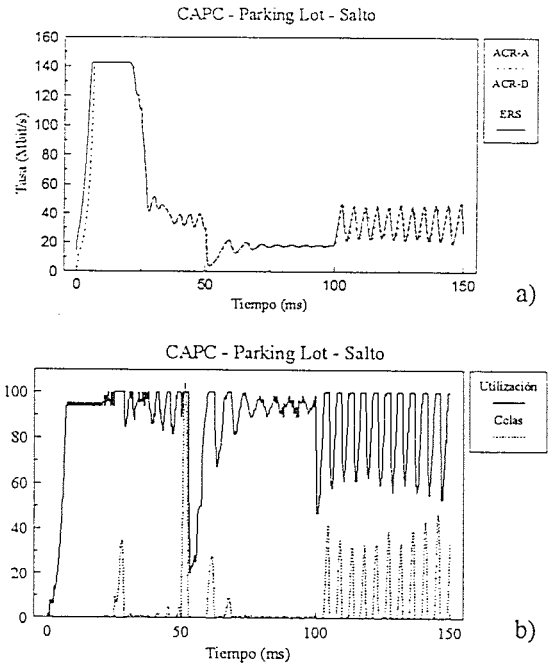


Fig. 5. a) ACRs instantáneos para el algoritmo CAPC; b) Utilización instantánea y Q(t) para el algoritmo CAPC.

En la Fig. 4 se observa la evolución del ACR de las conexiones en una configuración *Parking Lot*, para unos valores de los parámetros iguales a los definidos con anterioridad y con un factor de utilización del 95%. En esta figura, y en el resto de las que se presentarán posteriormente, se ha denominado ERS a la PABE. Como se puede

observar, el comportamiento es demasiado amortiguado, lo cual produce una pobre utilización del ancho de banda. En la Fig. 5 se ha repetido la misma simulación pero ahora para unos valores de Rup=0.5 y Rdn=1.0. Como se puede observar, en este segundo caso el algoritmo responde con mayor rapidez, puesto que el pico de Q(t) es menor, pero desgraciadamente, el algoritmo ahora oscila a partir de 100 ms.

6. El Algoritmo CAPC+.

El algoritmo CAPC+ fue propuesto por Martínez, Llop y Galindo en [Mar96] para mejorar el comportamiento del algoritmo CAPC básico. El CAPC permite que la PABE disminuya sin control en estado de congestión. En cambio, en el CAPC+ se limita la disminución de la PABE hasta un mínimo razonable, definido como:

$$\text{PABEmin} = \frac{\text{A.B. Disponible}}{\text{N}^{\text{o}} \text{ de Conex. Activas}}$$

La PABEmin representa, por tanto, el caso peor y corresponde a aquel en que todas las conexiones están limitadas en el enlace cuello de botella. Además, el algoritmo define el estado de Congestión Severa cuando la ocupación de la cola supera un umbral (Q(t)>*VeryCongested*). En este estado, el algoritmo deja que la PABE varíe libremente, es decir, que varíe igual que en el algoritmo CAPC. El razonamiento que sustenta este mecanismo es el siguiente. Si la disminución de la PABE hasta PABEmin no es suficiente para contener la congestión es porque, posiblemente, hayan fuentes muy alejadas del conmutador que tardan en reaccionar. Por ello, se pide la cooperación de las fuentes cercanas para que éstas, disminuyendo su PABE por debajo del límite razonable, ayuden a controlar la congestión*.

En la Fig. 6 se pueden observar los resultados de la simulación para el algoritmo CAPC+ con Rup=0.5, Rdn=1.0, VeryCongested=110 y el resto de parámetros iguales al escenario de las figuras Fig. 4 y Fig. 5.

Como se puede observar, el incremento en el valor de los parámetros Rup y Rdn permite acelerar la convergencia, pero gracias a los mecanismos introducidos, las oscilaciones se han limitado considerablemente.

* Puesto que la veracidad de este razonamiento puede ser limitada en algunos casos prácticos (por ejemplo, cuando todas las fuentes son lejanas), conviene limitar la celeridad con la que las fuentes lejanas ocupan ancho de banda, definiendo para ellas valores bajos de RIF.

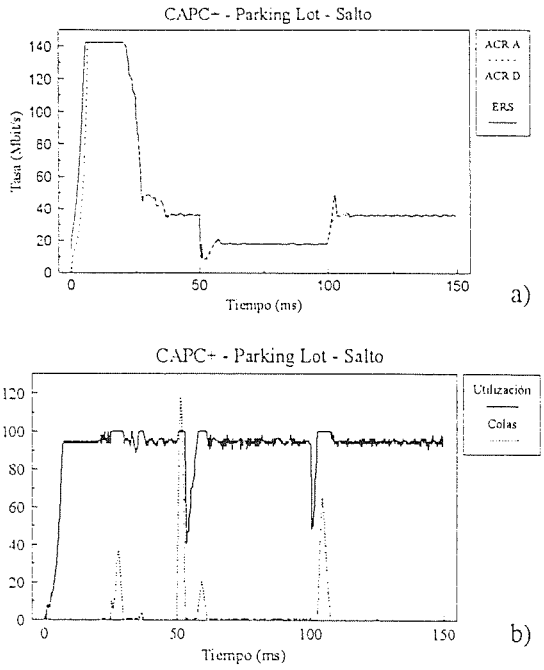


Fig. 6. a) ACRs instantáneos para el algoritmo CAPC+; b) Utilización instantánea y Q(t) para el algoritmo CAPC+.

Una de las limitaciones que se identificaron para el algoritmo CAPC+ fue que, tal y como se había definido, el Número de Conexiones Activas debía ser estimado cada intervalo de medida. Con un intervalo de medida de 100 células, esta estimación es claramente no escalable a un escenario real con miles de conexiones. Una solución a este problema ha sido propuesta recientemente por Jain *et al.* en [Jai96], donde se estima el número de medio de conexiones activas mediante una función promedio exponencial (*exponentially averaged estimate*).

7. El Algoritmo CAPAC.

El algoritmo CAPAC pretende resolver los problemas del ajuste de las constantes Rup y Rdn que presenta el algoritmo CAPC desde otra perspectiva.

El objetivo perseguido al incrementar el valor de las constantes Rup y Rdn no es otro que el de aumentar la rapidez de convergencia del algoritmo, pero desgraciadamente esto puede producir oscilaciones como las presentadas en la Fig. 5.

En realidad, lo que es deseable es que el algoritmo busque su punto de estabilidad mediante cambios de gran tamaño en la PABE cuando el LF está alejado de su valor óptimo (LF=1). Pero simultáneamente, es deseable que los cambios en la PABE sean de pequeña magnitud a medida que LF se aproxime a su valor óptimo. Es por ello que en el

algoritmo CAPAC, los valores de Rup y Rdn ya no son constantes, sino que se adaptan al valor de LF mediante las siguientes funciones lineales:

cuando $LF_{n-1} \leq 1$,

$Rup = \min [MaxRup, MinRup(1-LF_{n-1})KRup]$

cuando $LF_{n-1} > 1$,

$Rdn = \max [MaxRdn, MinRdn(1-LF_{n-1})KRdn]$

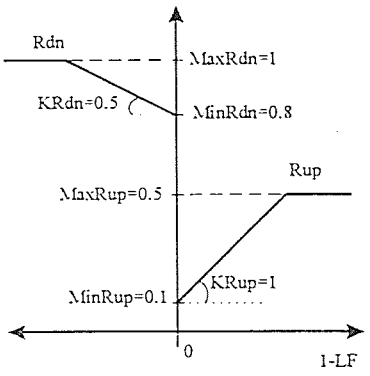


Fig. 7. ERX en función de IR

En la Fig. 7 se representan los valores de Rup y Rdn en función de LF para un conjunto de constantes de adaptación. Aunque el valor de estas constantes ha sido estimado empíricamente, se ha observado mediante simulación que la convergencia del algoritmo es independiente de las constantes de adaptación, siempre que el valor de éstas estén dentro de unos márgenes razonables, es decir, del orden del $\pm 50\%$. Esto representa una mejora substancial respecto al algoritmo CAPC, para el cual, como ya se ha descrito, la elección de los parámetros Rup y Rdn es crítica.

7.1 Evaluación de Prestaciones Mediante Fuentes Persistentes.

En la Fig. 8 se representa el resultado de la simulación para un escenario igual al del algoritmo CAPC y con los valores de las constantes de adaptación dados en la Fig. 7 .

Como se puede observar en la Fig. 8 los picos en Q(t) en los instantes de cambio son ligeramente inferiores a los obtenidos para el algoritmo CAPC+. En cambio, la utilización del enlace, tras la reducción del ancho de banda disponible, crece más lentamente de lo que sería deseable. Posteriormente se describe un mecanismo que podría eliminar esta limitación.

En la Fig. 9 se representan los valores de los ACRs de las conexiones que atraviesan el "enlace bajo estudio" en la configuración GFC1 de la Fig. 2 y las funciones Q(t) y utilización instantánea de ese mismo enlace. Esta configuración representa una situación más genérica, puesto que el ancho de banda del "enlace bajo estudio" está compartido por conexiones que están limitadas en enlaces anteriores (conexiones A), conexiones que están limitadas en enlaces posteriores (conexiones B) y conexiones que están limitadas en ese mismo enlace (conexiones C).

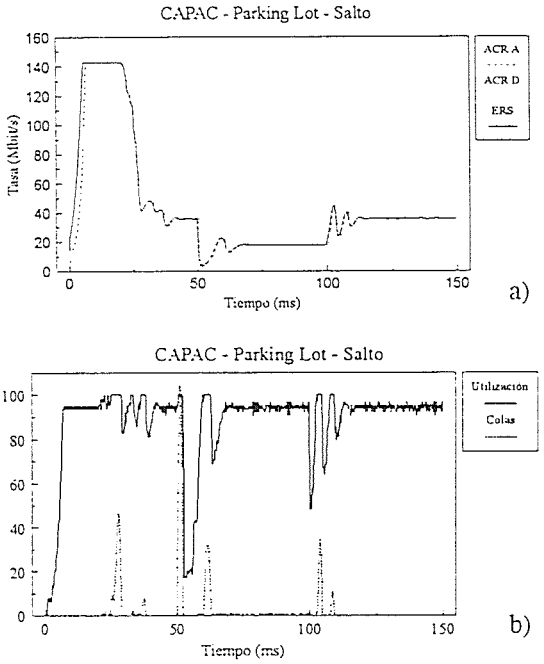


Fig. 8. a) ACRs instantáneos para el algoritmo CAPAC; b) Utilización instantánea y Q(t) para el algoritmo CAPAC.

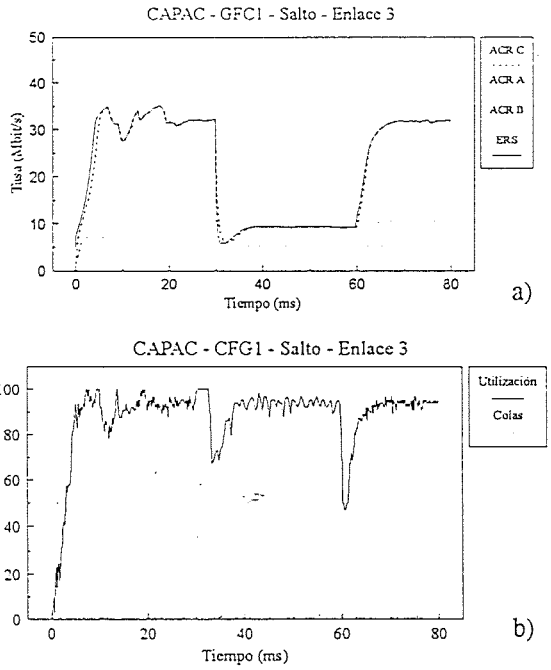


Fig. 9. a) ACRs instantáneos para el algoritmo CAPAC; b) Utilización instantánea y Q(t) para el algoritmo CAPAC.

Como se puede observar en la Fig. 9, las oscilaciones son en este caso prácticamente

inexistentes. Esto es debido, en parte, a la proximidad de las fuentes C al puerto congestionado. La existencia de 9 conexiones sobre el "enlace bajo estudio", cada una de ellas ocupando un ancho de banda de algo superior a 5 Mbit/s en el caso peor, asegura un flujo suficiente de células para obtener intervalos de medida cortos y por tanto actualizaciones rápidas de los parámetros del algoritmo.

7.2 Evaluación de Prestaciones Mediante Fuentes Tipo Ráfaga.

En un escenario más realista encontraríamos sobre un mismo enlace, conexiones con fuentes prácticamente persistentes (por ejemplo, aplicaciones como la transferencia de un fichero de gran tamaño) y conexiones con fuentes tipo ráfaga (por ejemplo, aplicaciones de tipo transaccional). Son estas últimas las que crean situaciones más críticas para los algoritmos de conmutador, puesto que pueden generar ráfagas que desaparecen antes de que el algoritmo haya podido conformarlas convenientemente.

Para simular el comportamiento de una aplicación transaccional, sería deseable que las fuentes no transmitieran la siguiente ráfaga hasta haber recibido contestación desde su ente homólogo en destino. Para simplificar la implementación de este comportamiento, en este trabajo se han configurado las fuentes para que no transmitan la siguiente ráfaga hasta que no haya transcurrido un retardo dado desde el instante en que la última ráfaga fue completamente recibida en destino. Este retardo simula el tiempo que tardan las respuestas en llegar a la fuente, incluyendo, por tanto, el retardo de procesado en el ente destino más el retardo de tránsito por la red de la respuesta hacia el ente origen.

Así pues, se generan a priori un conjunto de parejas longitud de ráfaga y retardo de forma independiente y con igual probabilidad a partir de los siguientes subconjuntos: i) las longitudes de ráfaga son: 100, 300, 1000 y 3000 células; ii) los retardos son: 1, 4, 10 y 40 ms. . Con las parejas generadas se construye una tabla de la que las fuentes van leyendo los valores. Cada fuente lee de la tabla las parejas de valores de forma secuencial y en orden creciente de índices. Para conseguir máxima aleatoriedad en el tráfico, cada fuente comienza a transmitir a partir de un índice de la tabla diferente.

En la Fig. 10 se ha representado Q(t) en el puerto del "enlace bajo estudio" en una configuración *Parking Lot* para dos valores del parámetro ADTF, 10 s. y 3 ms. .Recordemos que ADTF controla el mecanismo *use-it-or-loose-it* y define el tiempo máximo que una fuente puede retener el valor de ACR antes de inicializar éste a ICR. Es decir, define el tiempo máximo que puede ser válida la visión del estado de la red que una fuente adquirió antes de pasar a un estado de inactividad.

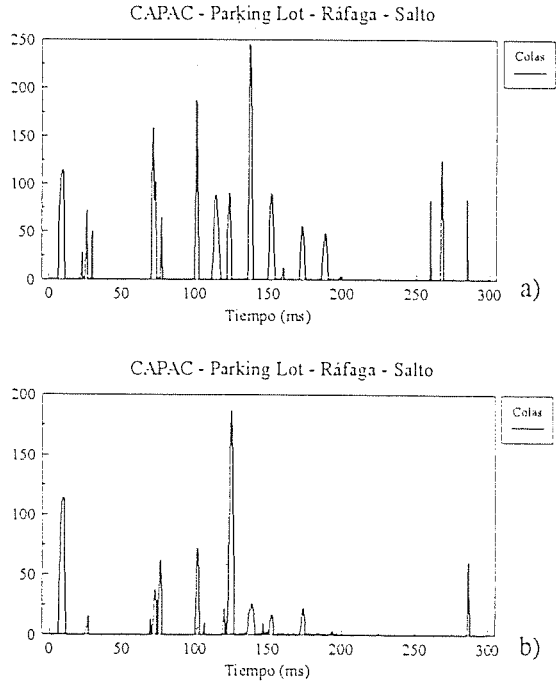


Fig. 10. Q(t))con CAPAC para diferentes configuraciones del mecanismo *use-it-or-loose-it* . a) ADTF=10 s.; b) ADTF=3 ms.

Un valor de ADTF=10 s. equivale prácticamente a desactivar el mecanismo *use-it-or-loose-it;* mientras que el valor de ADTF=3 ms. corresponde a un valor más apropiado para el escenario de simulación propuesto, del orden del doble del máximo RTT (*Round Trip Time*).

En este escenario, también se ha forzado un cambio en el ancho de banda disponible de la misma magnitud que en las simulaciones anteriores, pero ahora los instantes de cambio son 100 ms. y 200 ms. Sin embargo, las fuentes, en este caso, comienzan todas a transmitir en el instante inicial.

Para este escenario, no se representan los valores instantáneos de los ACRs ni la utilización del enlace por no ser significativos.

Como cabría esperar, valores diferentes de ADTF cambian radicalmente el comportamiento de las fuentes, obteniéndose gráficas de Q(t) completamente diferentes. Sin embargo, a la vista de los resultados, dos conclusiones podrían ser realizadas: i) con valores menores de ADTF se obtiene una ocupación media de la cola menor, lo

cual es, en principio, coherente con el hecho de que las fuentes, tras un periodo de inactividad, comienzan a transmitir un mayor número de veces con una tasa ICR*; ii) La disminución del valor máximo de Q(t) no es proporcional a la disminución de ADTF, es decir, los picos en Q(t) parecen estar más relacionados con la conjunción de eventos fortuitos adversos que ocurren como consecuencia de la propia aleatoriedad de las fuentes, es decir, la coincidencia de ráfagas.

Estos eventos adversos se ven favorecidos por el siguiente fenómeno. Cuando una fuente deja de transmitir, sus células RM, que todavía circulan por el bucle, recogen una información del estado de congestión de los conmutadores demasiado optimista, puesto que ésta se estima con la ausencia de su propio tráfico. Estas células RM al volver a la fuente, que ahora está inactiva, tienden a incrementar el ACR de ésta, de forma que si la fuente vuelve a transmitir antes de que haya vencido el temporizador ADTF, lo hacen con un ACR inadecuado, produciéndose los picos que se observan en la Fig. 10 .

El impacto de este fenómeno depende además del valor de ADTF, de muchos otros factores dependientes de la configuración y estado de la red, por lo que es difícil de cuantificar en un caso práctico.

7.3 Méritos y Limitaciones del Algoritmo CAPAC.

En las simulaciones realizadas se ha preferido utilizar unos valores para las constantes de adaptación de Rup y Rdn ligeramente agresivos, de forma que se premia la obtención de unos valores bajos para los picos en la ocupación de las colas y así prevenir las pérdidas. Esto, en algunos casos, se hace a costa de una disminución en la utilización del enlace en los instantes en los que ocurre la disminución del ancho de banda disponible, es decir, en los instantes en los que aparece una situación de fuerte congestión.

Este tipo de comportamiento es deseable puesto que ABR se ha pensado para el soporte de tráfico de datos. En este caso, la pérdida de una sola célula requiere la retransmisión del paquete entero. Dado que las PDUs de capa AAL5 pueden ser de hasta 64 Kbytes, el impacto que la pérdida de un paquete tiene sobre la utilización de los recursos de la red no es despreciable.

* Esto trae como consecuencia una disminución efectiva de *troughput* medio que una fuente puede conseguir en una ventana de tiempo dado.

No obstante, dos limitaciones del algoritmo podrían ser puestas de manifiesto:

1. La permanencia de la PABE en valores por debajo de los óptimos durante periodos excesivamente largos. Ello provoca que la utilización del enlace, tras la reducción del ancho de banda disponible, crezca más lentamente de lo que sería deseable. Esto es debido a la conjunción de dos fenómenos: i) a la agresividad con la que el algoritmo reduce la PABE al encontrar congestión; ii) a la forma en que se ha definido el intervalo de medida, el cual, cuando la tasa agregado (IR) es baja, se alarga demasiado tiempo y por tanto ralentiza el crecimiento de la PABE. Este problema podría ser resuelto haciendo que el valor del parámetro N (número de células recibidas), que define la finalización el intervalo de medida, se ajustase con el valor de IR, al menos, dentro de un margen.

2. Aunque la convergencia del algoritmo parece asegurada, incluso cuando se le somete a perturbaciones significativas, como la reducción del ancho de banda disponible a la mitad de forma instantánea o cuando se le somete a tráfico de tipo ráfaga, ésta debería de garantizarse analíticamente. Desgraciadamente, el número de parámetros susceptibles de ser configurados y que tienen impacto sobre ésta es tan grande, que la tarea no parece, a priori, sencilla.

Entre los méritos del algoritmo se podrían destacar los siguientes:

1. Es un algoritmo con una complejidad de implementación muy reducida. Al igual que el CAPC, no requiere de ningún tipo de parámetros por conexión, aunque el CAPAC requiere una multiplicación adicional por intervalo de medida.

2. El algoritmo tiene una respuesta muy agresiva en Estado de Congestión, asegurando que los picos en la ocupación de la cola serán reducidos. Al ser un algoritmo perteneciente a la clase de los que previenen la congestión, mantiene la cola normalmente vacía.

8. Soporte de Protocolos de Capa Superior.

El éxito de una nueva tecnología como ATM radica, en cierta medida, en que sea capaz de soportar eficientemente tecnologías y protocolos que se están utilizando en la actualidad de forma masiva. Por ejemplo, las redes LAN basadas en las

normas IEEE 802.x, o bien el conjunto de protocolos TCP/IP. Para el primer caso, el ATM Forum ha definido en [For95] la solución *LAN Emulation*, en cambio el segundo caso es todavía objeto de investigación activa.

El soporte de TCP sobre la clase de servicio UBR ha sido ampliamente analizada por numerosos trabajos de investigación, de entre los que cabe destacar el realizado por Romanow y Floyd en [Rom95]. En cambio, el soporte de TCP sobre ABR presenta un reto adicional puesto que en este escenario funcionan dos mecanismos de control de congestión a diferentes escalas de tiempo.

El objetivo de los trabajos de investigación en este caso es doble. Por una parte, analizar el impacto que sobre las prestaciones tiene la utilización de fuentes TCP, puesto que los algoritmos de conmutador fueron diseñados inicialmente para trabajar con fuentes persistentes. Por otra parte, optimizar el valor de los parámetros que definen el comportamiento de TCP y diseñar mecanismos adicionales que permitan a los algoritmos ABR soportar de forma eficiente el nuevo tráfico.

El objetivo de esta sección no es realizar un análisis de prestaciones del soporte de TCP sobre ABR, sino analizar a priori la idoneidad de algunos algoritmos ABR frente a otros para el soporte de fuentes TCP. En particular, estamos interesados en analizar el retardo introducido por diferentes algoritmos ABR.

El protocolo TCP detecta la congestión cuando se pierden paquetes. Una de las formas de detectar la perdida de un paquete es mediante temporizadores. Para evitar la toma de decisiones erróneas, TCP estima tanto el valor medio de RTT como su varianza. Una componente del RTT es el tiempo de espera en las colas de los conmutadores. Esta componente puede ser dominante, sobre todo en configuraciones WAN.

Cuando aparece una situación de congestión, puesto que el tiempo que las fuentes tardan en reaccionar no es nulo, se acumularán células en la cola del enlace congestionado hasta que se ajusten los ACRs de las fuentes. En este trabajo se han analizado un conjunto de algoritmos ABR, todos ellos con un comportamiento similar en cuanto a las acciones que se toman para salir de la congestión.

Básicamente, estos algoritmos siguen reduciendo su PABE mientras detectan congestión. Esto les lleva, en general, a realizar una estimación a la baja excesiva de la PABE óptima. Este efecto, al desaparecer el estado de congestión, tiene dos consecuencias: i) la disminución de la utilización; ii) el vaciado rápido de la cola. Es a esta segunda consecuencia a la que queremos prestar mayor atención.

El vaciado rápido de la cola, después de una situación de congestión, es debido a que durante un corto periodo de tiempo IR es considerablemente menor a TR, de forma que el servidor utiliza esta capacidad sobrante para vaciar la cola.

Este comportamiento no aparece en otros algoritmos ABR, como por ejemplo el ERICA. Este algoritmo tiende a ajustar de forma casi exacta los ACRs de las fuentes para que IR sea igual a TR lo antes posible. Por ello, cuando desaparece la congestión, el servidor no tiene apenas capacidad sobrante para vaciar la cola. Por ejemplo, cuando el factor de utilización es del 95%, sólo el 5% de la capacidad de servidor puede ser dedicado a vaciar la cola. En la Fig. 11 se representa el resultado de una simulación del algoritmo ERICA en una configuración *Parking Lot*. Tanto los parámetros de la configuración como de las fuentes se han mantenido iguales a los definidos con anterioridad. Para el algoritmo ERICA se ha utilizado un intervalo de medida de N=60 células.

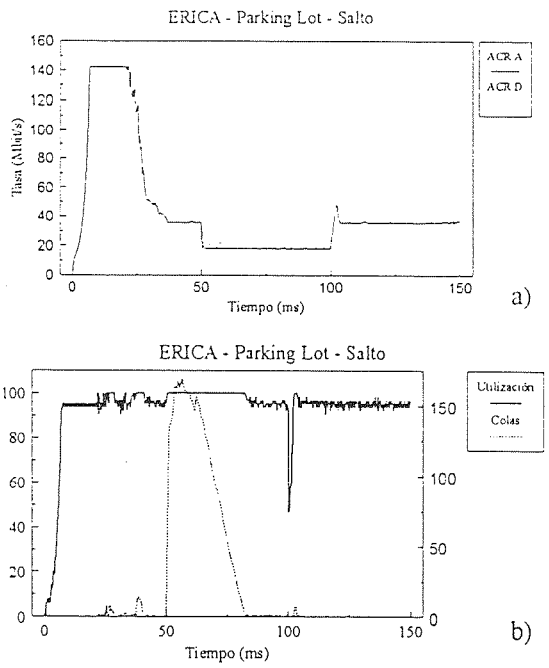


Fig. 11. a) ACRs instantáneos para el algoritmo ERICA; b) Utilización instantánea y Q(t) para el algoritmo ERICA.

Como se puede observar, el tiempo de vaciado de las colas en ERICA es considerablemente mayor al observado para el CAPAC. Es de esperar que este comportamiento tenga una repercusión negativa sobre el retardo de las tramas de la capa superior.

Por otra parte, la rapidez con la que el algoritmo ERICA ajusta el ACR de las fuentes después de un cambio de ancho de banda puede dar lugar a un comportamiento inestable del algoritmo. Este comportamiento ha sido observado recientemente por Vidal en [Vid97], con tráfico TCP y colas limitadas en los conmutadores. La inestabilidad observada se debe a que TCP puede generar ráfagas de corta duración. Mientras esto sucede, el algoritmo no es capaz de estimar de forma exacta la PABE, lo cual da lugar a oscilaciones que impiden obtener unas prestaciones aceptables. Como solución se ha propuesto un filtrado paso bajo de los parámetros que estima el algoritmo. Un efecto adicional observado al aplicar el filtrado es la reducción en el tiempo de vaciado de las colas*.

Agradecimientos

Este trabajo ha sido financiado por la Comisión Interministerial de Ciencia y Tecnología (CICYT) a través del proyecto CICYT TIC96-0680.

Referencias

[Bar94] A.W. Barnhart, "Explicit Rate Performance Evaluations," ATM Forum/94-0983, October 1994.

[Ber92] D. Bertsekas and R. Gallager, "Data Communications," second edition, Prentice Hall, 1992.

[Bon95] F. Bonomi and K.W. Fendick, "The Rate-Based Control Framework for the Available Bit Rate ATM Service," IEEE Network, pp. 25-39, March/April 1995.

[Cha94] A.Charny, D. Clarck and R.Jain, "Congestion Control with Explicit Rate Indication," ATM Forum/94-0692, July 1994.

[For95] The ATM Forum Technical Committee, "LAN Emulation over ATM," Version 1.0, Implementation Agreement, 1995.

[For96] The ATM Forum Technical Committee, "Traffic Management Specification. Version 4.0," Forum/95-0013R11, March 1996.

[I.371] ITU-T Draft Recommendation I.371, "Traffic Control and Congestion Control in B-ISDN," May 1996.

[Jai95] R. Jain et al., "A Sample Switch Algorithm," ATM Forum/95-0178R1, February, 1995.

[Jai96] R. Jain, et al., "ERICA Switch Algorithm: A Complete Description," ATM Forum/96-1172, August, 1996.

[Jai96a] R. Jain, et al., "Source Behabior for ATM ABR Traffic Management," IEEE Communications Magazine, pp. 50-57, November 1996.

[Mar96] J. Martinez, M. Llop and J.M. Galindo, "Support of Non-Real Time Networked Multimedia Systems in ATM Based Networks," Proceedings of the 3rd International Workshop on Protocols for Multimedia Systems (PROMS'96), October 1996.

[Rob94] L. Roberts, "Enhanced PRCA (Proportional Rate-Control Algorithm)," ATM Forum/94-0735R1, August 1994.

[Rom95] A. Romanow and S. Floyd, "Dynamics of TCP Traffic over ATM Networks," IEEE J. Select. Areas Commun., vol.13, n.4, pp.633-641, May 1995.

[Sim94] R.J.Simcoe, "Test Configurations for Fairness and Other Tests," ATM Forum/94-0557, July 1994.

[Siu94] K-Y. Siu and H-Y.Tzeng, "Adaptive Proportional Rate Control with Intelligent Congestion Indication," ATM Forum/94-0888, September 1994.

[Vid97] J.R. Vidal, "Evaluación de Prestaciones Mediante Técnicas de Descripción Formal de la Emulación de Red Local sobre ATM," Tesis Doctoral, E.T.S.I. de Telecomunicación, Universidad Politécnica de Valencia, 1997.

[Yin94] N. Yin., "Fairness Definition in ABR Service Mode," ATM Forum/94-0928R2, 1994.

* En un estudio posterior se pretende realizar un análisis del algoritmo ERICA con los nuevos mecanismos introducidos en un escenario similar al descrito en este trabajo y compararlo con los algoritmos aquí descritos.

Control de congestión en redes de banda ancha tipo ATM

Antonio Barba, Eulàlia Mèlich
Dpto. Matemática Aplicada y Telemática. ETSETB
Universitat Politècnica de Catalunya.
c/ Girona Salgado 1-3, Modulo C-3 Campus Nord, 08034 Barcelona.
email: telabm@mat.upc.es

Abstract:

Broadband area networks require new mechanisms related to resource control in order to allow in real time to manage the traffic with the objective to avoid congestion situations. In this article a switching node models integrated in a broadband network are presented and studied. These nodes are designed in order to be able to support specifically different types of traffic in ATM cell networks. With the idea to provide an answer to the network management problem with the ABR traffic (burst traffic) in this type of networks, different mechanisms of traffic and congestion control are described. Finally, it is shown a congestion control algorithm and is evaluated its behaviour according to the network model proposed and diverse scenarios of traffic requirements.

1. Introducción

En esta década, las redes de comunicaciones han experimentado un increíble aumento de la velocidad de transmisión que permite transmitir grandes volúmenes de información en tiempos relativamente pequeños. Esto requiere de un control preciso del tráfico que circula por estas redes.

Frente al peligro de congestión, que se produce cuando el tráfico a la entrada de la red excede la capacidad de ésta y, por lo tanto, hay una pérdida de información, son necesarios mecanismos que permitan la recuperación del estado de equilibrio de la red. Debido a esta elevada velocidad de transmisión, cuando se produce la congestión en un punto de la red, mientras se envia una notificación hacia la fuente, ésta continua emitiendo datos. Esto es especialmente crítico en las redes de gran tamaño donde las distancias son grandes y cuando la información de realimentación llega a la fuente, ésta ya ha transmitido una gran cantidad de datos, agravando aún más la situación de congestión.

La gestión de tráfico en las redes de banda ancha tipo ATM está relacionada con la habilidad de la red para proporcionar apropiadamente diferentes calidades de servicios para las aplicaciones, así como la protección de la red y de los sistemas finales frente a la congestión y a la promoción de un uso eficiente de los recursos de la red.

Con la intención de conseguir estos objetivos se dispone de una serie de mecanismos para el control del tráfico y de la congestión en la red. El control de tráfico consiste en un conjunto de acciones realizadas por el sistema de gestión de la red para evitar que se produzca la congestión, y el control de congestión son las acciones también realizadas por la red para minimizar la duración e intensidad de la congestión cuando ésta ya se ha producido.

El marco de funcionamiento de estos controles viene determinado por el modelo de la arquitectura de la red de banda ancha que se presenta a continuación.

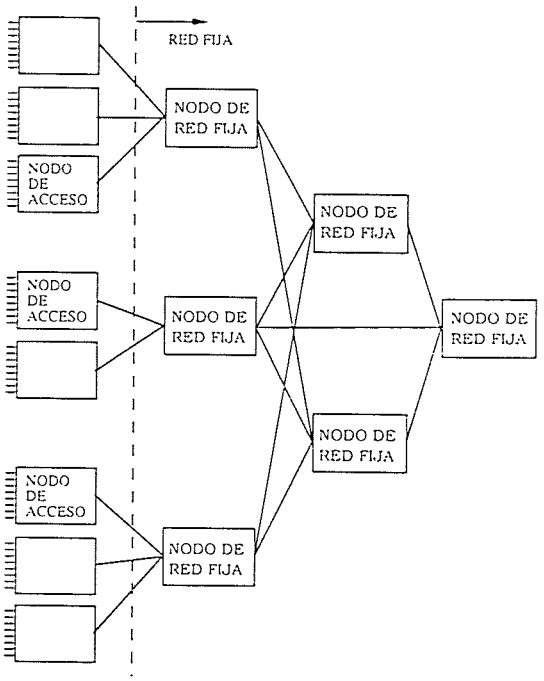


Fig. 1. Arquitectura de la red.

El sistema ofrece dos redes diferenciadas. Una sería la red de acceso formada por nodos con conexión directa a los usuarios abonados. Se trataría de comunicaciones que serían agrupadas de alguna forma hasta los 155 Mbps. Por otro lado está la red fija, con otro tipo de nodos, interconectados entre sí de forma mallada definiendo diversas jerarquías que permiten a su vez un aumento de la fiabilidad del sistema. Los enlaces de estas conexiones pueden transportar tráfico multiplexado hasta los 622 Mbps e incluso los 2,4 Gbps.

Los nodos se han diseñado de forma que puedan procesar de manera explícita diversos tipos de tráfico independientemente. A tal efecto se disponen

estructuras de colas M/D/1 con prioridades distintas según el tipo de tráfico. Ello permite definir tráfico CBR, VBR, ABR y también señalización (celdas RM inclusive).

Para obtener los objetivos definidos previamente, el resto del artículo se estructura de la siguiente forma. En el apartado dos se presenta el modelo matemático utilizado para definir los nodos de la red. En el apartado tres se muestran los mecanismos de control de congestión en las redes ATM. En el cuarto se presenta el algoritmo de control de congestión. Finalmente se muestran los resultados en forma de diversas gráficas y las conclusiones.

2. Modelado de los nodos de conmutación de una red ATM

Se ha utilizado un modelo de nodo diferenciado para implementar los nodos situados en la red de acceso respecto de los nodos de la red fija. En ambos nodos se han dispuesto colas del tipo M/D/1 para representar las diferentes etapas de procesado de la información en las celdas. Este modelo de cola, por su sencillez, es el más apropiado en el caso de redes ATM donde el tamaño de las celdas es fijo y por lo tanto el tiempo de servicio es determinista.

Se distinguen tres etapas dentro de un nodo. En la primera se procesan las cabeceras de señalización enrutando a diferentes colas y procesadores según el tipo y/o prioridad del tráfico. Las colas se han definido independientes para facilitar el cálculo y delimitar claramente los distintos efectos relativos al control de la congestión en el sistema. En la segunda etapa, se efectúa la conmutación propiamente dicha de las celdas a las rutas correspondientes, en este caso se ha utilizado un modelo de cola del mismo tipo con procesador para las funciones de procesamiento requeridas por las celdas. En la tercera fase se realiza la agrupación de la información y se establece de nuevo la señalización de las celdas según los trayectos y rutas definidos. El acceso al medio de transmisión se realiza según un sistema de prioridades.

El tipo de tráfico a la entrada de las conexiones sigue un proceso de Poisson modulado de Markov (MMPP) [4]. El tráfico en las conexiones CBR y VBR proviene de la multiplexación estadística de un conjunto de fuentes de Poisson. Para el tráfico ABR se trabaja con la suma de M minifuentes idénticas de dos estados (on, off) que tienen una tasa promedio de:

$E(\lambda) = MpA$, donde p es la probabilidad de que una minifuente esté en el estado on y A es el factor de compresión de la señal de video en bits/pixel. Esto permite disponer de una tasa variable para simular las ráfagas de celdas del tráfico ABR.

Para el caso de una cola M/D/1 el número medio de unidades (celdas) en la cola, incluyendo la que está en servicio, es el siguiente:

$$E(n) = \frac{\rho}{(1-\rho)}(1-\frac{\rho}{2}) \qquad (1)$$

y el retardo medio en el sistema:

$$E(T) = \frac{\frac{1}{\mu}}{(1-\rho)}(1-\frac{\rho}{2}) \qquad (2)$$

donde $\rho = \frac{\lambda}{\mu}$ es la intensidad de tráfico, lambda es la tasa promedio de celdas que llegan a la cola y $\frac{1}{\mu}$ es el tiempo medio de servicio de una celda. Ambas expresiones están relacionadas por la fórmula de Little: $E(n) = \lambda E(T)$

Para la elaboración de los modelos de los nodos se ha utilizado un sistema de prioridades. Se ha asignado la máxima prioridad al tráfico CBR ya que se trata de un tipo de tráfico interactivo que requiere una transmisión en tiempo real y que, por lo tanto, es extremadamente sensible al retardo. Dado que los otros tipos de tráfico (VBR y ABR) no tienen requerimientos de tiempo real se les ha asignado prioridades menores de forma que, en las etapas donde los tres tipos de tráfico comparten un mismo procesador, las unidades de menor prioridad no pueden ser servidas hasta que no lo hayan sido previamente las que tienen una prioridad más alta. Finalmente se ha considerado el tráfico de señalización.

Con todas estas hipótesis, haciendo uso de las fórmulas de las colas M/D/1 y a partir de la estructura de los nodos, se ha llegado a las siguientes expresiones: a la entrada, en la primera etapa del nodo de acceso es donde se separa la información en cuatro colas correspondientes a los cuatro tipos de tráfico distintos. Calculamos primeramente la tasa promedio de celdas a la entrada de cada una de estas colas:

$$\lambda_C = a_0\lambda_0 + a_1\lambda_1 + \ldots + a_n\lambda_n$$
$$\lambda_V = b_0\lambda_0 + b_1\lambda_1 + \ldots + b_n\lambda_n$$
$$\lambda_a = c_0\lambda_0 + c_1\lambda_1 + \ldots + c_n\lambda_n$$
$$\lambda_S = s_0\lambda_0 + s_1\lambda_1 + \ldots + s_n\lambda_n$$

donde λ_c es la tasa promedio total de tráfico CBR debido a todas las n conexiones de entrada al nodo. Igualmente $\lambda_v, \lambda_a, \lambda_s$ son las tasas promedio totales de tráfico VBR, ABR y de señalización respectivamente debidas a todas la conexiones de entrada al nodo.

λ_i con i = 0,.....,n. es la tasa promedio en la entrada "i" del nodo (donde llegan multiplexados los distintos tipos de tráfico). a_i con i = 0,.....,n. indica el porcentaje de tráfico CBR en la conexión de entrada "i". Igualmente b_i, c_i, s_i con i = 0,.....,n. indican el porcentaje de tráfico VBR, ABR y de señalización respectivamente en la conexión "i".

Las tasas a la entrada de las colas de la segunda etapa son de nuevo λ_c, λ_v, λ_a y λ_s en el caso de que no se produzcan pérdidas. En el caso de que se congestione una cola de una etapa, el tráfico a la salida de ésta será el máximo tráfico que es capaz de cursar el procesador del conjunto cola-procesador. El resto de las celdas se perderán. En las fórmulas que a continuación se detallan se considera el caso en el que no se producen pérdidas, para simplificar así la notación. No obstante, en el programa simulador implementado, se consideran las pérdidas en caso de congestión de una cola y se actualiza la tasa a la salida de la misma.

A la entrada de las colas de la tercera etapa volvemos a tener la información multiplexada, siendo λ_{Ti} la tasa total a la entrada de la tercera etapa para la salida "i" del nodo. El porcentaje de tráfico CBR, de señalización, VBR y ABR de usuario en cada conexión de salida viene dado por los parámetros x, y, z respectivamente.

Una vez detallado todo esto ya podemos calcular el retardo medio de conmutación de un nodo para cada tipo de tráfico, sumando los retardos medios de las tres etapas. En la primera y segunda etapa utilizamos la ya anteriormente mencionada fórmula (2).El cálculo para la tercera etapa requiere un estudio más detallado de las distintas prioridades de las colas. Se ha utilizado la siguiente fórmula del tiempo medio de espera en cola de una unidad en la cola k-esima de un conjunto de N colas que comparten un mismo procesador:

$$\overline{W}_{qk} = \frac{\bar{t}_{residual}}{(1-\sum_{i=1}^{k}\rho_i)(1-\sum_{i=1}^{k-1}\rho_i)}$$

donde $\bar{t}_{residual} = \sum_{i=1}^{N}\rho_i \frac{\overline{x^2}_i}{2\overline{x}_i}$ es el tiempo residual, es decir, es el tiempo de espera a que el procesador acabe de servir la unidad en curso.

Para el caso de una cola M/D/1 la $\sigma^2 = 0$,con lo que $\overline{x^2}_i = \overline{x}_i^{\,2}$ y el tiempo residual queda:

$$\bar{t}_{residual} = \sum_{i=1}^{N}\rho_i \frac{\overline{x}_i}{2}$$

Finalmente, particularizando para N=3 colas y sumando los términos de las tres etapas, se obtienen la siguientes expresiones para el nodo de acceso.

El retardo medio de conmutación para el tráfico CBR es:

$$E(T)_{a,C} = \frac{1}{(u1-\lambda_c)}(1-\frac{\lambda_c}{2u1})+\frac{1}{(u11-\lambda_c)}(1-\frac{\lambda_c}{2u11})+ \frac{\lambda_{Ti}}{2(u4^2-u4(x)(\lambda_{Ti}))}+\frac{1}{u4}$$

Para el tráfico VBR:

$$E(T)_{a,V} = \frac{(\lambda_v+\lambda_a)}{2(u2^2-u2(\lambda_v))}+\frac{1}{u2}+\frac{1}{(u22-\lambda_v)}(1-\frac{\lambda_v}{2u22})+ \frac{\lambda_{Ti}}{2(u4^2-u4(x+y+1)(\lambda_{Ti})+\lambda^2_{Ti}\ (x+y))}+\frac{1}{u4}$$

Para el tráfico ABR de usuario:

$$E(T)_{a,A} = \frac{(\lambda_v+\lambda_a)}{2(u2^2-u2(\lambda_a+2\lambda_v)+\lambda^2_v+\lambda_v\lambda_a)}+\frac{1}{u2}+ \frac{1}{(u33-\lambda_a)}(1-\frac{\lambda_a}{2u33})+ \frac{\lambda_{Ti}}{2(u4^2-u4(x+y+1)(\lambda_{Ti})+\lambda^2_{Ti}\ (x+y))}+\frac{1}{u4}$$

Para el tráfico ABR de señalización:

$$E(T)_{a,S} = \frac{1}{(u3-\lambda_s)}(1-\frac{\lambda_s}{2u3})+\frac{1}{(u44-\lambda_s)}(1-\frac{\lambda_s}{2u44}) + \frac{\lambda_{Ti}}{2(u4^2-u4(2x+y)(\lambda_{Ti})+\lambda^2_{Ti}\ (x^2+xy))}+\frac{1}{u4}$$

El número medio de unidades en toda etapa se obtiene multiplicando el retardo medio de cada etapa por la tasa promedio de entrada a dicha etapa (fórmula de Little).

Para el nodo de la red fija seguimos el mismo procedimiento. La diferencia radica en la primera etapa, donde no concentramos el tráfico de un mismo tipo de todas la conexiones de entrada en una misma cola (como hacíamos en el nodo de acceso), sino que establecemos tres colas diferentes para cada

conexión de entrada. Esto es debido a que los nodos de la red fija soportan un tráfico elevado y no nos interesa concentrarlo porque entonces tendríamos rápidamente problemas de congestión.

Al igual que en el caso anterior, llegamos a las siguientes expresiones para el nodo de la red fija:

El retardo medio de conmutación para el tráfico CBR es:

$$E(T)_{f,C} = \frac{1}{(v1 - a_i\lambda_i)}\left(1 - \frac{a_i\lambda_i}{2v1}\right) + \frac{1}{(v11 - \lambda_c)}\left(1 - \frac{\lambda_c}{2v11}\right) + \frac{\lambda_{Ti}}{2(v4^2 - v4(x)\lambda_{Ti})} + \frac{1}{v4}$$

Para el tráfico VBR:

$$E(T)_{f,V} = \frac{1}{(v2 - b_i\lambda_i)}\left(1 - \frac{b_i\lambda_i}{2v2}\right) + \frac{1}{(v22 - \lambda_v)}\left(1 - \frac{\lambda_v}{2v22}\right) + \frac{\lambda_{Ti}}{2(v4^2 - v4(2x+y)(\lambda_{Ti}) + \lambda^2_{Ti}\,(x^2 + xy))} + \frac{1}{v4}$$

Para el tráfico ABR de usuario:

$$E(T)_{f,A} = \frac{1}{(v3 - \lambda_i(c_i + s_i))}\left(1 - \frac{\lambda_i(c_i + s_i)}{2v3}\right) + \frac{1}{(v33 - \lambda_a)}\left(1 - \frac{\lambda_a}{2v33}\right) + \frac{\lambda_{Ti}}{2(v4^2 - v4(x+y+1)(\lambda_{Ti}) + \lambda^2_{Ti}\,(x + y))} + \frac{1}{v4}$$

Para el tráfico de señalización:

$$E(T)_{f,S} = \frac{1}{(v3 - \lambda_i(c_i + s_i))}\left(1 - \frac{\lambda_i(c_i + s_i)}{2v3}\right) + \frac{1}{(v44 - \lambda_s)}\left(1 - \frac{\lambda_s}{2v44}\right) + \frac{\lambda_{Ti}}{2(v4^2 - v4(x+y+1)(\lambda_{Ti}) + \lambda^2_{Ti}\,(x + y))} + \frac{1}{v4}$$

El número medio de unidades en el sistema se calcula como ya se ha indicado en el caso anterior.

Por otra parte el diseño de los nodos que componen la red consiste también en el dimensionado adecuado de sus procesadores y buffers.

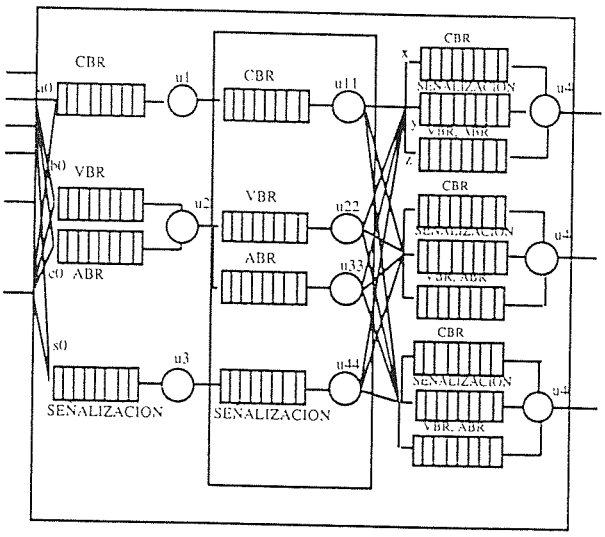


Fig. 2. Esquema del nodo de acceso

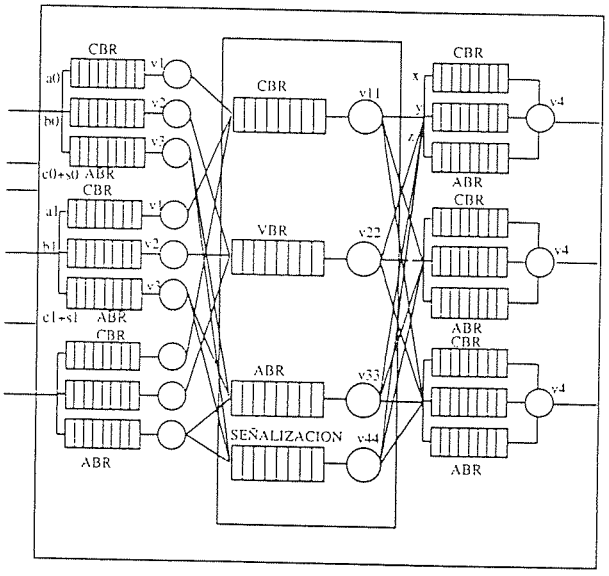


Fig. 3. Esquema de un nodo de la red fija.

Se han diseñado los procesadores para un escenario de tráfico en concreto que se ha considerado razonable. Los desequilibrios causados por variaciones en este tráfico son los que deberán ser corregidos por el algoritmo de control de congestión. Dicho escenario consta de un 50% de tráfico CBR respecto del total a la entrada, un 30% de tráfico VBR, un 10% de tráfico ABR y un 10% de tráfico de señalización [2, 3]. También se ha previsto que el tráfico ABR que contiene la información de usuario puede tener ráfagas de hasta 10 veces el valor preestablecido con el operador.

La capacidad de procesamiento definida para los procesadores depende del número de entradas del nodo. En el caso de un nodo de acceso de 8 entradas se han tomado los valores siguientes para garantizar un buen funcionamiento en el escenario propuesto de tráfico [5]:

u1 = u11 = 1,419.927 celdas/s, u2 = 1,135.942 celdas/s
u3 = u44 = 500.000 celdas/s, u22 = 851.956 celdas/s
u33 = 283.985 celdas/s, u4 = 354.981 celdas/s.

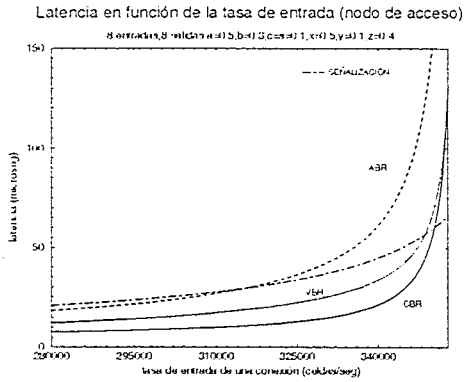
Se ha tomado un valor del procesador para el tráfico de señalización suficientemente potente para que no se pierdan celdas de señalización incluso en los casos de un fuerte desequilibrio.

De igual forma, se han tomado los siguientes valores para garantizar el buen funcionamiento de un nodo de la red fija con ocho entradas:

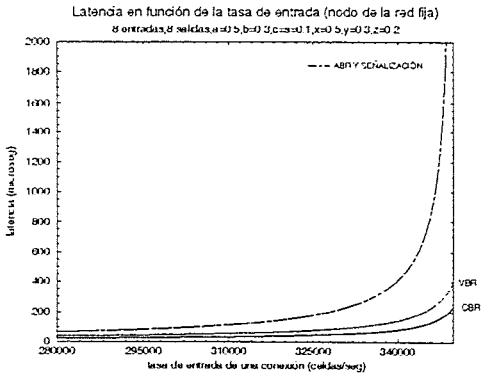
v1 = 177.490 celdas/s, v2 = 106494 celdas/s, v3 = 70.996 celdas/s
v11 = 1,419.927 celdas/s, v22 = 851.956 celdas/s,
v33 = 283.985 celdas/s, v4 = 354.981 celdas/s, v44 = 470.000 celdas/s.

Con los nodos dimensionados hemos estudiado su comportamiento mediante simulación para ver si se corresponde al comportamiento esperado para conmutadores reales ATM.

A continuación se presentan las gráficas de la latencia del nodo de acceso y de la red fija para los distintos tipos de tráfico.

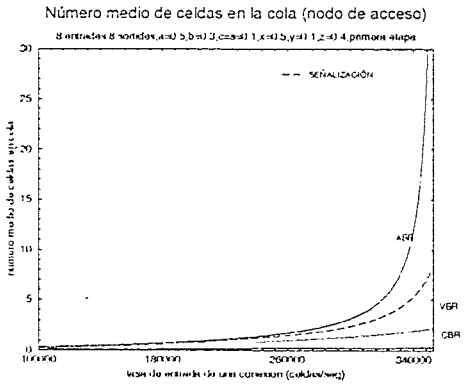


Gráfica 1. Latencia del nodo de acceso

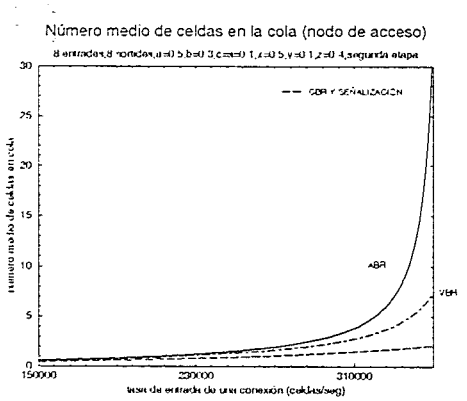


Gráfica 2. Latencia del nodo de la red fija

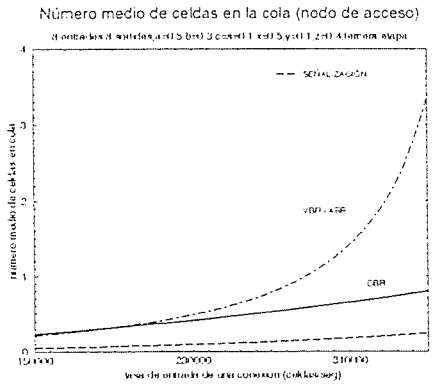
El diseño de los buffers se ha establecido en función de un retardo máximo y mínimo tolerable para los distintos tipos de tráfico y las calidades de servicio especificadas en las recomendaciones. Como que los tráficos CBR y VBR no toleran retardos grandes disponen de procesadores más potentes, con lo cual el número de celdas que esperan en las colas es pequeño y es suficient utilizar buffers de tamaño reducido (de unas 10 celdas, como puede observarse en las gráficas del número de celdas en las colas que se presentan a continuación). En cambio, para el tráfico ABR son necesarios buffers mayores ya que presenta fuertes desviaciones respecto su valor medio, con lo cual el tamaño del buffer debe ser suficiente para absorber estas ráfagas.



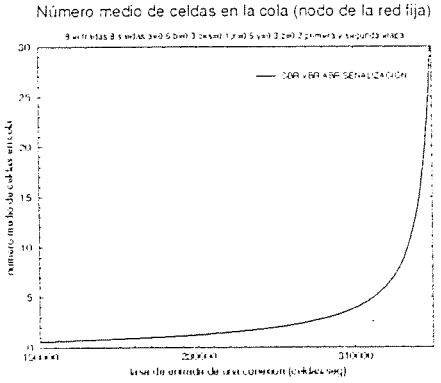
Gráfica 3. Número medio de celdas en las colas de la primera etapa del nodo de acceso.



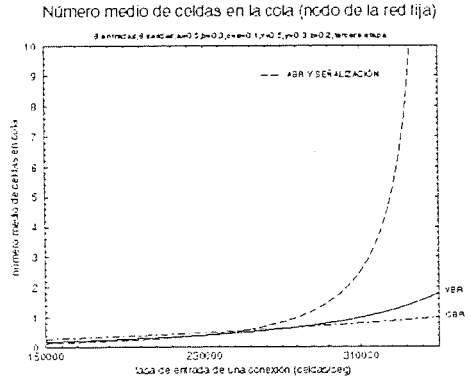
Gráfica 4. Número medio de celdas en las colas de la segunda etapa del nodo de acceso.



Gráfica 5. Número medio de celdas en las colas de la tercera etapa del nodo de acceso.

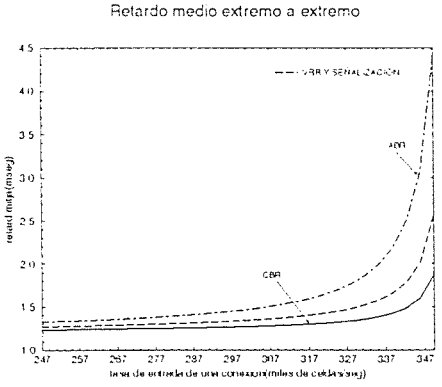


Gráfica 6. Número medio de celdas en las colas de la primera y segunda etapa del nodo de red fija.



Gráfica 7. Número medio de celdas en las colas de la tercera etapa del nodo de red fija.

La siguiente gráfica muestra la evolución del retardo extremo a extremo entre dos puntos de la red a medida que se aumenta la tasa de tráfico a la entrada. El retardo mayor se obtiene para el tráfico ABR que resulta ser el que tiene menores restricciones. El retardo menor es el del tráfico CBR, ya que es el que tiene requerimientos de tiempo real.



Gráfica 8. Retardo extremo a extremo en función de la tasa de entrada.

De la observación de todas estas gráficas se concluye que el comportamiento de los nodos que hemos diseñado se corresponde con el comportamiento de conmutadores reales.

3. Control de congestión

Mientras que en las comunicaciones en tiempo real, como es el caso de les servicios CBR y VBR se usa un mecanismo de control de congestión preventivo y de lazo abierto, para el tráfico a ráfagas de tipo ABR, que resulta ser el mayor problema para disponer de una red sin congestión, el ATM Forum ha propuesto un mecanismo reactivo de lazo cerrado, el cual regula dinámicamente la tasa de emisión de celdas de cada conexión ABR, usando la información de realimentación que le llega de la red.

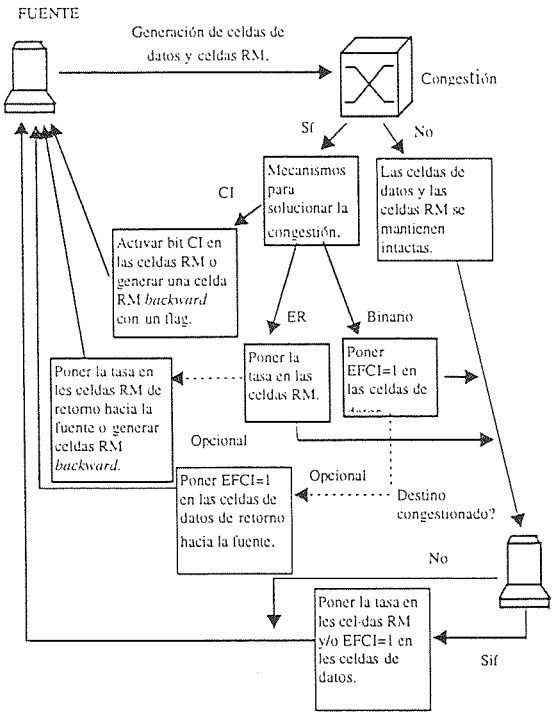


Fig. 4. Mecanismos de control de congestión.

La red puede enviar a los usuarios la información de su estado de congestión de distintas formas. Una sería proporcionando una indicación binaria sobre si se ha producido o no la congestión activando un bit en la cabecera de las celdas de datos. Este bit es llamado EFCI (*Explicit Forward Congestion Indication*) si se activa en las celdas que se dirigen en dirección al destino y EBCI (*Explicit Backward Congestion Indication*) si se dirigen en el otro sentido del enlace bidireccional, es decir, hacia la fuente.

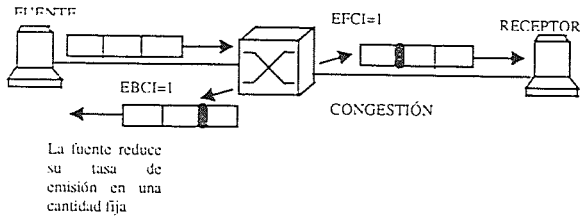


Fig. 5. Notificación binaria de la congestión.

Otra forma sería utilizando las celdas RM (*Resource Management*) que pueden ser generadas periódicamente por la fuente y cuyos campos pueden ser modificados por los nodos conmutadores por los cuales pasa. Si un nodo se halla congestionado bien puede activar el bit CI de la celda RM que pasa a través suyo o puede indicar la tasa máxima que puede soportar en el campo ER (*Explicit Rate*) de esa misma celda.

En ambos casos, cuando la fuente recibe una notificación de congestión adapta su tasa de emisión de acuerdo con la información recibida y se reduce así el flujo de tráfico.

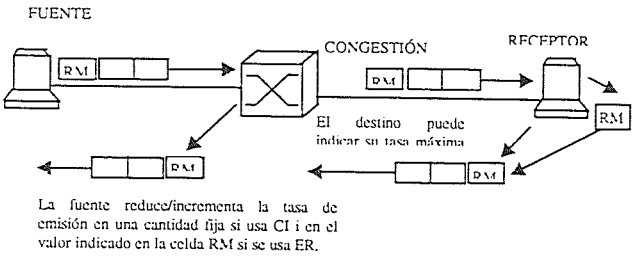


Fig. 6. Empleo de las celdas RM.

4. Algoritmo de control de congestión

Para reducir el nivel de congestión en la red se ha implementado una algoritmo en los nodos de acceso que actua sobre el tráfico a ráfagas de tipo ABR, cuyo funcionamiento se expone a continuación. En cada nodo se analiza la ocupación de los buffers de tráfico ABR. Si en algún caso la longitud de la cola sobrepasa un determinado umbral, el nodo descarta las celdas de baja prioridad (CLP=1). Estas celdas pueden ser marcadas inicialmente por la fuente emisora como celdas de baja prioridad, (usadas por ejemplo para aumentar la calidad de una imagen), o pueden ser celdas que no cumplan el contrato de tráfico establecido entre el usuario y la red. Frente al peligro de congestión, dichas celdas pueden ser selectivamente eliminadas por el nodo congestionado para proteger las prestaciones del tráfico de mayor prioridad.

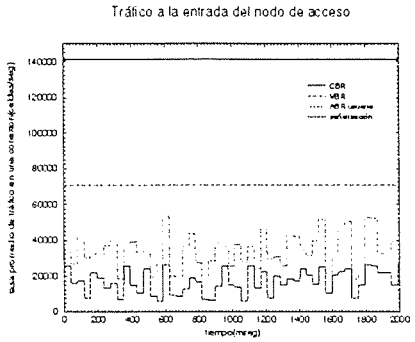
En el caso de que el tráfico a la entrada del nodo de acceso consista en ráfagas que saturen completamente los buffers, el citado mecanismo no es suficiente para solucionar la congestión en el nodo y evitar la pérdida de celdas prioritarias. Entonces se aplica una realimentación hacia las fuentes de información con el fin de controlar el flujo de celdas a la entrada del nodo de acceso. En ésta, el nodo saturado activa el bit EFCI/EBCI en la cabecera de las celdas de datos de las conexiones que pasan por ese nodo. Cuando la fuente recibe celdas con EBCI=1 reduce su tasa de emisión. Esta reducción del tráfico a la entrada, debe asegurar siempre la calidad de servicio contratada por el abonado, (no estando nunca por debajo de la mínima tasa contratada por éste), y permite solucionar en buena medida la congestión en la red. Se ha adoptado el método de notificación binaria para indicar la presencia de congestión en la red porque es la forma más rápida, ya que utiliza las propias celdas de datos.

Los dos parámetros críticos en el funcionamiento del algoritmo son el retardo de la realimentación hacia la fuente y la mayor o menor brusquedad con la que ésta reduce su tasa de emisión. Aquí se ha implementado una reducción brusca hasta el mínimo valor contratado por el usuario porque se pretende obtener la máxima eficacia.

La velocidad de recuperación de la congestión depende del retardo de la realimentación. Éste, a su vez, es función del tiempo de procesado del algoritmo de gestión y de los parámetros característicos de la propia red: potencia de los procesadores de los nodos y velocidad de transmisión y propagación de la información.

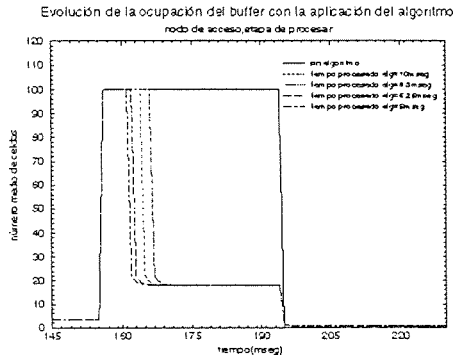
Cuanto mayor es la rapidez de la realimentación, menor es el número de celdas perdidas ya que se tarda menos en notar el efecto del algoritmo. Éste se percibe en el incremento de la calidad de servicio (QOS): reducción en el porcentaje de pérdidas de celdas y en el retardo de transferencia de éstas.

A continuación se ilustra el comportamiento del algoritmo con una serie de gráficas.

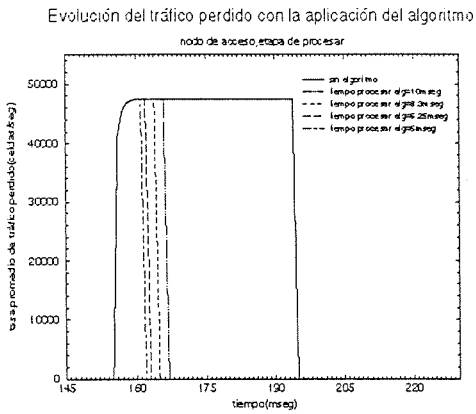


Gráfica 9. Tráfico a la entrada del nodo de acceso

En la gráfica 9 se muestra la evolución del tráfico a la entrada del nodo de acceso.



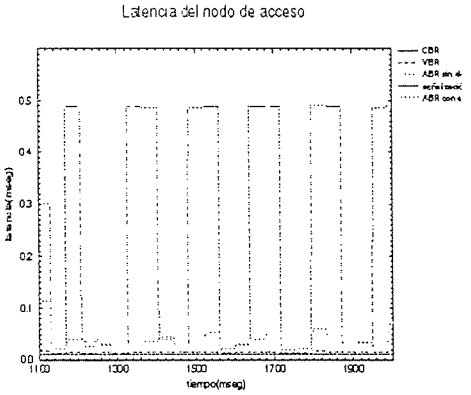
Gráfica 10. Evolución de la ocupación del buffer ABR



Gráfica 11. Evolución del tráfico perdido

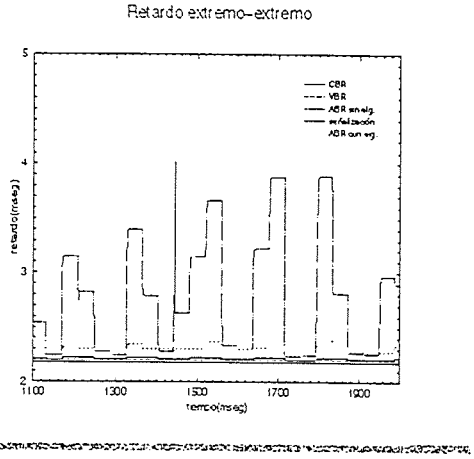
En la gráfica 10 se observa la reducción del número de celdas en el buffer del tráfico ABR de usuario al aplicar el algoritmo. Estos resultados corresponden a la simulación realizada con un tamaño de buffer ABR de 100 celdas. Cuando no se aplica el algoritmo, el buffer permanece saturado durante todo el intervalo de tiempo que dura la ráfaga de entrada. Cuando se aplica el algoritmo, después del tiempo necesario para que llegue a la fuente la notificación de congestión y de que ésta disminuya la tasa de celdas enviadas, el número de celdas en el buffer disminuye, pasando de 100 (saturación del buffer) a 20. Las distintas curvas corresponden a valores diferentes del retardo de realimentación. El buffer permanecerá más tiempo saturado cuanto más lenta sea dicha realimentación y esto supondrá unas pérdidas mayores como puede observarse en la gráfica 11. Cuando no actúa el algoritmo la tasa de celdas perdidas se mantiene durante todo el intervalo que dura la ráfaga de tráfico a la entrada. Cuando actúa el algoritmo, en el momento en que la fuente disminuye la tasa enviada, disminuyen también las pérdidas. Debido a que la reducción de la tasa de la fuente es brusca, también lo es la caída de la tasa de pérdidas.

De aquí, se deduce que el parámetro crítico de funcionamiento del algoritmo es el retardo de realimentación. Dicho retardo consta de varios términos. El más significativo es el tiempo de procesado del algoritmo de gestión que varia en función de la potencia del procesador de la señalización de la segunda etapa del nodo de acceso. Este término es del orden de algunos milisegundos, mientras que los términos debidos a los retardos de transmisión y propagación son del orden de los microsegundos, ya que la velocidad de transmisión es elevada y la distancia entre la fuente y el nodo de acceso es del orden de kilómetros. Si el algoritmo de control de congestión se implementase en los nodos de la red fija, el término del retardo de propagación tendría mayor peso ya que las distancias serían mayores.



Gráfica 12. Latencia del nodo de acceso

En las gráficas 12 y 13 se observa la latencia del nodo de acceso y el retardo extremo a extremo para un determinado origen y destino dentro de la red, cuando no se aplica el algoritmo y después de su aplicación. .La reducción de estos retardos contribuye a aumentar la calidad de servicio (QOS) del sistema.



Gráfica 13. Retardo extremo a extremo

Finalmente se han realizado diversas simulaciones para estudiar el porcentaje de celdas perdidas en el nodo de acceso en función del tamaño de los buffers (de la segunda etapa) del tráfico ABR y también en función del umbral de congestión establecido en los mismos. Se define un ratio de pérdida de celdas como el cociente entre el número total de celdas pérdidas y el número total de celdas a la entrada del nodo de acceso.

Ratio de pérdida de celdas: Sin algoritmo

Buffer	100 celdas	84 celdas	72 celdas	63 celdas	50 celdas	5 celdas
Ratio %	6'15	6'18	6'2	6'23	6'29	8'59

Ratio de pérdida de celdas: Con algoritmo

Buffer	100 celdas	84 celdas	72 celdas	63 celdas	50 celdas	5 celdas
Ratio %	0'09	0'11	0'13	0'14	0'18	0'684

Ratio de pérdida de celdas: de acuerdo a un nivel de disparo en el buffer (tamaño de 100 celdas)

Nivel disp.	20%	30%	40%	50%	60%
Ratio %	0'132	0'094	0'058	0'039	0'039

5. Conclusiones

Se han presentado los esquemas correspondientes al modelo de nodo definido para una red de acceso y para una red fija ATM. En ambos casos se han realizado simulaciones a fin de estimar las condiciones de carga y los retardos que soportan los nodos para diferentes escenarios de tráfico. Se ha comprobado que los parámetros obtenidos están dentro de los márgenes de funcionamiento requeridos para este tipo de redes.

Se ha expuesto un marco de funcionamiento para los mecanismos de control de tráfico y congestión utilizados en redes de tipo ATM. Se ha hecho una valoración de los distintos procedimientos existentes y se ha propuesto un nuevo algoritmo que permite el control de congestión para tráfico ABR con unos ratios de prestaciones muy buenos.

6. Referencias

[1] Traffic Management Specification v. 4.0. The ATM Forum. 1996.

[2] Hiroshi Suzuki et al. *A burst traffic control strategy for ATM networks.* Globecom 1990, San Diego.

[3] Gillian M. Woodruff and Rungroj Kositpaiboon. *Multimedia traffic management principles for guaranteed ATM network performance.* IEEE Journal on selected areas in communications. April 1990.

[4] Mischa Schwartz. *Broadband Integrated Networks.* Prentice Hall PTR. 1996

[5] Othmar Kyas. *ATM Networks.* International Thompson Publishing. 1995.

[6] Jerry Banks et al. *Discrete event system simulation.* Prentice Hall. 1996.

[7] Antoni Barba, Eulàlia Mèlich. *Algoritmo para el control de congestión en redes ATM.* Yuforic 1997. Barcelona.

Red de Acceso

Soluciones de Red de Acceso para Operadores de Cable

Eva Parrilla, Belén Carro, Judith Redoli, Rafael Mompó
Dpto. de Teoría de la Señal y Comunicaciones e Ingeniería Telemática
ETSI Telecomunicación – Universidad de Valladolid
C/ Real de Burgos s/n 47011 Valladolid
{evapar; belcar}@tel.uva.es, {jredoli; jyr}@dvnet.es

Abstract:

Cable TV networks can be enhanced to be used not only for TV broadcasting but for bi-directional high speed digital data communications as well. A recently appeared device called 'cable modem' transmits digital data through the cable television plant. A cable modem connects the user's computer to the coaxial CATV cable providing high speed access to the Information Superhighways for residential users.

In this paper we explore the cable modem technology and propose it for linking the residential users to broadband networks.

We will first explain the performance characteristics of cable modems and then we will look more in depth into their physical and link layer characteristics. We will finally talk about the standardization efforts that are being driven and the actual position of these devices in the marketplace.

But , there are places where CATV networks are difficult to arrive. The solution to serve these places is radio access. Some research on these alternative methods is applied to rural areas.

1.-Introducción

Los módems de cable son una alternativa eficaz y económica para conectar a usuarios residenciales y pequeñas empresas a las autopistas de la información, cuyo mayor exponente se encuentra hoy en día en la red Internet.

Los módems de cable están en disposición de ofrecer el mejor acceso a Internet, pues ofrecen:

- *gran velocidad*: las redes de cable construidas en fibra óptica y coaxial explotan un gran ancho de banda
- *economía*: la capacidad de la red es compartida dinámicamente entre varios usuarios y aprovecha una infraestructura puesta en marcha para otros servicios (TV)
- *conexión permanente*: no se requiere marcado

Suelen ser dispositivos externos, que se conectan a una tarjeta Ethernet instalada en el PC y a la toma de cable. El módem es autoconfigurable y se gestiona remotamente. Nada más conectarlo a la red de cable, busca un canal que le permita entrar en comunicación con el órgano central de control que le configurará remotamente todos los parámetros de operación. Algunos de los cuales además se modifican periódicamente para adaptarse a las condiciones cambiantes de la red y del tráfico.

Estos dispositivos han sido concebidos para utilizarse en entornos particulares, por este motivo, a pesar de su complejidad resultan extremadamente sencillos de manejar, todos los detalles de operación son transparentes al usuario.

Antes de pasar a comentar las bases de la tecnología de los módems de cable, haremos una breve introducción de la topología de las redes sobre las que operan.

2.-Las Redes de TV por Cable

Las redes de TV por cable han sido diseñadas inicialmente para la distribución de canales de TV. Aunque algunas están construidas enteramente en cable coaxial, las más recientes cuentan con una topología híbrida de fibra-coaxial o HFC (Hybrid Fiber-Coax).

Una red típica de CATV consta de tres etapas (Fig 1.):

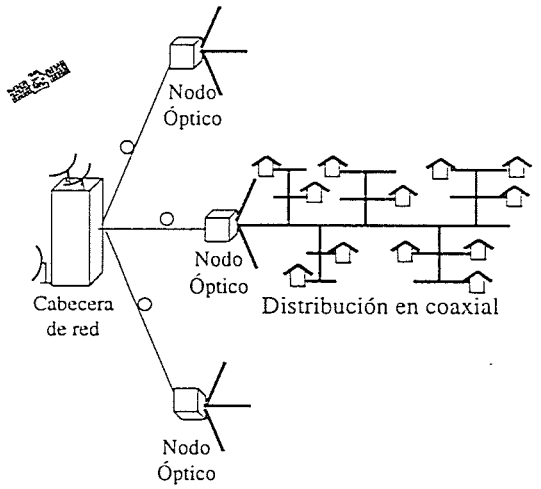


Fig. 1: Topología típica de red de CATV

1. Cabecera o *Head End*: es el lugar donde se recopilan todos los canales de TV, provenientes de satélites, enlaces terrestres o de estudios de producción propia, y se combinan para ser transmitidos por la red

2. Estrella en fibra óptica: varias ramas de fibra parten de la cabecera, llevando la señal hasta los nodos ópticos, donde la señal se transforma de óptica a eléctrica

3. Árbol-rama en cable coaxial: conjunto de ramificaciones de cable que lleva la señal hasta los abonados

Tradicionalmente estas redes se han explotado de forma unidireccional, difundiendo señales de TV desde cabecera hacia los usuarios en la banda comprendida entre 50 y 860MHz aproximadamente. Sin embargo, existe la posibilidad de utilizar estas redes de forma bidireccional, activando lo que se conoce como *canal de retorno*, que comprende aproximadamente de 5 a 45MHz y permite transportar señales procedentes de los usuarios hasta la cabecera de red.

Es importante llamar la atención sobre el hecho de que el *canal descendente* es punto a multipunto, mientras que el *canal ascendente* o de retorno es multipunto a punto. Las señales procedentes de los usuarios se sumarán en su camino hacia la cabecera al recorrer la estructura árbol-rama del coaxial.

Este factor puede acarrear algunos problemas en cuanto a ruido o capacidad, que se comentarán más adelante.

Buena parte de los mismos se pueden solucionar con un buen diseño de la red HFC, que le permita evolucionar al ritmo de la demanda de los nuevos servicios. Disminuyendo el tamaño de los nodos ópticos o multiplexando los canales de retorno de cada rama de coaxial podemos llegar a tener un canal de retorno (40MHz) para cada 200 usuarios, por ejemplo [1].

3.-Nivel Físico de los Módems de Cable

Los *módems de cable* pueden hacer un uso eficiente de esta infraestructura permitiendo comunicaciones de datos a alta velocidad.

Los canales de TV se difunden multiplexados en frecuencia en el canal descendente, ocupando 6 u 8 MHz (en América y Europa respectivamente) cada canal analógico. En el espacio de cada canal analógico se pueden ubicar 4 o más canales digitales, dependiendo de la compresión utilizada.

Los módems de cable transmiten por una portadora situada en el canal ascendente y escuchan otra portadora ubicada en el canal descendente. Un órgano situado en la cabecera comunica estas dos portadoras, permitiendo que los módems se puedan comunicar entres sí.

Todos los módems son sintonizables, pueden colocar sus portadoras en cualquier zona del espectro que se encuentre libre. Algunos sistemas permiten incluso variar la frecuencia a la que están funcionando los módems dinámicamente sin interrumpir el servicio. Esto puede resultar muy útil para esquivar el ruido en el canal ascendente o simplemente para repartir la carga de tráfico entre distintas portadoras.

Se pueden distinguir dos tipos de sistemas:

- **Sistemas simétricos:** las portadoras ascendente y descendente son del mismo ancho (Fig.2)
- **Sistemas asimétricos:** las portadoras ascendentes son más estrechas que las descendentes. Este sistema se adapta mejor a las características de la red HFC, que intrínsecamente es asimétrica (Fig.3)

En los sistemas simétricos se suelen encontrar portadoras de 6MHz con modulaciones BPSK o QPSK, lo que supone velocidades de 4 o 10Mbps respectivamente en ambas direcciones [7].

Los sistemas asimétricos, que se están imponiendo masivamente, utilizan portadoras ascendentes de ancho variable entre 0,2 y 3,2MHz y modulación QPSK o 16QAM. En estas condiciones la tasa binaria ascendente oscila entre 320Kbps y 10Mbps, pudiendo tomar valores intermedios entre estos extremos [4].

Se persigue que el ancho y modulación de la portadora puedan variar dinámicamente adaptándose a las características cambiantes del canal de retorno.

El canal descendente es de 6MHz y modulación 64QAM, lo que supone 30Mbps de la cabecera hacia los usuarios. Si el canal está en buenas condiciones, con 256QAM se pueden alcanzar los 40Mbps [4].

Habitualmente en los sistemas asimétricos encontramos varias portadoras ascendentes asociadas a una misma portadora descendente (Fig.3), mientras que en los sistemas simétricos esta asociación es siempre uno a uno. En cualquiera de los dos casos, a medida que aumente el número de usuarios y el tráfico cursado se pueden ir habilitando portadoras adicionales para darle cabida.

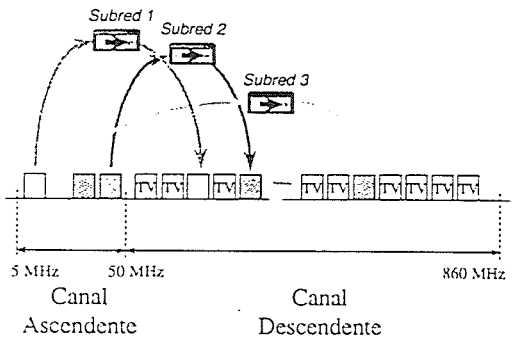


Fig. 2: Organización del espectro en un sistema simétrico

Por tanto, los sistemas de módems de cable permiten comunicaciones entre distintos puntos de la red de CATV a velocidades de Mbps. Estas redes suelen ser de ámbito metropolitano, con lo que podemos conectar puntos situados a varias decenas de kilómetros. Estas son características diferenciales de este servicio, tipo red de área local pero con extensión de área metropolitana.

Aunque sin duda constituyen una ventaja importante, las grandes distancias añaden un grado más de complejidad al diseño de estos dispositivos. Por una parte el nivel de potencia se tendrá que ajustar en cada caso a la distancia que tenga que recorrer la señal hasta llegar a la cabecera para llegar a la misma con un nivel aceptable de C/N. Por otra, será necesario establecer algún mecanismo para mantener el sincronismo a estas velocidades y distancias. De forma similar a los sistemas por satélite, cada estación deberá conocer el retardo de su señal hasta cabecera para transmitir con suficiente antelación.

Como hemos visto, los módems requieren información precisa acerca de su modo de funcionamiento. El órgano de cabecera se encarga de enviarles periódicamente información detallada acerca de los parámetros que deberán utilizar.

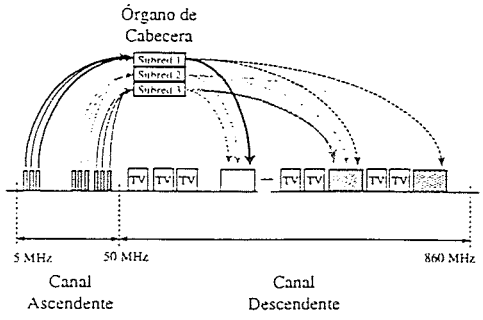


Fig. 3: Organización del espectro en un sistema asimétrico

4.-Acceso a un Medio Común

La estructura árbol-rama de las redes de CATV permite utilizar el medio de transmisión de forma compartida. Varios módems estarán sintonizados a cada portadora ascendente y descendente repartiéndose dinámicamente la capacidad de las mismas. Esto constituye una solución económica para disponer de una red de acceso de alta velocidad.

Sin embargo, esto sólo será posible si se utiliza un método de acceso al medio eficiente. La estrategia aplicada para compartir el ancho de banda determinará la calidad del servicio final ofrecido.

Partimos de la base de que la capacidad deberá ser asignada dinámicamente bajo demanda. Asignaciones estáticas reducirían las prestaciones, especialmente para las comunicaciones de datos, que inicialmente son el principal objetivo de estos sistemas.

Las dos soluciones que se han planteado para solucionar el problema del acceso al medio son la contienda y el método de solicitud-reserva.

5.-Acceso Basado en Contienda

Los primeros módems de cable se diseñaron para conectar redes de área local situadas en distintos puntos de una misma ciudad aprovechando la infraestructura de la red de CATV.

Con este objetivo en mente, los primeros módems basaban su funcionamiento en adaptar el protocolo de una red de área local tipo *Ethernet* a la red de cable.

Las redes tipo Ethernet utilizan un esquema CSMA/CD (*Carrier Sense Multiple Access/Collision Detection*) para controlar el acceso al medio. Según este método, cuando una estación tiene datos para transmitir, en primer lugar escucha el medio, si percibe que no está ocupado, comienza a transmitir. Su paquete de datos tardará un tiempo en llegar a las estaciones más alejadas de la red, que durante ese período perciben el medio como libre y pueden empezar a transmitir también. Si esto ocurriese, ambas emisiones colisionarían, perdiéndose los dos paquetes, que tendrían que ser retransmitidos.

Este método resulta muy eficiente para comunicaciones de datos en redes de área local, pero al extrapolarlo a las redes de CATV surgen varios problemas adicionales.

En primer lugar hay que lograr que los terminales transmitan y escuchen un mismo medio para poder detectar las colisiones. Esto se consigue colocando un repetidor en cabecera que retransmite transparentemente el contenido del canal ascendente

en el descendente simulando de esta forma el comportamiento de un bus (Fig.4). Esto supone un desperdicio del ancho de banda del canal descendente, que frecuentemente transportará datos colisionados o destinados a terminales de otras redes [5].

Por otra parte, las largas distancias de las redes de cable hacen que la resolución de las colisiones sea menos eficiente:

- aumenta el tiempo de riesgo de colisión: cuando una estación comienza a transmitir, su señal tiene que llegar hasta cabecera y luego ser retransmitida por el canal descendente para que las restantes estaciones perciban el medio como ocupado
- la duración mínima de cada transmisión aumenta: cada vez que una estación transmite debe hacerlo durante el tiempo suficiente para saber que su paquete no colisionó. Esto supone un tiempo igual al doble del retardo máximo de propagación de la red. Todas las transmisiones menores que este tiempo deberán ser 'rellenadas' para alcanzar esta duración mínima, desperdiciándose así un valioso ancho de banda.
- puede darse trato discriminatorio entre estaciones: una vez que una estación detecta que su transmisión colisionó, pone en marcha un mecanismo para volver a intentarlo. Las estaciones más alejadas de la cabecera siempre tardarán más en detectar las colisiones y por tanto reaccionarán más tarde ante una colisión. Deberá ponerse en marcha algún mecanismo para que todas las estaciones tengan las mismas oportunidades independientemente de su ubicación

Limitando la extensión de la red o la tasa binaria, podemos amortiguar estos efectos, pero no erradicarlos, y además estaremos empobreciendo el servicio prestado.

Por otra parte, los sistemas basados en contienda presentan un comportamiento impredecible ante elevadas cargas de tráfico y no permiten garantizar al usuario una determinada calidad de servicio, lo que prácticamente eliminará la posibilidad de prestar otro tipo de servicios que no sea de datos.

6.-Acceso al Medio Mediante Solicitud-Reserva

Aunque los primeros módems utilizaban el método de contienda, las últimas generaciones tienden unánimemente hacia un método de acceso al medio radicalmente diferente: el modelo de solicitud-reserva, que se adapta muy bien a las características de las redes de CATV.

En estos sistemas el canal descendente es de difusión y el órgano de cabecera es el único transmisor, por tanto no presenta problemas de acceso y los datos que transmita serán recibidos por todos los usuarios sintonizados a ese canal.

La dificultad radica en el canal ascendente, por el que varios terminales intentarán transmitir sus datos hacia cabecera. El mecanismo de solicitud-reserva pretende establecer un procedimiento para regular eficientemente el acceso múltiple a ese canal.

Para ello se estructura el canal ascendente en *slots* de tiempo. El órgano de cabecera controlará la utilización que se haga de cada uno de esos slots.

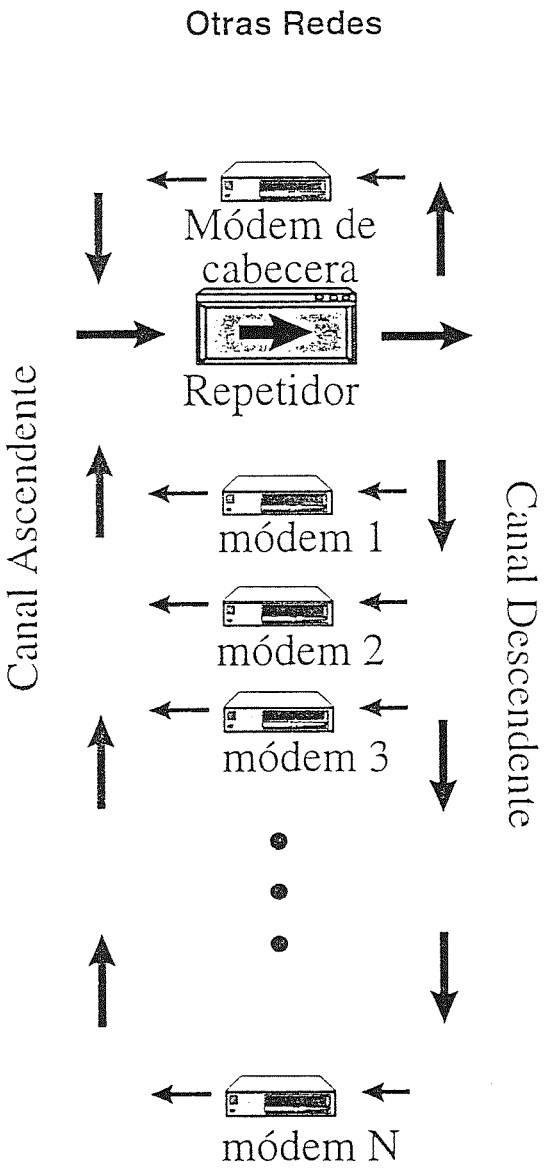


Fig. 4: Sistema simétrico basado en contienda. En cabecera se retransmite el canal ascendente sobre el descendente.

El órgano de control de cabecera puede asignar un conjunto de slots a una estación para que trasmita sin riesgo de colisión (*slots reservados*), o bien permitir que sean accedidos por todas o un conjunto de estaciones (*slots de contienda*) [4].

Las estaciones están en permanente comunicación con el órgano de cabecera que se encarga de la asignación y gestión del ancho de banda del canal ascendente. Cuando una estación tenga datos para transmitir, solicitará el número de slots necesarios a la cabecera, que se los asignará mediante un mensaje enviado por el canal descendente.

El órgano de control deberá confirmar todos los datos que reciba por los slots de contienda. Si una estación no recibe confirmación de los datos que envió en un slot de este tipo, sabrá que colisionaron e intentará retransmitirlos. Así no es necesario que sean transmitidos por el canal descendente. Además el órgano de cabecera analiza quien es el destinatario de cada paquete recibido y lo encamina hacia el canal descendente correspondiente o hacia otras redes externas a la de cable.

Este esquema logra un mejor aprovechamiento la capacidad tanto del canal ascendente como del descendente. Además al ser un órgano central el encargado de asignar el ancho de banda, se puede garantizar una determinada calidad de servicio (QOS, *Quality of* Service) a los usuarios, y definir distintas clases de servicio (cada una basada en un ancho de banda y una prioridad, por ejemplo) para ofertar a los usuarios [7].

Se han descrito las características más generales de este método, pero a la hora de llevarlo a la práctica se puede implementar de formas muy diversas. Por ejemplo, se puede imponer que los paquetes sean transmitidos íntegramente en slots consecutivos o por el contrario fragmentarlos en pequeños trozos y transmitirlos intercaladamente con los de otras estaciones [2]. Las solicitudes se pueden transmitir en slots de contienda, minislots especiales para solicitudes, o adjuntos a paquetes de datos. Una estación puede estar limitada a tener una sola solicitud pendiente simultáneamente o por el contrario poder ir solicitando ancho de banda para todos los datos que tenga en cola. Las asignaciones en el canal descendente pueden ir acompañadas de una marca temporal que indique el comienzo del período asignado, o simplemente la propia emisión de la misma indicar el comienzo del período de transmisión [2] [3]. Por otra parte, los algoritmos que se apliquen en cabecera para asignar el ancho de banda pueden ser también muy diversos.

7.-Proceso de Estandarización

Como hemos visto, los módems no son entidades independientes que se comunican entre sí como hacen los módems telefónicos. Para poder funcionar correctamente, un módem ha de estar en concordancia con el módem destino, pero también órgano de cabecera de la red.

Los módems de cable hasta ahora disponibles obedecen a tecnologías propietarias desarrolladas por las empresas fabricantes. Esto significa que los módems de un determinado fabricante sólo se podrán comunicar con equipos de cabecera de ese mismo fabricante. Esta situación limita enormemente la implantación de estos dispositivos, por lo que se están llevando a cabo importantes esfuerzos destinados a elaborar un modelo de referencia estándar, que permita comunicarse a módems de distintos fabricantes, a través de una misma red.

Este modelo de referencia será un sistema asimétrico basado en solicitud-reserva. Sin embargo, los detalles acerca de la implementación del mismo aún no se han esclarecido y actualmente son varias las soluciones propuestas.

El IEEE802.14 fue el grupo designado para elaborar este estándar. Propone un sistema basado en el transporte de celdas ATM y que permitirá la integración de servicios de datos junto con otros en tiempo real. El problema es que este grupo, formado principalmente por fabricantes ha venido actuando demasiado lento y tras más de dos años de trabajo, aún no ha hecho pública ninguna relación de especificaciones. Aunque ha anunciado que lo hará antes del fin de 1997, mientras tanto se han puesto en marcha otras iniciativas.

El grupo MCNS (*Multimedia Cable Network Systems*) promovido por CableLabs está formado por un grupo de operadores norteamericanos (80% del mercado de EE.UU.) que apremiados por la competencia de otros sectores (telefónicas, satélite) se han movido ágilmente para elaborar cuanto antes un conjunto de especificaciones, que se hicieron públicas el pasado mes de marzo. Actualmente están siendo revisadas por los fabricantes, que se espera comiencen a ofrecer módems MCNS a finales del 97. El modelo se basa en el transporte de tráfico IP de forma transparente a través de la red de CATV [4].

Las especificaciones del IEEE802.14 y MCNS, aunque basadas en modelos asimétricos de solicitud-reserva, siguen caminos diferentes y en principio no serán compatibles.

Ambas organizaciones emitirán sus propuestas al ITU (MCNS de la mano de algún

organismo con capacidad de estandarización como podría ser el SCTE, *Society of Cable Telecommuniction Engineers*) que decidirá finalmente si alguna de las dos es reconocida como estándar oficial.

MCNS por el momento parece estar ampliamente aceptado en EE.UU., donde podría convertirse en un estándar 'de facto', sin embargo a nivel internacional parece que la expectación en torno al IEEE802.14 es importante.

8.-Situación Actual del Mercado

A pesar de esta situación inestable producida por la carencia de un estándar y la incertidumbre acerca de la aparición del mismo, se contabilizan más de 22.000 módems de cable instalados en todo el mundo, a la fecha de redacción de este escrito.

Todos ellos obedecen a tecnologías propietarias. LANcity-Bay Networks y Zenith fueron las pioneras y se encuentran a la cabeza con módems instalados en más de 600 redes entre los dos. Ambos son sistemas simétricos, Zenith basado en contienda y LANcity, que por el momento es quien lleva vendido el mayor número de módems, ofrece un protocolo basado en solicitud-reserva en su última generación de producto. Anuncia que tendrá un módem MCNS listo antes del fin del 97.

Motorola, cuyo módem se puede considerar ya de segunda generación, tiene características comparables con el modelo de MCNS. El número de módems instalados está a punto de igualar al de Zenith.

Otras empresas pioneras en esta tecnología son Hybrid Networks, Com21, Terayon Corp., General Instruments entre otras.

El panorama del mercado es cambiante, siguiendo, entre otras pautas, la corriente del proceso de estandarización., Algunas grandes compañías (Intel, HP) han abandonado el mercado de los módems de cable, pero es mayor el número de las que se han sumado (USRobotics, Hayes, Cisco, Panasonic, NEC, etc.). Además han surgido recientemente gran cantidad de pequeñas empresas con el objetivo de fabricar y comercializar estos dispositivos en entornos locales. Hoy en día resulta difícil contabilizar el número total de fabricantes, pero se estima que sobrepasan los 30.

Es importante notar que de las cerca de 1000 redes donde han sido instalados módems de cable a modo de prueba, más de medio centenar han pasado a ofrecer el servicio comercialmente. Aunque los primeros sistemas surgieron en EE.UU., de donde proceden la mayor parte de fabricantes, este producto ha tenido más repercusión de la que se esperaba en otros mercados. Actualmente hay varios sistemas en marcha en Europa, Australia y algunos países asiáticos.

Pero el despegue total de estos sistemas hacia grandes volúmenes de producción no se producirá hasta que no aparezca un estándar medianamente consolidado.

9.-Acceso Vía Radio

Pero las infraestructuras de cable no llegarán, en general, hasta todos los habitantes. Es decir, un servicio considerado universal, deberá utilizar forzosamente otras tecnologías alternativas al cable para llegar a ciertos usuarios de no fácil acceso.

Como cabe esperar, las tecnologías más adecuadas a estas condiciones son las tecnologías vía radio.

Puesto que en España no está permitido el funcionamiento de sistemas a 2.5 GHz, pues parte de la banda ya está ocupada en otros servicios, y los de 42 GHz no presentan ventajas, ni de costes ni técnicas, frente a la banda de 28 GHz, nos quedaremos con estos últimos en la discusión de este artículo.

Estudiaremos, en concreto, las ventajas de implantar un sistema digital frente a uno analógico, no solo por sus mayores prestaciones en cuanto a implantación de servicios sino también por su menor coste.

10.-Arquitectura de Red

La arquitectura de un sistema de microondas es similar a la de telefonía móvil celular. Una cabecera es el centro de procesado, tratamiento y control de la señal. El transmisor de cabecera envía la señal por enlaces punto a punto a otros transmisores, que la radiarán (omnidireccionalmente) hasta los límites de su célula de cobertura. Para evitar interferencias entre células adyacentes se utilizan métodos diversos, como polarización cruzada e interleaving en frecuencia.

Los transmisores se sitúan en lugares altos, y las antenas de usuario preferiblemente en las azoteas, para que exista línea directa de visión. Pero en las ciudades no siempre se consigue. Se recurre entonces a repetidores pasivos (espejos) o activos de muy baja potencia y bajo coste, para cubrir las zonas de sombra. En cuanto a servicios, los sistemas de radio a 28 GHz o LMDS disponen de 1 GHz de ancho de banda, que proporciona la suficiente capacidad para ofrecer, al igual que el cable: televisión multicanal, transmisión de datos a alta velocidad y comunicaciones interactivas en general.

11.-Zonas rurales

En la mayor parte de España, si salimos de las grandes localidades nos encontramos con un paisaje rural caracterizado, en general, por una gran extensión con poblaciones pequeñas y dispersas. Es decir, se trata de zonas con baja densidad de población. En estas circunstancias no resulta rentable tender kilómetros de cable para llegar a unos pocos abonados. Un sistema LMDS parece, en principio capaz de solucionar el problema. La disyuntiva está en si usar un sistema analógico (aparentemente más barato) o uno digital (dotado de mayores prestaciones).

12.-Sistemas Analógicos versus Sistemas Digitales

La duda está de actualidad: ¿qué es mejor, un sistema analógico cuya principal ventaja es el menor coste del set-top box de abonado, o uno digital, con mayor capacidad de canales gracias a la compresión MPEG-2 y con facilidad para transmisión de datos a alta velocidad (Internet)? Para poder dar una respuesta es necesario además comparar los costes que resultan de dar cobertura a una zona con ambos sistemas.

Utilizamos un método de transmisores esclavos: se tiene una célula central con un transmisor omnidireccional, y bordeando por fuera la célula central se sitúan otros transmisores esclavos, que reciben la señal del central gracias a una antena receptora de mayor ganancia. Estos transmisores son sectoriales (su área de cobertura es aproximadamente elíptica), y el objetivo de colocarlos de esta manera es que puedan recibir la señal directamente del transmisor principal evitando así la necesidad de radioenlaces punto a punto.

Para nuestra discusión nos basamos en un modelo de población regular, en el que las poblaciones distan entre sí 2.4 Km y cuentan con una densidad de población de 27 habitantes por Km^2,datos típicos de Castilla y León.

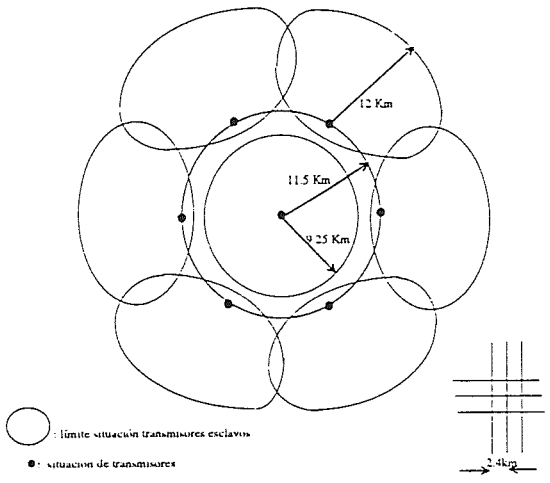


Fig. 5: Cobertura con señal analógica FM.

En el caso analógico, para una modulación FM de 18 MHz de ancho de banda por portadora y canal, se tiene una capacidad total de 50 canales, en un ancho de banda disponible de 1 GHz. Con una potencia de 24 dBm por canal y en zona seca, el radio de alcance de la señal del transmisor omnidireccional puede llegar a 9.25 Km, llegando hasta los 12 Km para los transmisores esclavos (ver Figura 5). Estos se pueden situar a una distancia máxima del transmisor central de 11.5 Km. Si aproximamos el área que queda cubierta por un círculo de radio 11.5 + 12 Km, se tiene que:

- *Area de cobertura total = 1734 Km^2*
- *Total pueblos cubiertos = 301 pueblos*

Suponiendo un coste orientativo de 25 millones de ptas. por transmisor y 10 millones para el radioenlace de 1GHz(que comunica el transmisor central con otro transmisor central o con la cabecera), el coste asciende a:

- *coste: 7×25 + 10 = 185 millones de ptas.*

Considerando una media de 3 habitantes por hogar pasado, y suponiendo un 60% de penetración del servicio, calculamos el coste por hogar pasado:

$$\frac{coste/km^2}{mhab/km^2} \bullet \frac{nhab/hogar}{\%h.p.} = 20.000\,ptas/h.p.$$

donde h.p. = hogar pasado

Para la señal digital se tiene una modulación QPSK, con 24 MHz de ancho de banda por portadora. La capacidad del sistema es de casi 200 canales digitales, al incluir 5 canales comprimidos MPEG-2 en cada portadora. El radio de alcance del transmisor omnidireccional es de 11 Km, con una potencia por portadora de 27 dBm, el de los transmisores esclavos de 14.5 Km, pudiéndose situar a 13.75 Km del centro de la célula (Figura 6), con lo que, calculando de forma análoga al caso anterior:

- *Area cobertura total = 2507 Km^2*
- *Total pueblos cubiertos = 435 pueblos*
- *Coste = 8×25 + 10 = 210 millones de ptas.*

Entonces:

$$\frac{coste/km^2}{mhab/km^2} \bullet \frac{nhab/hogar}{\%h.p.} = 15500\,ptas/h.p.$$

donde h.p. = hogar pasados

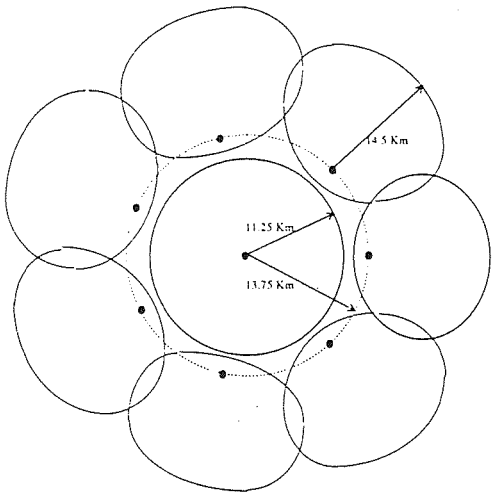


Fig. 6: Cobertura con señal digital QPSK.

13.-Conclusiones

Los módems de cable actualmente están en situación de ofrecer un servicio no disponible a través de ninguna de las infraestructuras existentes.

El acceso a los hogares residenciales es un tema complejo y caro que hasta ahora ha estado limitado a la línea telefónica. Pero el par trenzado que ésta proporciona resulta insuficiente para los nuevos servicios que plantea la Sociedad de la Información [6].

La llegada de las redes de CATV abre un panorama nuevo por explorar. Los sistemas basados en módems de cable ofrecen una alternativa flexible y económica, con capacidad de evolucionar paulatinamente al ritmo que lo hagan los nuevos servicios.

Por otra parte, no todas las zonas son susceptibles de disponer, a un coste razonable, de este tipo de servicios a través de cable.

El acceso radio es la solución idónea para solventar esta problemática. se han estudiado las dos posibilidades de acceso radio: analógica y digital, y en base al estudio realizado se concluye que un acceso radio digital es mejor y más barato que su homólogo analógico por diversas causas.

El radio de alcance de la señal digital es mayor que el de la señal analógica, en igualdad de condiciones. Por tanto, al cubrir una zona mayor, el coste por hogar pasado es un 22%menor. Además, un usuario que reciba televisión analógica, y que por lo tanto tiene un set-top box analógico, si quiere tener acceso a Internet o a otro servicio de datos a alta velocidad necesitará adicionalmente un modem, y esto implica un nuevo coste para el usuario, con lo que la ventaja de menor coste del set-top box analógico desaparece. Como consecuencia de todo lo visto, y en previsión de que los precios de los equipos digitales tanto de usuario como de la cabecera disminuyan, es más recomendable un sistema LMDS digital que uno analógico.

Referencias

[1] Judith Redoli, Rafael Mompó, "Estrategia de Arquitectura de Red para Operadores de CATV", *Cable y Satélite Profesional*, N°21, Enero-Febrero 1996

[2] John O. Limb, Dolors Sala, Georgia Institute of Technology, "A Protocolo For Efficient Transfer of Data Over Hybrid Fiber/Coax Systems", *Proceedings of INFOCOM'96, pp. 904-911*, San Francisco, CA, Marzo 24-28 1996.

[3] John O. Limb, Dolors Sala, "An Access Protocol to Support Multimedia Traffic over Hybrid Fiber/Coax Systems", *Second International Workshop in Community Networking*, pp. 35-40, Junio 20-22, Princeton 1995

[4] Arthur D. Little, Inc., MCNS Holdings, "Radio Frequency Interface Specifications", SP-RFII01-970326, Data-Over-Cable Interface Specifications

[5] Sharon Eisner Gillett, "Connecting Homes To The Internet: An Engineering Cost Model Of Cable Vs. ISDN", *Master Thesis*, Laboratory for Computer Science, Massachusetts Institute of Technology, Junio 1995

[6] David Gingold, "Integrated Digital Services For Cable Networks", *Master of Science in Technology and Policy* at The Massachrsetts Institute of Technology, Agosto 1996

[7] Eva Parrilla, Judith Redoli, Rafael Mompó, "Acceso Veloz a Internet por las Redes de TV por Cable", *Serie Breve de Ciencia y Tecnología*, Sección de Publicaciones de la Universidad de Valladolid, Enero 1997.

Protocolo de acceso múltiple con control de asignación y calidad de servicio para redes sin hilos en Modo de Transferencia Asíncrono

Lluís Casals [1] y Josep Paradells [2]

Departament de Matemàtica Aplicada i Telemàtica
Universitat Politècnica de Catalunya

(1) EUPVG, Avda. Eduard Maristany s/n, 08800 Vilanova i la Geltrú, España
Tf: (93) 896 77 16, Fax: (93) 896 77 00, E-mail: casals@mat.upc.es
(2) Campus Nord UPC, Mòdul C3, C/ Jordi Girona, 1 - 3, 08034 Barcelona, España
Tf: (93) 401 60 24, Fax: (93) 401 59 81, E-mail: teljpa@mat.upc.es

Abstract:

In this work we propose and study a multi-access protocol with channel assignment control and several quality of service control mechanisms to be used in ATM wireless networks. The proposed protocol is slotted-basis with no fixed slot length and no frame structure. Protocol simulations have been made to study the protocol behaviour under several traffic mixtures, such as ABR and CBR traffic types. We present results relatives to the delay-throughput behaviour, the delay variation and the minimum guaranteed bandwidth.

1. Introducción

En este trabajo proponemos y estudiamos un protocolo de acceso múltiple con control de asignación de canal y varios mecanismos de control de calidad de servicio para su uso en redes sin hilos con transmisión en modo de transferencia asíncrono (ATM). Este estudio incluye simulaciones del comportamiento del protocolo para determinar su rendimiento y ajustar los parámetros clave. Se ha estudiado el protocolo en escenarios correspondientes a redes de alta velocidad con mezcla de tráficos que tienen diferentes exigencias de calidad de servicio. En este contexto, se han estudiado métodos para ofrecer el tráfico con una calidad de servicio garantizada. Los parámetros que se han examinado para validar el grado de consecución de estos objetivos son, por ejemplo, el comportamiento retardo-caudal, la variación del retardo, el ancho de banda medio garantizado o la probabilidad de descarte de celdas.

En el diseño del protocolo, los objetivos principales han sido obtener un retardo bajo de servicio de celdas y un mecanismo eficiente de uso compartido del medio que permita ofrecer diferentes calidades de servicio frente a un número variable de terminales y con una carga variable. Estos objetivos se alcanzan usando un método eficiente de acceso múltiple y dotando a la estación base con métodos de asignación de ranuras a los diferentes terminales.

Existen ya otros protocolos que cumplen estos objetivos en mayor o menor grado, de los cuales podemos destacar los siguientes: En el protocolo PRMA (Packet Reservation Multiple Access) [1], las ranuras de tiempo están agrupadas para formar tramas, donde cada ranura pude estar "reservada" o "libre". Los terminales móviles pueden reservar una ranura de tiempo transmitiendo el primer paquete del mensaje en aquella ranura de tiempo. Si hay una colisión se pierde todo el tiempo correspondiente a esa ranura, lo cual conlleva una pérdida de capacidad importante. Otra desventaja es la estructura fija de las tramas; esto no permite conseguir un uso eficiente del medio bajo condiciones de carga asimétrica en el sistema. En el protocolo RMAV (Reservation-based Multiple Access with Variable frame length) [2], posterior al PRMA, se propone una trama con longitud variable para alcanzar un uso mas eficiente del medio, así como un mecanismo, libre de colisiones, para mantener la reserva de ranuras. En el protocolo DQRUMA (Distributed-Queueing Request Update Mutiple Access) [3], cada ranura está compuesta por un intervalo de contienda a través del cual los terminales móviles pueden pedir una ranura para poder transmitir. Los terminales móviles solamente necesitan mandar una petición a la estación base cuando en ese móvil no hay otros paquetes esperando para ser transmitidos, ya que, los paquetes que llegan cuando el móvil tiene otros paquetes esperando para ser transmitidos pueden hacer las peticiones sin colisiones mediante "piggybacking" de la petición en un paquete a transmitir. En este protocolo, además, se prevé que la estación base tenga la responsabilidad de administrar la asignación de las ranuras de tiempo a partir de las peticiones recibidas.

Actualmente, el campo de las redes sin hilos está en constante crecimiento en muchos países. En particular, las redes ATM sin hilos tienen un especial interés debido a su capacidad para llevar tráfico multimedia a alta velocidad. Hay varios proyectos importantes trabajando para conseguir realizaciones prácticas de redes ATM sin hilos. En el marco del programa ACTS podemos encontrar el proyecto "The Magic WAND" [4] que, junto a otros trabajos, están influyendo en la futura estandarización de la red sin hilos de alta velocidad (HIPERLAN tipos 2,3 y 4) que está llevando a cabo en Europa el grupo RES 10 de la ETSI.

Al protocolo MAC propuesto en el proyecto "The Magic WAND" se le ha denominado protocolo MASCARA [5], y está basado en una

estructura de trama de longitud variable, con método de acceso basado en TDMA y que combina mecanismos de reserva y de contienda.

2. Descripción del protocolo

El protocolo propuesto en este trabajo está basado en una estructura de ranuras temporales pero sin formar una trama. Así, con esta estructura basada en ranuras, por el canal de comunicación se transmite una secuencia continua de ranuras. Se ha considerado que la estación base y los terminales móviles comparten un único canal mediante un esquema "Time Division Duplex" (TDD). Cada ranura está compuesta por tres intervalos de tiempo que permiten la transmisión de información de control, realizar el mecanismo de acceso y la transmisión de datos.

El primero de ellos es un intervalo de control (C, en la fig. 1). En este intervalo la estación base transmite a los móviles información de control como es la confirmación de las peticiones de ranura y de las transmisiones de celdas realizadas en la ranura anterior y, también, la asignación de la ranura actual a una fuente (móvil o estación base). El intervalo de control también podría ser usado para permitir a los móviles sincronizarse con cada ranura temporal.

El segundo intervalo, llamado intervalo de petición (R, en la fig. 1), se utiliza para que los móviles hagan peticiones de asignación de ranura en la que poder transmitir. En el intervalo de petición, los diferentes móviles que quieren obtener una ranura compiten de manera similar a lo que sucede en un protocolo aloha ranurado: con acceso aleatorio. Por cada paquete que un móvil quiere transmitir, éste solo necesita realizar una petición, de manera que cada celda que forma un paquete es transmitida en una ranura asignada por la estación base. Esta asignación de ranuras que lleva a cabo la estación base sigue una cierta política de servicio que tiene en cuenta los requerimientos de las peticiones recibidas y la calidad de servicio comprometida en la conexión.

El tercer intervalo de cada ranura es el intervalo de transmisión de celda (T, en la fig. 1). Como se ha mencionado anteriormente, en este intervalo solamente puede transmitir un móvil o la estación base. La estación base tiene la responsabilidad de asignar cada intervalo T a una de las peticiones que están esperando a ser servidas. Mas adelante se discuten algunos posibles algoritmos de asignación.

El intervalo de transmisión de celda puede tener una duración variable en función de las celdas consecutivas que la estación base deje transmitir a cada móvil. La duración máxima de este intervalo está fijada por un parámetro del sistema y depende de la duración del grupo de celdas y de la eficiencia de canal fijada. Por otra parte, el intervalo de transmisión tendrá una duración nula cuando no haya peticiones que servir (ni la base ni los terminales móviles quieren transmitir), y, en consecuencia, la ranura temporal queda reducida a los intervalos C y R. Esta característica aporta dos importantes ventajas: por un lado, permite un uso eficiente del canal y, por otro lado, el tiempo entre apariciones del intervalo de petición será más pequeño (el mínimo) dando un retardo de acceso pequeño cuando el sistema funciona con cargas bajas.

Tal como se referenció anteriormente, los móviles y la estación base solo transmiten cuando la estación base les asigna el intervalo T de una ranura. Por tanto, todos los paquetes generados están esperando en sus respectivas fuentes hasta que obtienen el permiso para ser transmitidos. Así pues, este protocolo tiene una característica de cola distribuida (distributed-queuing). Por otro lado, la estación base conoce las necesidades del sistema en todo momento a través de las peticiones mandadas por los móviles y la estación base, que habrán sido negociadas y, posteriormente, aceptadas por la estación base. Este hecho permite a la estación base planificar la asignación del intervalo T de cada ranura, para compartir la capacidad del canal, dando a cada móvil (enlaces ascendentes) y a la estación base (enlace descendente) una parte de ese canal, asignándoles un cierto número de ranuras, en función de la carga del sistema y de las características del tráfico procedente de las peticiones tanto de los terminales como de la estación base, mediante una política de servicio adecuada. El control centralizado que realiza la estación base también le permitirá llevar a cabo controles policía sobre las conexiones admitidas.

En el presente trabajo se han estudiado varias alternativas de política de servicio con el propósito de conseguir estos objetivos. En particular se ha estudiado el comportamiento de mecanismos de tipo "Round-robin" para la gestión igualitaria de tráficos con características estadísticas similares, y mecanismos con prioridad para gestión de la calidad de servicio comprometida por la estación base frente a mezclas de tráfico del tipo ABR y CBR.

3. Funcionamiento del protocolo

En esta sección se describirá el funcionamiento del protocolo de acceso que se propone describiendo las principales situaciones que se pueden presentar.

3.1. Petición de ranura y asignación de ranura

Cuando un móvil tiene un paquete para transmitir, el móvil tiene que hacer una petición de ranura a la estación base. El móvil esperará el comienzo de un intervalo R y entonces transmitirá una petición de ranura (Rm1 en la fig. 2). En el caso

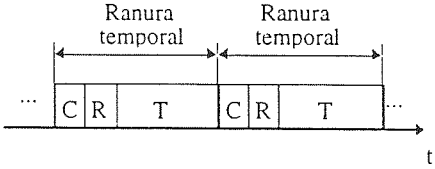


Fig. 1. Estructura de una ranura temporal.

de que sea la única petición que se transmite en aquel intervalo R, el móvil recibirá un reconocimiento (o confirmación) (RAm1 en la fig. 2) en el intervalo C de la ranura siguiente. El móvil transmitirá (Tm1 en la fig.2) cuando reciba una asignación de ranura (SAm1) en el intervalo C de la misma ranura en la que hizo la petición (si no hay otras peticiones pendientes) o en una de las siguientes ranuras temporales. En el intervalo de petición, el móvil transmite información de control a la estación base, como es la longitud del paquete. Esta información se almacena en la tabla de peticiones que es mantenida por la estación base y puede ser usada como un parámetro más para resolver el algoritmo de asignación de ranuras.

En lo relativo al uso del canal de comunicación, la estación base se considerada como otro móvil. Pero, en este caso, cuando la estación base quiere transmitir un paquete hace la petición directamente al inicio de la ranura. Debido a que la estación base no tiene que compartir el intervalo R para realizar sus peticiones, éstas son un caso especial de petición, para las que no se darán colisiones. Otra particularidad del funcionamiento de la estación base como fuente de paquetes es que puede obtener el permiso de transmisión en la misma ranura en que realizó la petición, siempre que no haya otras peticiones pendientes que tengan mayor prioridad.

3.2 Colisión

Se producirá una colisión cuando uno o más móviles realicen una petición en el mismo intervalo R (fig. 3). En este caso la estación base no "entenderá" ninguna de las peticiones transmitidas y, por tanto, no transmitirá ningún reconocimiento de petición en la próxima ranura. Los móviles interpretarán el hecho de no recibir un reconocimiento como que se produjo una colisión. Después de esto, los móviles llevan a cabo un proceso de "backoff" y más tarde volverán a intentar la petición de nuevo.

3.3. No hay peticiones esperando para ser servidas

Como se mencionó anteriormente, la estructura de una ranura temporal no tiene una longitud fija. Esta situación se puede dar, en particular, cuando no hay peticiones pendientes en la tabla de peticiones de la estación base, ja que, entonces, el intervalo T no se usa (fig. 4). Por tanto, la ranura temporal solo consiste en un intervalo C y un intervalo R. En este caso un móvil puede acceder rápidamente con un tiempo de acceso bajo.

3.4. Mecanismo de ventana y mecanismo de prioridades para servicios ABR y CBR

Para el estudio del comportamiento del protocolo propuesto para el caso de servicios ABR se utilizan los parámetros de QoS siguientes: MCR (Minimum Cell Rate) y MBS (Maximum Burst Size). El MBS es usado por la estación base para establecer un mecanismo de control de policía para asegurar que la fuente correspondiente no sobrepasa el valor MBS negociado en la fase de conexión. Para ello, se utiliza un mecanismo de ventana temporal deslizante de una longitud equivalente a (MBS + 1) ranuras completas (que contienen los tres intervalos: C, R y T). En el momento de decidir la asignación de una nueva ranura, la estación base comprueba que no se supere el valor de MBS fijado. En el caso de ser superado no se asignará la ranura.

El MCR se usa para contabilizar el tanto por ciento de celdas transmitidas que no cumplen este parámetro de QoS. Para ello se utiliza la misma ventana temporal usada en el tratamiento del MBS, y se compara el MCR con el número de celdas transmitidas dentro de la ventana. Esta comparación se realiza para cada celda transmitida excepto para las MCR primeras celdas transmitidas de cada paquete.

Tal como se discute en la sección siguiente, se ha estudiado el comportamiento del protocolo para el caso de transmisión de mezclas de tráfico con servicio ABR y CBR. Para este caso, la política de servicio usada trata al tráfico con servicio CBR con prioridad sobre el tráfico con servicio ABR.

4. Rendimiento del protocolo

Con el objetivo de evaluar el rendimiento del protocolo se han realizado un conjunto de simulaciones para diferentes situaciones de tipos y mezclas de tráfico. Inicialmente, se presentan los parámetros usados en el protocolo y los correspondientes valores usados en las simulaciones. En la tabla 1 se muestran los parámetros del canal y de la ranura que se usaron en cada simulación. Para la elección de la velocidad de transmisión se consideró que el presente protocolo se usará a través de un interfaz radio DECT. La longitud asignada al intervalo T se escogió considerando que se va a transmitir una secuencia de sincronización, de 32 bits, seguida de una celda ATM (53 bytes) y un campo de control de errores, de 32 bits.

En las secciones siguientes se presentan algunos resultados de simulaciones agrupados en tres modos de operación: modo paquete, modo servicio ABR y modo servicio CBR.

Las condiciones generales de simulación se han establecido considerando un canal de comunicación libre de errores y sin efecto captura.

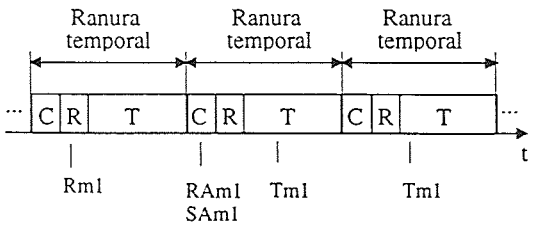


Fig. 2. Petición y asignación de ranura.

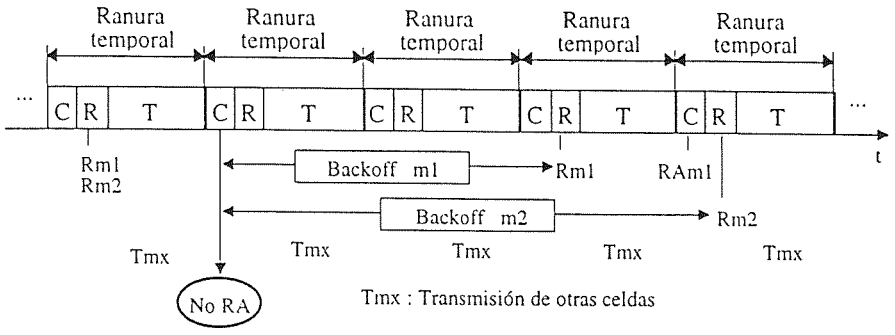


Fig. 3. Colisión entre peticiones procedentes de dos móviles.

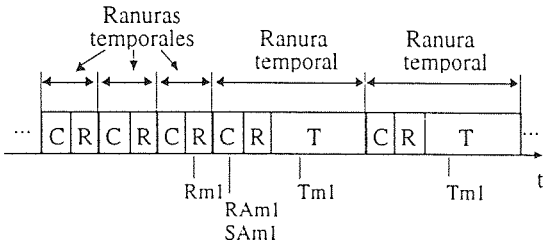


Fig. 4. Reducción de la longitud de las ranuras cuando no hay peticiones pendientes de servicio.

4.1. Modo paquete

En el modo paquete se ha considerado que todas las fuentes generan paquetes con una distribución geométrica, de media P paquetes/s, y con longitud de paquete constante. En este modo se han estudiado las cuatro políticas de servicio siguientes: FIFO (First In First Out), que sirve primero a la petición que lleva primero y no sirve a la siguiente hasta que no ha completado el servicio de la anterior, "Round-robin", que asigna una ranura a cada petición, repartiendo por igual la capacidad del canal entre todas las peticiones pendientes, "Short-first", que sirve primero al paquete más corto, y "Short-first with Discrimination Control" (SFDC), que sirve primero al paquete más corto pero asegurando que los paquetes grandes también obtendrán servicio.

En la figura 5 se muestra el comportamiento del retardo de servicio frente al caudal (throughput) para 2 y 10 móviles, más la base. En este caso se usa una política de servicio FIFO. Por otro lado, la estación base y los móviles generan paquetes de longitud constante que requieren un tiempo de transmisión igual a T unidades de tiempo (donde T es la duración del intervalo de transmisión, T, de una ranura temporal). La carga total del sistema se genera a partes iguales por todos los móviles y la estación base. El mecanismo de "backoff" usado asigna un tiempo de espera a cada móvil, equivalente a un número aleatorio de ranuras temporales, con una distribución uniforme. Este mecanismo de "backoff" tiene un funcionamiento estático, aunque también podría usarse otra versión con comportamiento dinámico para que se adaptara a condiciones de carga altas en las que se producen muchas colisiones.

Los resultados se comparan con un sistema ideal, que consiste en un sistema que conoce las peticiones inmediatamente y que no tiene colisiones en el canal de transmisión y que utiliza una política de servicio de tipo FIFO. Este sistema ideal puede ser modelado como una cola M/D/1.

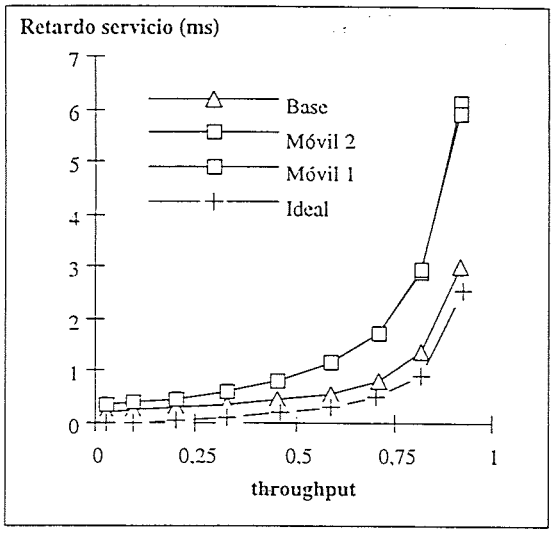
En la figura 5 se puede ver que los móviles tienen un retardo más alto que la estación base. Esto es debido a que la estación base realiza las peticiones sin colisiones y al hecho de que la estación base pueda ser servida en la misma ranura en la que ha realizado la petición. También se puede ver que, con un número de móviles alto el retardo de servicio es mayor que en el caso de 2 móviles. Este resultado es una consecuencia directa de la limitación del procedimiento de acceso que solo afecta a los móviles y deja la capacidad sobrante a la estación base.

En la figura 6 se comparan las cuatro políticas de servicio mencionadas anteriormente en un sistema con 3 móviles y una estación base. Se han tomado los mismos parámetros de carga para las cuatro políticas de servicio: la estación base genera 100 paquetes/s de 10 celdas, el móvil 1 genera 25 paquetes/s de 4 celdas y los móviles 2 y 3 generan 25 paquetes/s de 1 celda. Podemos ver que la estación base sufre ligeras variaciones en el retardo de servicio y en el tiempo de servicio en función de la política de servicio.

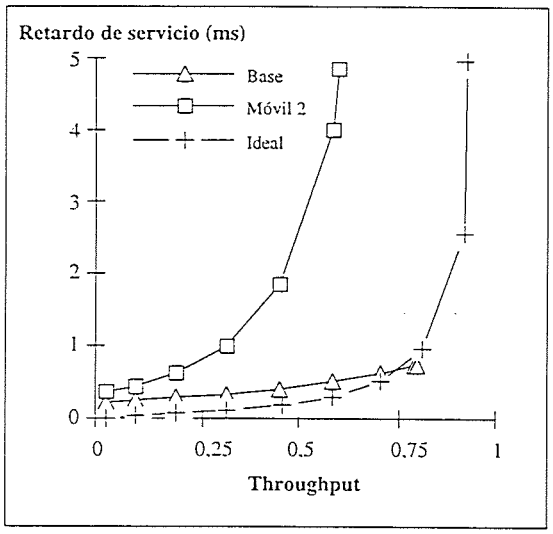
Por otra parte, con los móviles 2 y 3 se obtiene mejor QoS con las políticas de servicio "Round-robin", "Short-first" y SFDC que con FIFO.

Tabla 1. Parámetros del canal y de la ranura.

Parámetro	Valor	Unidades
Intervalo C	11	bytes
Intervalo R	9	bytes
Intervalo T	61	bytes
Gap entre intervalos	0,7	μs
Velocidad de transmisión	1152000	bps



(a) Base y 2 móviles.



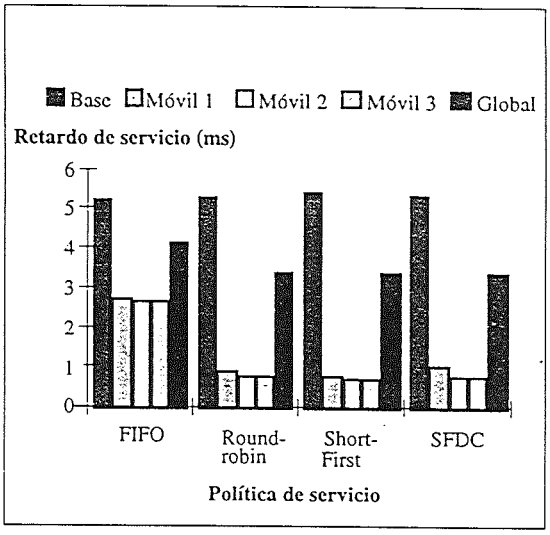
(b) Base y 10 móviles

Fig. 5. Retardo de servicio frente a throughput. (a) Base y 2 móviles. (b) Base y 10 móviles.

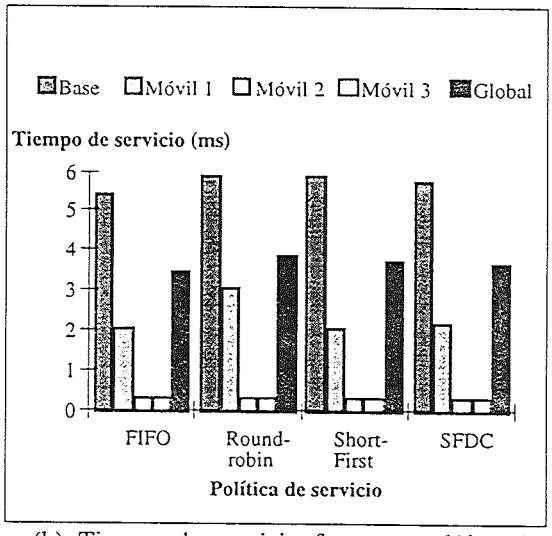
Para estos móviles la mejor política es "Short-first". Para el móvil 1, la mejor política desde el punto de vista del retardo es "Short-first" y la peor es la FIFO, mientras que para el tiempo de servicio la mejor es la "Round-robin". Para la política de servicio FIFO las fuentes que dan una carga baja tienen un retardo alto a causa de la estación base, que transmite paquetes de longitud mucho mayor.

4.2. Modo servicio ABR

En el modo servicio ABR se considera que cada fuente genera paquetes con una distribución geométrica, de media P paquetes/s, con una longitud de paquete constante y con una actividad aleatoria de la fuente modelada con un proceso de Markov tipo on-off.



(a) Retardo frente a política de servicio.



(b) Tiempo de servicio frente a política de servicio.

Fig. 6. Comparación del comportamiento de las políticas de servicio FIFO, Round-robin, Short-first, SFDC. (a) Comportamiento del retardo de servicio. (b) Comportamiento del tiempo de servicio.

Como se ha mencionado anteriormente, la estación base usa los parámetros MBR para limitar las ráfagas que la fuente intenta introducir en la red y usa el parámetro MCR para evaluar el grado de QoS en forma de tanto por ciento de celdas que son transmitidas sin cumplir este parámetro de calidad. La política de servicio usada asigna una ranura a la petición que tiene una velocidad de celdas actual menor. En la figura 7 se presenta el efecto sobre los paquetes generados por fuentes con servicio ABR cuando este tráfico es mezclado con el tráfico procedente de un número creciente de fuentes con servicio CBR. El móvil 1 es un móvil con servicio ABR que genera una carga constante de 10 paquetes/s de 10 celdas cada paquete. El valor del parámetro MBS es 9 celdas y el valor de parámetro MCR es 4 celdas por ventana.

De los resultados obtenidos se puede apreciar (fig. 7(i), trazo (a)) que el retardo de servicio de los paquetes generados por el móvil 1 se incrementa cuando la carga en el sistema aumenta. El aumento de la carga en el sistema es debido al incremento del número de fuentes con servicio CBR (concretamente, 2, 4, 6, 8 y 10 fuentes), todas ellas transmitiendo a una velocidad de 167 celdas/s. Otro efecto es el incremento del número de celdas con servicio ABR que son transmitidas sin cumplir el requerimiento especificado por el parámetro MCR que se había negociado en la fase de conexión (fig. 7(ii)).

En la figura 7(i), trazo (b), se puede ver un comportamiento similar con la mezcla de 2 móviles con servicio ABR y un número creciente de fuentes con servicio CBR. En este caso, los móviles con servicio ABR (1 y 2) generan, cada uno, 5 paquetes/s de 10 celdas y el número de fuentes con servicio CBR se ha tomado como 2, 4, 6 y 8, y con una velocidad de transmisión de 167 celdas/s, cada una.

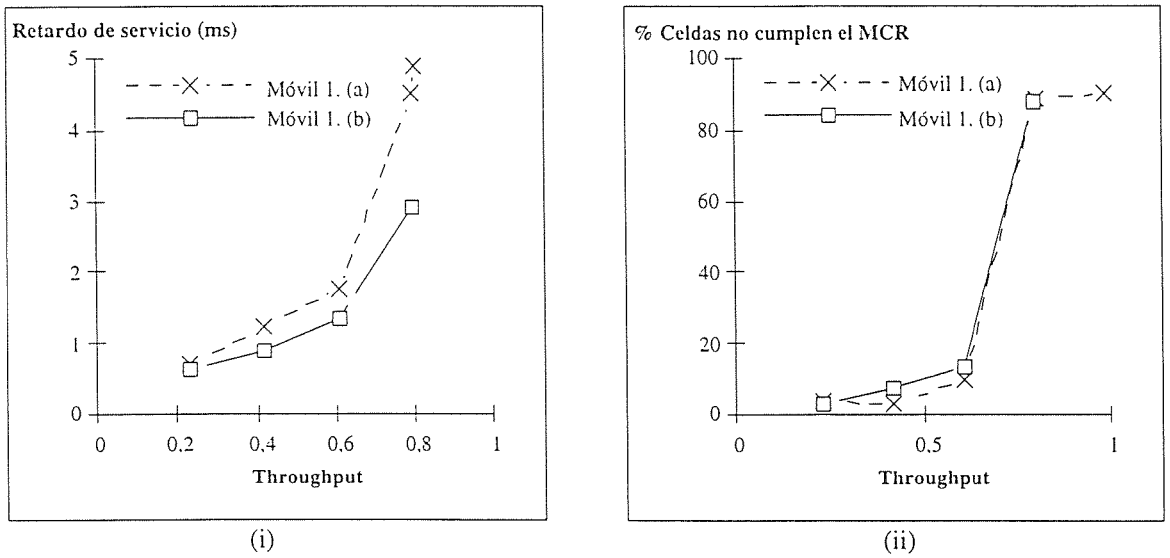
4.3. Modo servicio CBR

En el modo servicio CBR las fuentes generan un flujo constante de celdas. Los parámetros de QoS usados en este modo son el PCR (Peak Cell Rate) y el CDD (Cell Delay standard Deviation). El CDD es un parámetro que se ha usado para obtener una estimación de la fluctuación del retardo a través del cálculo de la desviación estándar del retardo de transmisión de las celdas. Las fuentes transmiten una celda solamente cuando la estación base les da permiso. A su vez, la estación base reparte los permisos de transmisión a las fuentes CBR de acuerdo con el PCR acordado en la fase de conexión. Si varias celdas tienen que ser transmitidas en la misma ranura, la estación base asignará la ranura a la fuente que tenga un retardo mayor (determinado como la diferencia entre la aparición de la ranura y el instante de transmisión teórico de la celda). En la evaluación del método propuesto se ha contabilizado el tanto por ciento de celdas que exceden un valor umbral de variación de retardo, dado como parámetro de la simulación.

En la figura 8, trazo (a), se muestra el valor del CDD para un número variable de fuentes con servicio CBR (1, 2, 4 , 5, 6, 8 y 10 fuentes), que tienen el mismo valor de PCR (167 celdas/s) y el mismo valor umbral de variación de retardo (0,3 ms).

En la figura 8, trazos (b), se muestra el efecto que produce una fuente con servicio ABR, que genera una carga creciente, sobre 2 fuentes con servicio CBR (estación base y móvil 1). La variación del retardo se incrementa tanto en la estación base como en el móvil. En este caso, la carga mínima corresponde a la carga generada por la estación base y el móvil (sin fuentes con servicio ABR), en donde se obtiene un mínimo para el valor del CDD debido a la baja carga y, por tanto, debido a que cada celda con servicio CBR se transmitirá, mayoritariamente, después de una ranura de tamaño mínimo (ranura sin intervalo T). También podemos ver que el valor del CDD está acotado a un valor máximo debido a que para cargas altas la mayoría de ranuras tienen un intervalo T y, entonces, la ranura tiene su duración máxima.

En la figura 8, trazos (c) también se muestra el valor de CDD para 2 fuentes con servicio CBR (estación base y móvil 1), cuando existen dos fuentes con servicio ABR que generan una carga creciente sobre el canal. En este caso, cuando los móviles con servicio ABR incrementan la carga ofrecida, estos móviles colisionan con mayor frecuencia y, por tanto, llegan menos peticiones a la estación base. En consecuencia, se transmiten menos ranuras que tengan el intervalo T presente y, entonces, el valor del CDD tiende a disminuir cuando la carga en el canal, debida al servicio ABR, disminuye.



(i) (ii)

Fig. 7. Mezcla de 1 (a) i 2 (b) fuentes con servicio ABR y un número variable de fuentes con servicio CBR. (i) Retardo de servicio frente al throughput. (ii) % de celdas que no cumplen el parámetro de QoS utilizado, MCR, frente al throughput.

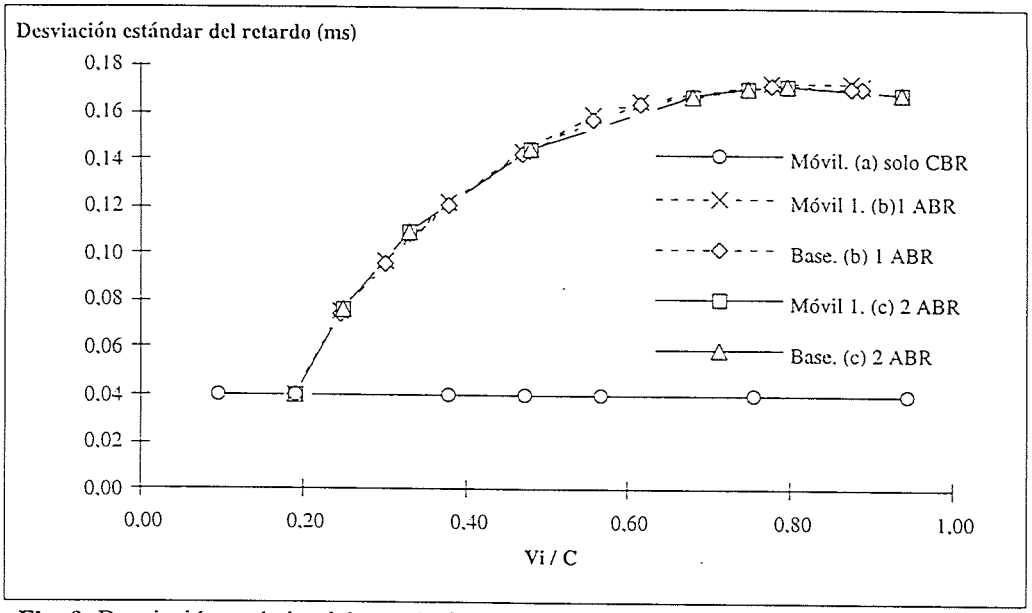


Fig. 9. Desviación estándar del retardo frente a cargas variables. (a) Carga variable debida a fuentes con servicio CBR. (b) Carga variable debida a 1 móvil con servicio ABR. (c) Carga variable debida a 2 móviles con servicio ABR.

5. Conclusiones

El protocolo propuesto tiene buenas características como método de acceso para redes sin hilos. Se han evaluado cuatro políticas de servicio que pueden ser usadas en la transmisión de paquetes y se ha visto que las políticas de servicio Short-first y SFDC son políticas buenas cuando la carga en el sistema es debida a paquetes con longitudes diferentes.

Para el modo de servicio CBR se ha evaluado la desviación estándar del retardo y se ha visto que está acotada a un valor que depende de la duración de la ranura temporal.

Finalmente, con mezclas de tráficos debidos a servicio ABR y CBR se ha visto que el retardo, para el tráfico CBR, está acotado a medida que las fuentes ABR incrementan la carga.

Teniendo en cuenta la definición y el rendimiento podemos concluir que el protocolo propuesto puede ser una buena alternativa como protocolo MAC para su uso en redes sin hilos en ATM o, des de un punto de vista más general, como protocolo de acceso a redes integradas sin hilos.

Referencias

[1] Googman, D.J., Valenzuela, R.A., Gayliard, K.T., y Ramamurthi, B., "Packet reservation multiple access for local wireless communications", *IEEE trans. Commu.* COM-37 (1989)885-890.

[2] Jeong, D.G., Choi, C.H., y Jeon, W.S., "Design and performance evaluation of a new medium access control protocol for local wireless data communications", *IEEE/ACM trans. Networking*, vol. 3, no. 6, pp. 742-752, Dec. 1995.

[3] Karol, M.J., Liu, Z., y Eng, K.Y., "An efficient demand-assignment multiple access protocol for wireless packet (ATM) networks", *Wireless Networks* 1 (1995)267-279.

[4] Mikkonen, J., y Kruys, J., "The Magic WAND: a wireless ATM access system", *ATCS Mobile Telecommunications Summit*, Granada, Spain, Nov. 1996.

[5] Bauchot, F., Decrauzat, S., Marmigère, G., Merakos, L., y Passas, N., "MASCARA, a MAC Protocol for Wireless ATM", *ATCS Mobile Telecommunications Summit*, Granada, Spain, Nov. 1996.

Contributions to the XDQRAP MAC Protocol over HFC Access Networks*

Cèsar Fernàndez†, Sebastià Sallent‡, Eva Vilaginés†
† Departament d'Informàtica i Enginyeria Industrial
Universitat de Lleida (UdL)
Ap. de correus 471, 25080 Lleida, Spain
e_mail: cesar@eup.udl.es
‡ Departament de Matemàtica Aplicada i Telemàtica
Universitat Politècnica de Catalunya (UPC)
C/. Gran Capità, s/n. Mòdul C3 08034 Barcelona
e_mail: matssr@mat.upc.es

Abstract

XDQRAP (eXtended Distributed Queuing Random Access Protocol) MAC protocol is one of many solutions proposed by IEEE 802.14 Working Group in order to define a common background for broadband/multimedia residential services emerging in a near future based on HFC (Hybrid Fiber Coax) access networks. In this paper we first present an improvement of such a protocol in order to deal with round trip delays larger than the slot time, preserving the excellent performance of throughput and delay of XDQRAP protocol and avoiding to increase the complexity of the access mechanisms. Finally, a methodology for CBR support is exposed, analyzing the cell delay variation time for different slot allocation policies.

1 Introduction

The ever increasing demand of new applications such as fast internet access, video on demand and others, in conjunction with the convergence of computer, communication, cable and television industries are the essential factors that boost the growing interest on extending the access of the broadband services to the user's home [6],[11].

Several alternatives have been heavily debated during the last years [8], from cooper loops as ADSL (Asymmetric Digital Subscriber Line) to FTTH (Fiber To The Home). One of this viable alternatives that might be taked into account is HFC (Hybrid Fiber Coax) due to several reasons, such as high bandwidth of fiber optics and a large number of existing networks of coaxial cable where an important amount of bandwith is not being used [9].

*Research partially supported by the project SMASH (TIC96-1038-C04-03) funded by the CICYT.

In this sense, the IEEE 802.14 working group has been working in aim to define the Physical and Medium Access Control layer protocols of a bi-directional Cable TV network based on HFC technology. Several MAC protocol proposals have been submitted to the 802.14 Working Group in order to accomplish the functional requirements defined in [3], and other proposals have been published in technical press.

One of these proposals is the XDQRAP (eXtended Distributed Queuing Random Access Protocol) MAC protocol of Scientific-Atlanta Inc. [13] as an evolution of a LAN/MAN MAC protocol called DQRAP [14].

This paper is organized as follows: Section 2 reports the delay versus throughput performance of XDQRAP obtained by simulation where no mechanisms to deal with long delays are considered. The best number of slots choice according to the actual scenarios and the use of techniques which allow the operation under long delays is justified. Section 3 describes an alternative to the classical interleaving solutions applicable to XDQRAP and delay results are compared to those of Section 2. Section 4 proposes and analyzes some techniques to add CBR support and finally, Section 5 presents concluding remarks and future work.

2 On the issue of frame format

One of the main factors that determines the performance of a XDQRAP protocol is the size of the frame format and the contention/data slot ratio. In this section we present some results obtained by simulation of the performance of a XDQRAP system under certain conditions discused below. We also comment performance measurements compiled from the avalaible literature and we expose the reasons that guided us to propose an improvement of the original protocol.

N_{ms}	S_{max}
2	0.74
3	0.99

Table 1: Maximun throughput of XDQRAP with no delay nor overhead.

Results obtained from IEEE 802.14 Working Group documents [5] report performance of 3 different protocols. XDQRAP is one of them, under a common scenario. This common scenario has two important factors to be highlighted.

The first one is the number of minislots considered and their size. In [5] simulations have been performed with 2 minislots with a length of 8 bytes for a upstream channel rate of 3 Mbps (megabits/second), meanwhile in [4] the size of the minislots was proposed to be 16 bytes. In this sense, simulations performed under poisson traffic assumption for a XDQRAP protocol with a variable number of minislots (N_{ms}) [2] reported the maximum throughput (S_{max}) figures showed in table 1, not considering delay nor overhead due to minislot length. For values of $N_{ms} > 3$ it seems clear that the benefits introduced by a larger number of minislots will be lower than the loss of throughput due to a larger overhead. Moreover, behaviour of measured delay reported in [2] shows how XDQRAP with $N_{ms} = 3$ is close to a M/D/1 queue, consequently, the use of 2 or 3 minislots will be the more reasonable choice depending on network parameters such as capacity, delay and minislot and dataslot length.

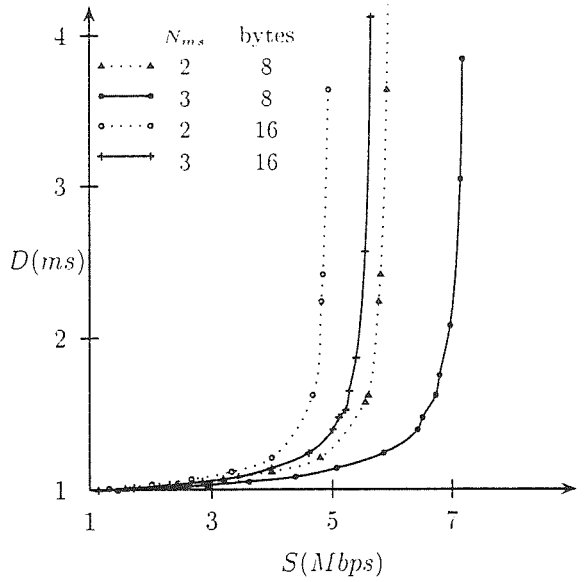


Figure 1: Access delay for 2 and 3 minislots of 8 and 16 bytes length.

Considering this values, figure 1 shows the access delay (D) in front of throughput for diferent values of the number of minislots (2 or 3) and their size (8 or 16 bytes). The access delay is measured as the time elapsed from the arrival of a packet at the station to the arrival of such a packet to the HeadEnd. The parameters used for simulations are the following: 15 stations, 10 Mbps for the upstream capacity, dataslot length equal to 64 bytes, mean propagation delay equivalent to 50 Km and a poisson distributed traffic. It is important to note that simulations have been performed without considering the propagation delay. This means that no tecniques as interleaving [12] have been applied, and that mean propagation delay has been added to the obtained results. Below, in Section 3, we will take up again this subject and will introduce a different solution to those reported until now.

Results from Figure 1 show the improvement of delay versus throughput due to the use of a third minislot. It does not matter if the size of the minislots is 8 or 16 bytes. Nevertheless this improvement for sizes of 8 bytes (1.2 Mbps) is larger than for minislots of 16 bytes (0.7 Mbps).

The second factor to be taken into account of the results from [5] is the round trip/slot time ratio. In this case, this ratio is equal 2, but considering requeriments from [3] where a maximum distance of 80 Km from furthest station to HeadEnd are determined and considering also the possibility of faster rates at the upstream channel, such as 10 Mbps, a slot size composed of 64 bytes of data and 8 bytes foreach of the three minislots will lead to a round trip/slot time ratio larger than 100.

This ratio, also known as interleaving factor, will produce a severe degradation on the throughput of an XDQRAP network if the transmission procedures are carried out according to the requests over a certain slot. We mean that it is not an optimal procedure do nothing while expecting the results of a request that will arrive several slots later. A solution reported to such an effect [12], proposes the use of N request queues, where N equals the interleaving factor, and a transmission queue for a determined MAC level. We argue that this solution is feasible for small values of N, but complicates excessivelly the MAC layer for large interleaving factors. In this sense, we propose in the next section a modification of the XDQRAP protocol procedures in order to use only two queues as the original proposal, one for request and another for transmission, but allowing activity to the station over the request an transmission queues for everyone of the N slots.

3 An implementation of the interleaving solution

As mentioned in the previous section, interleaving is a general procedure which allows to a MAC protocol

working in a slotted basis to operate under the same conditions as if the network was N times shorter. The main problem introduced by this general approach is that the complexity of the MAC layer is multiplied by a factor depending on N. The principal advantages are twofold:

- The delay access is the same as it was for a N times shorter network plus the propagation delay.
- Throughput is not affected by N.

In this section we propose an alternative implementation of interleaving which is applicable to the XDQRAP protocol and preserves its advantages but mantains the number of queues by MAC layer (2) despite introducing a little overhead on downstream slots and increasing slightly the complexity of the queue management algorithms. Figure 2 is an exemple of the XDQRAP operations considering 2 stations, A and B, which are located to 1 and 3 slots far from HeadEnd, and two minislots. The interleaving factor considered in this example is 6.

Without loss of generality and for for the sake of simplicity, the HeadEnd to station delay has been suposed to be an integer number of slot time.

Downstream slots reflect 3 fields. The most right is the slot number. The rest are the feedback information that reflects the status of the upstream minislots, S: single, C: collided and $-$: null reservation. A fourth field will be needed to inform the stations if the corresponding upstream slot has carried a data slot or not. This case has been reflected shadowing the downstream slot.

Upstream slots are also depicted with 3 fields. The left most is the slot number, not necessary but included here for a best understanding. The remaining fields correspond to the minislots. Reservations in minislots are indicated with the lower case station letter plus a number.

Finally, the request and transmission queues contain the reservation and its corresponding slot number. A description of the involved events in this exemple, with the same time notation as in Figure 2 (numbered by time slot), is the following:

1,3 HeadEnd sends numbered slots in downstream.

4 B makes a reservation in the first minislot.

6 A makes a reservation in the first minislot. A collission is produced. B reserves for another arrival in slot 3.

7 A detects the collission by means of the downstream slot, identifies it as own, and inserts it in the request queue for further reservations.

8 A makes another reservation corresponding to a second arrival and B performs a reservation for a third arrival.

9 B detects an own collission on slot 1 and A on slot 3. Both RQ are updated.

10 B retries a reservation on slot 1 according to RQ.

11 B detects the own collission on slot 3. A detects a foreign reservation on slot 5, updating the transmission queue.

12 A and B reserve over slot 1 and 3 respectivelly following RQ.

13 A updates TQ when a foreign and a own reservation are detected on first and second minislot respectivelly. B is also informed of its succesful reservation on slot 5.

14 A performs a reservation on slot 3 as indicated in RQ. B transmits its third arrival in slot 5 according to TQ.

15 Both, A and B include accepted request in TQ.

17 A transmission in slot 1, produced in t=16, is depicted. When slot 1 has to be transmitted, B searchs for the first reservation in TQ corresponding to slot 1. Also, A deletes from TQ the foreign reservation on slot 5 because on downstream, slot 5 is marked as containing data slot (b3).

19 A and B perform the same operation than the above point. They search for the first reservation in TQ than equals the downstream number slot and deletes it.

Actions performed from time 20 to 23 are equivalent to those described above.

Appendix A contains a more formal description of the XDQRAP procedures from the point of view of a generic station. To close this section, Figure 3 depicts the access delay measured under simulation of the interleaving solution described above for 2 and 3 minislots, being compared with those reported in Section 2. As shown, the delay performance for the interleaving algorithm is quite good, taking into account that a large network (100 Km) has been choosen.

4 Support for CBR service

The Constant Bit Rate (CBR) service category is used by connections that request a fixed amount of bandwith, continuosly available during the connection time and characterized by a Peak Cell Rate (PCR) value. Consequently, CBR service is intended to support real-time applications, such as voice, video and circuit emulation services, where the network commits to provide a maximum Cell Transfer Delay (CTD) and Cell Delay Variation (CDV) [7],[1].

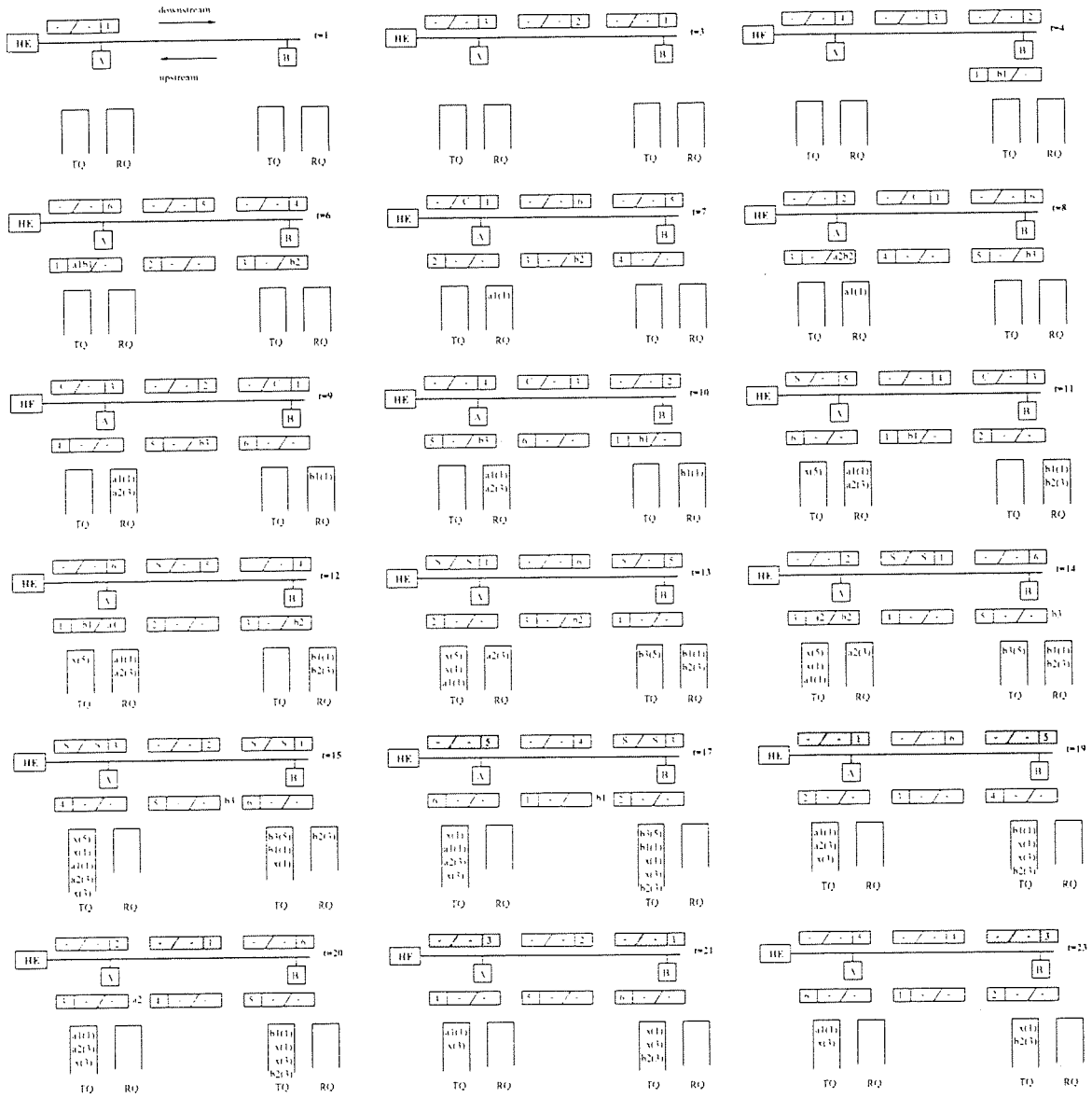


Figure 2: Exemple of the XDQRAP operation using interleaving.

In this section we present a method which offers CBR service where requested by stations. Next we expose some results of the delay variation and fairness of CBR traffic under different simulation scenarios.

At this point, with interleaving operation and no CBR support, downstream frame format requires two kind of fields as shown in figure 4, a field containing the sequence number and as many fields as minislots to indicate the status of the upstream minislots. In this sense, a sequence number of 8 bits will be enough to arrange in sequence as many slots as needed under conditions reflected in Section 2. Finally, 2 bits foreach minislot will indicate the reservation performed in the upstream channel, empty, single or collission.

Our solution to allow CBR services on XDQRAP networks is the following:

- Stations will ask to HeadEnd for CBR bandwith allocation using the upstream channel bymeans tranmission of some MAC control packets. Reservation will be specified on a preestablished unit of measurement, for exemple, number of CBR slots / slot time.

- HeadEnd will accept or reject the CBR bandwith request according to a simple rule. Being CBR_i the number of CBR slots in a slot time reserved by station$_i$, and being C the fraction of the total bandwith reserved for CBR services, the following

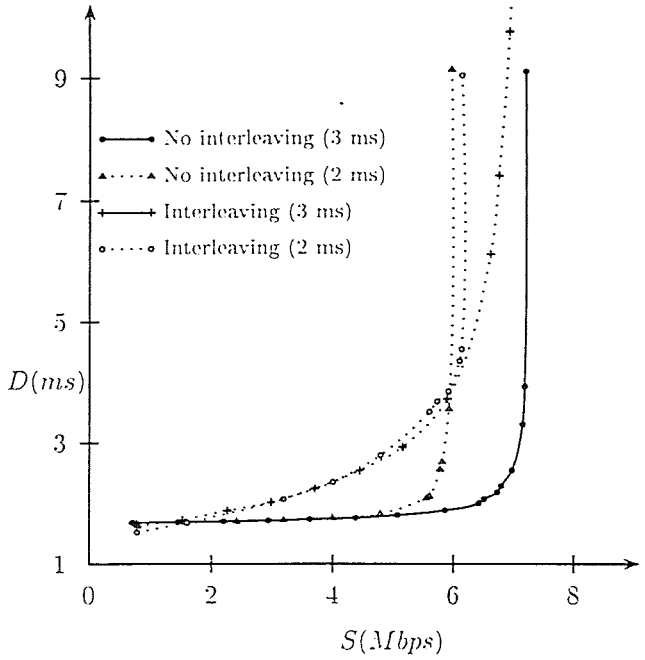


Figure 3: Acces delay comparison for a 100 Km network with and without interleaving

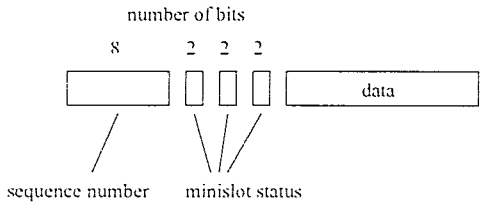


Figure 4: Downstream frame format with no CBR support

condition must be applied,

$$\sum_i CBR_i \leq C.$$

- Once a reservation is accepted by the HeadEnd, it will indicate in the downstream channel, at every slot, if the current slot is reserved or not for CBR transmission, and if applicable, which is the station that will transmit in this slot.

This solution implies two considerations. The first one is that a policy to decide slot reservations must be performed by the HeadEnd. The second one, is that the frame format in the downstream channel must be modified, containing a new field of 1 bit to indicate if the slot is reserved for CBR transmission or not, and a second field, if slot is reserved, to indicate the station MAC address. These changes are depicted in Figure 5. A description of three HeadEnd policies to schedule CBR reservations, called sequential, Round-Robin and Golden-Ratio, are detailed in Appendix B.

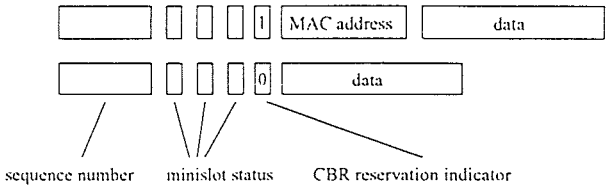


Figure 5: Downstream frame format with CBR support

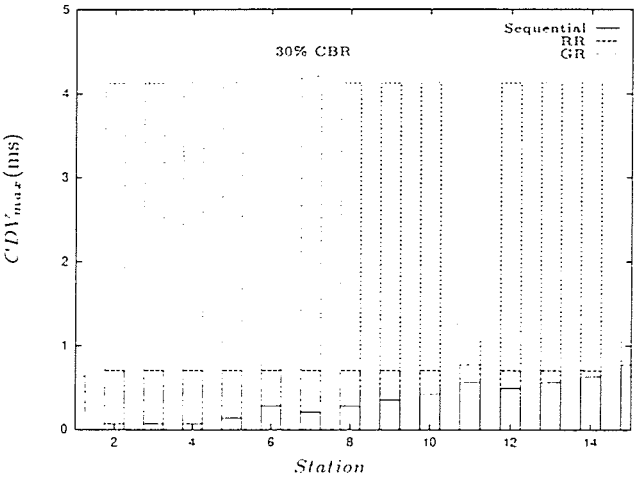


Figure 6: Maximum cell delay variation for 30% of CBR load.

The guideline to examine the jitter performance of the XDQRAP protocol has been the Peak Cell Delay Variation (CDV_{max}) under different loads of CBR traffic. Figures 6, 7 and 8 shows CDV_{max} and fairness for the three policies mentioned above, corresponding to a load of 30, 60 and 90% respectively. Such a loads have been achieved taking 15, 30 and 45 stations, asking for CBR reservations of 384 and 64 Kbps.

From the results obtained, it can be concluded that the best choice is RR or GR because the Sequential policy seems to perform better only for loads less than 30% but unfairnessly.

The CDV_{max} obtained for GR is not dependent of the load, giving 4 ms of maximum CDV_{max} for the allocations corresponding to the 64 Kbps channels and 1 ms for the 384 Kbps channels. On the other hand, the measured CDV_{max} for the RR policy shows no differences between channels of 64 and 384 Kbps, being at a load of 90% almost 3 ms.

These results show an excellent performance of CDV_{max} and fairness, even at high levels of CBR load such as 90%, mantaining this performance independently of the total network load.

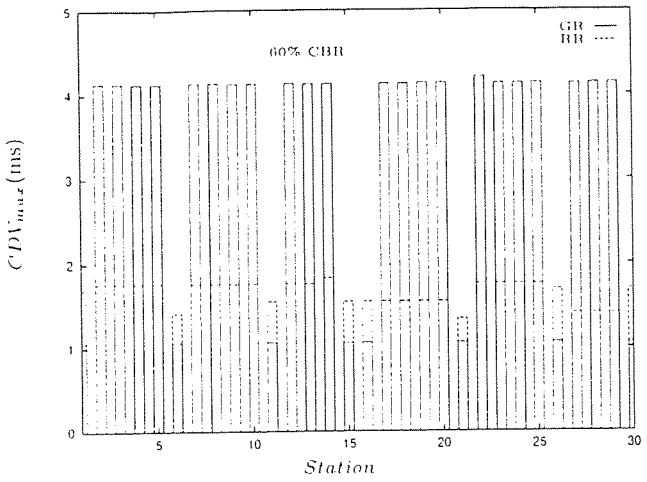


Figure 7: Maximum cell delay variation for 60% of CBR load.

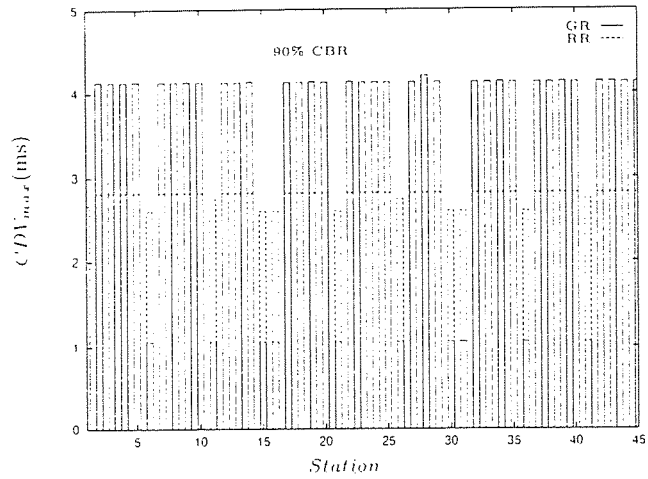


Figure 8: Maximum cell delay variation for 90% of CBR load.

5 Conclusions and future work

This paper presents two major contributions to the XDQRAP protocol. The first one is an alternative solution to the interleaving procedures which is demonstrated to perform well with long range networks. The measured acces delay under traffic poisson models shows an increment of such a delay due only to the propagation delay, exactly in the same way as it is suposed to be applicable to the traditional interleaving solution, but in this case, the number of queues is mantained at two.

In this sense, we are in process to obtain exhaustive access delay results under different models of traffic sources, video, audio and data in order to determine a more realistic performance of the XDQRAP protocol.

The second contribution has been to determine a mechanism which allows to offer CBR service over XDQRAP networks. This mechanism is very simple and it is able to operate with many allocation policies giving excellent results of fairness and CDV. Much work has to be done at this point to define the possible mechanisms which will allow XDQRAP to support variable and available rate services, determining the corresponding quality of service parameters in order to stablish comparisons with the actual proposed MAC protocols.

A Station transmission algorithms

This appendix details the XDQRAP algorithms, explained above, for a generic station j. The global procedure is started when the station receives a slot on the downstream channel. As can be noted, the procedure is splitted into two actions, both dependent on the downstream feedback which reflects the status of the upstream reservations and transmissions. The first one dedicated to prepare transmission over the minislots and data slot of the upstream channel, and the second one devoted to update the transmission and requests queues.

If we take τ and τ_j as the network propagation and the HeadEnd/Station$_j$ delays respectively, and taking

$$D = \left\lceil \frac{2\tau}{st} \right\rceil,$$

being st the slot time on the upstream channel, downstream slots will be numbered in a cyclic manner from 1 to D. Taking also

$$D_j = \left\lceil \frac{2\tau_j}{st} \right\rceil,$$

we can define the back round trip time of station$_j$ (brt_j) as

$$brt_j = 2(\tau - \tau_j).$$

Back round trip is important because if station$_j$ wants to transmit in slot number x, the transmission will be performed when slot number y is received in the downstream, being

$$y \equiv x + \lceil brt_j \rceil \pmod{D}.$$

constant
  D,D_i, N_{ms}: integer;
  $brt = 2(D - D_i)$;
structure job_queue
  pk : packet;
  oi : {0,1}; /* Owner indicator */
  n_s : [1..D];
structure slot_down
  n_s : [1..D];
  da: **array**[1..N_{ms}] **of** $\{C, S, E\}$;

```
    ti : {0,1};          /* transmission indicator */
variable
    ua: array[1..N_ms, 1..D] of {0,1};
    i: integer;
    RQ, TQ : queue;
    job_TQ, job_RQ : job_queue;

action transmission in (sd: slot_down)
    if (empty(RQ)) then
        if something to transmit then
            i := random(1,N_ms);
            reserve minislot i;
            ua[i][sd.ns + brt] := 1;
        if (not empty(TQ)) then
            search first job in TQ such that
                job_TQ.ns = sd.ns + brt;
            if (job_TQ.oi = 1) then
                transmit job_TQ.pk;
    else
        if (not empty(TQ)) then
            search first job in TQ such that
                job_TQ.ns = sd.ns + brt;
            if (job_TQ.oi = 1) then
                transmit job_TQ.pk;
        search first job in RQ such that
            job_RQ.ns = sd.ns + brt;
        if (job_RQ.oi = 1) then
            i := random(1,N_ms);
            reserve minislot i;
            ua[i][sd.ns + brt] := 1;

action queue_update in (sd: slot_down)
    if (empty(RQ)) then
        for i:= 1 to N_ms
            if (sd.da[i] = S)
                if(ua[i][sd.ns + brt] = 1)
                    job_TQ.pk := new packet;
                    job_TQ.oi := 1;
                else
                    job_TQ.pk := NULL;
                    job_TQ.oi := 0;
                job_TQ.ns := sd.ns;
                insert job_TQ in TQ;
    else
        for i:= 1 to N_ms
            if (sd.da[i] = S)
                if(ua[i][sd.ns + brt] = 1)
                    search first job in RQ such that
                    job_RQ.ns = sd.ns + brt;
                    copy job_RQ into TQ;
                else
                    insert new job in TQ with
                    (oi := 0, pk:=NULL, ns := sd.ns)
        if (sd.ti = 1) then
            search first job in TQ such that
            job_TQ.ns = sd.ns;
            delete job;
        if (empty(RQ)) then
            for i:= 1 to N_ms
                if (sd.da[i] = C)
                    if(ua[i][sd.ns + brt] = 1)
                        insert new job in RQ with
                        (oi := 1, pk:=new packet, ns := sd.ns)
                    else
                        insert new job in RQ with
                        (oi := 0, pk:=NULL, ns := sd.ns)
        else
            for i:= 1 to N_ms
                if (sd.da[i] = C)
                search first job in RQ such that
                job_RQ.ns = sd.ns + brt;
                    if(ua[i][sd.ns + brt] = 1)
                        insert new job in RQ with
                        (oi := 1, pk:=job_RQ.pk, ns := sd.ns)
                    else
                        insert new job in RQ with
                        (oi := 0, pk:=NULL, ns := sd.ns)
                delete job_RQ;

algorithm xdqrap
    variable sd: slot_down;
    while slot_down received(sd)
        queue_update in (sd);
        transmission in (sd);
```

B HeadEnd CBR reservation scheduling policies

The first proposed policy to schedule CBR reservations (sequential) is simple. Taking the number of CBR slots reserved by station$_j$ in a time slot, (CBR_j), and considering N the number of stations asking for CBR services, HeadEnd defines a N-dimensional array containing,

$$\left\lceil \frac{1}{CBR_j} \right\rceil,$$

expression which reflects the number of slots between two successive CBR reservations for station$_j$.

HeadEnd decrements by one the reservation array at every slot time, and a reservation is made when a member of the array falls below zero. The following pseudo-code details better such an algorithm.

```
algorithm HeadEnd_CBR_control
constant
    N: integer;
    reservation: array[1..N] of integer;
variable
    control: array[1..N] of integer;
    t,i: integer;
```

```
t:=0;
for i:=1 to N
   if (control[i] >= reservation[i]-1 and t=0) then
      t:=1;
      CBR reservation for station_i
   else
      control[i]++;
```

The two remaining policies, Round-Robin and Golden-Ratio are detailed in [10]. Round-Robin is a refinement of the sequencial algorithm as far as the loop is initiated with the next allocated station instead of 1. This method improves the fairness as proved in section 4. Golden-Ratio tries to find an optimal sequence of allocation which minimizes the cell delay variation based on the known Golden Ratio $\phi^{-1} = (\sqrt{5} - 1)/2$.

References

[1] The ATM Forum Technical Committee. *Traffic Management Specification. Version 4.0*, April 1996.

[2] Cèsar Fernàndez, Sebastià Sallent, Carles Mateu, and Lluís Gutiérrez. Análisis del protocolo MAC XDQRAP para redes de acceso de banda ancha multiservicio IEEE 802.14. *XI Symposium Nacional de la Unión Científica Internacional de Radio*, I:432–435, Septiembre 1996.

[3] IEEE Project 802.14 Working Group. *IEEE P 802.14 Cable-TV Functional Requirements and Evaluation Criteria*. March 1995. Document IEEE802.14/94-002R2.

[4] IEEE Project 802.14 Working Group. *A MAC proposal for 802.14*. November 1995. Document IEEE802.14/95-134.

[5] IEEE Project 802.14 Working Group. *On the Issue of Frame Fromat (Size and Contention/Data Slots Ratio*. May 1996. Document IEEE802.14/96-159.

[6] Tim Kwok. A vision for residential broadband services: ATM-to-the-home. *IEEE Network*, 9(5):14–28, Septembre/October 1995.

[7] Livio Lambarelli. ATM service categories: The benefits to the user. CSELT, Torino, Italy, 1996. Available at `http://www.atmforum.com/atmforum/whitepapers.html`.

[8] William Pugh and Gerald Boyer. Broadband acces: Comparing alternatives. *IEEE Communications Magazine*, 33(8):34–46, August 1995.

[9] Srinivas Ramanathan and Riccardo Gusella. Toward management systems for emerging Hybrid Fiber-Coax access networks. *IEEE Network*, 9(5):58–68, Septembre/October 1995.

[10] Z. Rosberg. Cell multiplexing in ATM networks. *IEEE Transactions on Networking*, 4(1):112-122, February 1996.

[11] Mario Vecchi. Broadband networks and services: Architecture and control. *IEEE Communications Magazine*, 33(8):24-33, August 1995.

[12] Chien-Ting Wu and Graham Campbell. Interleaving DQRAP with global TQ. Illinois Institute of Technology. DQRAP Reseach Group Report 94-4, December 1994.

[13] Chieng-Ting Wu and Graham Campbell. Extended DQRAP (XDQRAP). A cable TV protocol functioning as a distributed switch. Illinois Institute of Technology. DQRAP Reseach Group Report 94-2, 1994.

[14] Wenxing Xu and Graham Campbell. A distributed queuing random access protocol for broadcast channels. Illinois Institute of Technology. DQRAP Reseach Group Report 90-1, 1990.

Modelos y Simulación de Redes de Banda Ancha

MODELADO DE VÍDEO VBR ORIENTADO A ESCENA

E. Casilari, M. Lorente, A. Reyes, A. Díaz Estrella, F. Sandoval
Dpto. Tecnología Electrónica, E.T.S.I. Telecomunicación,
Universidad de Málaga, Campus de Teatinos, 29071 Málaga.
Tfno.: 34-5- 2132755; FAX 34-5-2131447; E-mail: casilari@dte.uma.es

Abstract

In this paper a new model for variable bit rate (VBR) video traffic is presented. The model, which could be used as a traffic generator, considers three time scales: scenes (for periods of several minutes) , groups of pictures or GOPs (for periods of half a minute) and frames (for periods of some miliseconds). To model the scene changes a Markov chain is used. For the GOP level a modification for the projected autorregresive (PAR) model is proposed so that the fitting of the autocorrelation function is improved. The GOP is divided into frames following three different strategies. The model is utilised to imitate a real MPEG video signal, showing that it is able to accurately capture the behaviour of the real traffic in a queue.

1. Introducción

Entre los servicios con más proyección que se transmitirán por las redes de banda ancha, sobre tecnología ATM, se encuentran los diversos servicios de vídeo (teleconferencia, vídeo bajo demanda, HDTV ...). Dichos servicios, de naturaleza multimedia, constituirán, pues, una fuente de tráfico de gran importancia, por lo que, para el diseño, dimensionado o simulación de las futuras redes, así como para validar las diversas estrategias de vigilancia (UPC), control de admisión (CAC) o asignamiento dinámico de recursos, se hace necesario establecer generadores multiplexables y estadísticamente independientes que puedan aproximar de una manera precisa las características estadísticas del tráfico de vídeo.

Una secuencia de vídeo consiste en una serie de fotogramas (*frames* o *pictures*) conteniendo una matriz bidimensional de pixels. Los fotogramas pueden ser codificados con velocidad constante y calidad variable (vídeo CBR) o, por el contrario, se puede decidir fijar cierta calidad (una relación S/N constante) a cambio de que la cantidad de información por fotograma cambie en función de la complejidad y el movimiento de la imagen. Este último tipo de vídeo, que genera un tráfico variable (VBR), puede beneficiarse del multiplexado estadístico que ofrece ATM, optimizando la utilización del ancho de banda disponible. De ahí que, en los últimos tiempos, su análisis y modelado haya suscitado diversos estudios, cuyos resultados a menudo se extienden a tareas tales como los controles de vigilancia y admisión o la asignación dinámica de ancho de banda.

En este artículo se propone un modelo completo de tres niveles para imitar el comportamiento del tráfico de vídeo con codificación MPEG, una de las normas más extendidas y que más aceptación está teniendo en el ámbito de los servicios de vídeo. El esquema propuesto incluye el modelado de cambios de escenas, ajustando así dependencias a largo plazo (LRD), y propone una mejora al modelo PAR [1] para generar la señal, en este caso número de bits/GOP (*Group of pictures*), dentro de cada escena. Asimismo se proponen diversas estrategias para dividir la información correspondiente a cada GOP en los diversos fotogramas que lo componen, extendiendo a continuación el modelo a una tercera escala de tiempo (el fotograma), que tiene en cuenta la estructura intrínseca del codificador. En la fase de simulación, el modelo se utiliza para aproximar una película real con codificación MPEG, mostrándose que no sólo aproxima con exactitud sus estadísticos sino también su comportamiento al ser encolado.

2. Esquema de codificación MPEG

El comité MPEG (siglas inglesas para el "*Moving Pictures Experts Group*") surgió allá por el año 1988 con el objeto de crear un estándar de codificación de vídeo y audio para discos compactos. La sintaxis de codificación que se alcanzó al cabo de tres años de trabajo se mostró válida no sólo para dicho tipo de vídeo (MPEG-I) sino que rápidamente fue extendida a otros servicios cuya información contiene imágenes móviles: televisión de radiodifusión (MPEG-2), televisión de alta definición (MPEG-3) e incluso servicios que exigen un alto nivel de compresión a costa, si es preciso, de pérdidas fuertes de calidad en la señal, como puede ser la videotelefonía (MPEG-4).

La norma MPEG propone un algoritmo genérico de codificación de audio y vídeo con el que se consiguen relaciones de compresión que pueden oscilar desde 6:1 a 100:1. Como ya se ha dicho, la norma está abierta a cualquier servicio de vídeo, independientemente de la velocidad binaria que impliquen y la calidad de servicio que exijan. En el caso concreto de la especificación MPEG-1, el algoritmo es fuertemente asimétrico en la medida en que la complejidad computacional es mucho mayor a la hora de comprimir que a la hora de descomprimir. Esto lo hace especialmente atractivo para el almacenamiento de películas en CD-ROM así como para servicios de vídeo bajo demanda, dos ejemplos típicos en los que los procesos de descompresión son muy frecuentes en tanto que la compresión sólo se produce una vez.

La codificación MPEG basa sus altas tasas de compresión, principalmente, en el hecho de que es capaz de eliminar dos tipos de redundancias existentes siempre en una señal de vídeo: por un lado la fuerte redundancia espacial (*intraframe*) que existe entre pixels contiguos dentro de un mismo fotograma y, por otra parte, la redundancia temporal (*interframe*) que se establece en pixels situados en la misma posición dentro de fotogramas consecutivos. Con el objeto de esta doble compresión la norma define tres tipos de fotogramas (genéricamente *frames* o, más estrictamente, *pictures* según denominación propia de la norma) con codificaciones distintas:

-En primer lugar, tendríamos los fotogramas tipo *I* (o *intraframe*) que, mediante técnicas como la transformada discreta del coseno (DCT), suprimen únicamente la redundancia espacial existente en ellos mismos. Constituyen la unidad mínima decodificable en tanto que su decodificación no precisa de otros fotogramas. Por lo tanto pueden ser entendidos como "puntos de enganche" a la secuencia de vídeo que se esté transmitiendo.

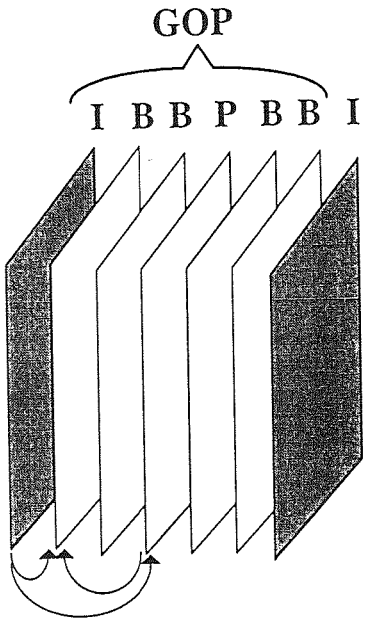
-En segundo lugar, existen también los denominados fotogramas tipo *P* (o *predictive*) que, añaden a la codificación propia de los *I* la supresión de la redundancia temporal existente con el fotograma *I* o *P* anterior. Esto se lleva a cabo mediante técnicas que incluyen la definición de vectores de movimiento (*motion vectors*).

- Finalmente, la norma define fotogramas *B* (o *bidirectional*) cuya filosofía de funcionamiento es similar a la de los *P*, con la diferencia de que en este caso la compensación de movimiento se hace en ambos sentidos, es decir, con respecto al pasado (fotograma *I* o *P* previo) y al futuro (fotograma *I* o *P* siguiente)

Para decodificar tanto los fotogramas *P* como *B*, se requiere, pues, de la presencia de otros fotogramas anteriores e incluso (caso de los *B*) posteriores.

Estos tres tipos de fotogramas se ordenan, habitualmente, en secuencias fijas denominadas GOPs (*Group of Pictures*). El GOP se puede definir como la cadena de fotogramas existentes entre dos fotogramas *I* consecutivos. La norma MPEG no define la estructura del GOP sino que deja al fabricante libertad para decidirla. Una estructura muy extendida en USA y Japón es la del tipo *IBBPBBPBBPBB*. No obstante, se habla de que en un futuro los codificadores MPEG optimizarán la estructura del GOP en función de la señal a codificar. Para ilustrar la secuenciación de fotogramas en un vídeo MPEG la figura 1 representa lo que podría ser la estructura de un GOP. En dicha figura las flechas representan dos ejemplos (uno para un fotograma *P* y otro para uno B) de las relaciones existentes con otros fotogramas a la hora de la decodificación.

Figura 1. Estructura de una secuencia MPEG



Como se puede comprender por lo hasta ahora dicho las características de la información transportada por una secuencia de vídeo MPEG dependen tanto de la propia naturaleza de la señal codificada (por ejemplo, una película con escenas de distinto nivel de movimiento o complejidad de las imágenes) como de la propia estructura del codificador. Así, hay que hacer notar que puesto que cada GOP comienza con un fotograma *I* (que no elimina la redundancia temporal) existe una fuerte correlación en la información contenida en GOPs sucesivos. Y por otro lado es destacable también el hecho de que al poseer diversas filosofías de codificación, el volumen de tráfico de cada fotograma viene muy determinado por el tipo del mismo. De este modo, los fotogramas *I* transportan más información (en bits) que los *P*, y éstos más que los *B*. En consecuencia, cualquier modelo general que se quiera establecer para el tráfico generado por un codificador MPEG debe ser consciente de la existencia de diversas escalas de tiempo (escenas, GOPs, fotogramas) y de las particularidades de las mismas. En este trabajo se plantea un modelo estructuralista o de "caja blanca", en el que se trata de imitar directamente la estructura propia del tráfico a cada una de las escalas, teniendo en cuenta la realidad física que se esconde detrás de cada escala, justificando su comportamiento.

3. Sistema Propuesto

El objetivo final de modelar cierta fuente de tráfico, considerada como una señal aleatoria *s[n]*, que representa la cantidad de información en cierta escala de tiempo (en este caso el GOP posteriormente dividido en fotogramas), es la

generación de otra señal *s'[n]* que sea capaz de aproximar el comportamiento en una cola (pérdidas, retrasos, jitter) que posee la señal real. La bibliografía existente sobre análisis de vídeo VBR propone diversos modelos [1-5], entre los que podemos destacar el modelo TES, las cadenas de Markov o los modelos AR, ARMA y PAR. En general estos modelos son diseñados para ajustar estadísticos de primer orden (la función de distribución estadística $F_S(x)$, media, varianza) y algunos de los puntos iniciales de la función de autocorrelación $R_S[k]$. Su comportamiento en una cola suele imitar razonablemente bien el del tráfico a simular cuando el tamaño del buffer es pequeño, pero infraestiman la probabilidad de pérdidas en cuanto el buffer aumenta. Esto es debido a que estos modelos sólo ajustan las correlaciones a corto plazo (modelos SRD) y no consideran las dependencias a más largo plazo existentes en una secuencia de vídeo, es decir, el hecho de que la señal presenta una evolución por estados (escenas) con distinto nivel de actividad, que no es apreciable para una escala pequeña de tiempos [6-11].

Analíticamente, estas dependencias se pueden observar, entre otras maneras, en valores relativamente altos de la función de autocorrelación para retrasos k grandes. Dichos valores a largo plazo no pueden ser ajustados por los modelos anteriores cuyas autocorrelaciones son exponencialmente decadentes para retrasos suficientemente largos y, por tanto, tienden rápidamente a 0 conforme k aumenta. Para superar este problema se requiere, pues, un modelo que no sólo sea capaz de simular las variaciones a corto plazo (dentro de una escena) sino que también incluya un proceso estocástico que imite el tránsito por escenas con distinta actividad [5] [6] . En concreto, en este artículo se propone un modelo de tres niveles: escenas, GOP y fotogramas. Partiendo de una señal *s[n]* que indica la información contenida en cada GOP y que se podría estimar como:

$$s[n] = \sum_{j=1}^{G} p[(G \cdot (n-1) + j]$$

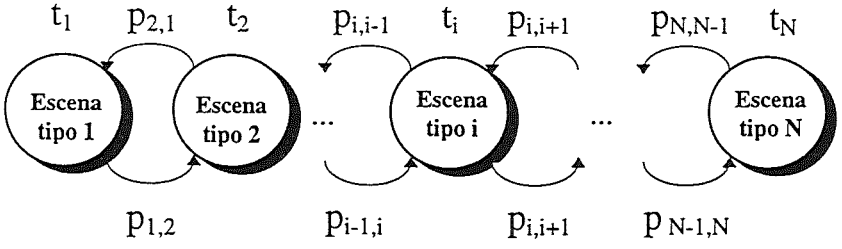
siendo *p[j]* la información del fotograma j de la señal a modelar y G el tamaño fijo del GOP, el modelo incluye por un lado los cambios de escena, para después volver a dividir la señal generada en fotogramas.

3.1 Modelo a nivel de escena.

Como ya se ha mencionado anteriormente, una de las características esenciales de una señal de vídeo (especialmente cuando se trata de películas) es la existencia de escenas de distintos niveles de actividad. El bajo o alto nivel de actividad de una escena dentro de una película con codificación MPEG puede ser debido a dos razones principales: la complejidad de la imagen y el grado de movimiento de la propia escena. Imágenes muy complejas (con escasa redundancia espacial) generarán fotogramas I con un volumen de tráfico grande mientras que escenas con alto grado de movimiento contendrán una redundancia temporal menor (en tanto que los fotogramas varían mucho de unos a otros) y, en consecuencia, la codificación de fotogramas P y B obligará también a volúmenes de tráfico más alto. Desde el punto de vista del nivel de GOP o superiores, cualquiera de estas dos razones provoca que la unidad de tráfico a considerar sufra una serie de fluctuaciones a más largo plazo, ya que una escena, definida no tanto en su sentido real de cambio de plano como por un entorno de GOPs consecutivos con un nivel de actividad similar, puede representar órdenes de tiempo del orden de varios minutos. Esto obliga a establecer un proceso subyacente al nivel de GOP que caracterice esta componente de muy baja frecuencia que presenta la señal. Diversas estrategias [6][8][10] se han propuesto para solventar este problema. En nuestro caso, para modelar las variaciones de escena se propone utilizar una cadena discreta de Markov $M_K=\{1, 2,...N\}$ donde el número de estados N coincide con el número de grados de actividad que se pretende distinguir (Fig. 2). En dicha cadena las conmutaciones entre estados se producirán según una matriz de conmutación P, mientras que el tiempo de permanencia para cierto estado I se aproximará mediante una distribución exponencial de media $\bar{t}_i$.

La elección de una distribución estadística para modelar la duración de las escenas no es fácil en tanto que, si consideramos que una película dispone de un número más que finito de escenas, carecemos de suficientes muestras para una correcta caracterización de las duraciones de cada tipo. Por

Fig. 2. Modelo a nivel de escena



tanto es ésta una cuestión todavía abierta en este ámbito. Para conocer tanto P como $\bar{t}_i$ para cada estado, se ha de dividir la señal $s[n]$ a nivel de GOP en una secuencia de escenas $m[n]$, donde $m[i] \in \{1,...N\}$ indica el tipo de escena a la que pertenece el GOP i. Para calcular $m[n]$ se debe baremar de algún modo no sólo el grado de actividad (bits) que posee el propio GOP sino también el de los GOPs adyacentes. Con ese objeto se genera la señal $s_w[n]$ como el resultado de pasar $s[n]$ por un filtro de media móvil con una ventana de tamaño W:

$$s_w[n] = \frac{1}{W+1} \sum_{n-W/2}^{n+W/2} s[n]$$

Cuantizando dicha señal $s_w[n]$ en N niveles uniformes y numerándolos desde 1 a N, se obtendría $m[n]$.

$$m[n] = 1 + Parte\ entera\{\frac{s_w[n] - min(s[n])}{B}\}$$

siendo $B = \frac{max(s[n]) - min(s[n])}{N}$

A partir de $m[n]$ se podría computar la probabilidad p_{ij} de transición entre dos estados diferentes i y j:

$$p_{ij} = \frac{No.\ de\ Transiciones\ desde\ i\ a\ j}{No.\ de\ Transiciones\ desde\ i\ a\ cualquier\ estado}$$

Sabiendo que en el fotograma k perteneciente al tipo de escena i se produce un salto de escena (transición desde i) si:

$$m[k] = i \quad y \quad m[k+1] \neq i$$

y, en concreto, se produce un salto desde las escena tipo i a la tipo j si:

$$m[k] = i \quad y \quad m[k+1] = j$$

De este modo quedaría definida la matriz de transición P:

$$P = \begin{vmatrix} 0 & p_{12} & \cdots & p_{1j} & \cdots & p_{1N} \\ p_{21} & 0 & \cdots & p_{2j} & \cdots & p_{2N} \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ p_{i1} & p_{i2} & \cdots & p_{ij} & \cdots & \cdot p_{iN} \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ p_{N1} & p_{N2} & \cdots & p_{Nj} & \cdots & 0 \end{vmatrix}$$

En la práctica se comprueba que, si la ventana es suficientemente grande, $p_{ij} \neq 0$ sólo si $j=i+1$ ó $j=i-1$, lo cual implica que las transiciones sólo se producen entre estados adyacentes.

Se establece una probabilidad nula de permanencia en cada estado ($p_{ii} = 0 \quad \forall i$) puesto que para cada estado el tiempo de permanencia (en GOPs) se modelará mediante series numéricas independientes.

De igual manera se podría hallar la secuencia de tiempos de permanencia $t_i[l]$ para cada estado i con $i \in \{1,...,N\}$ y $l \in \{1,...,$n° de escenas de tipo $i\}$. A partir de estas secuencias se puede calcular los tiempos medios de permanencia $\bar{t}_i$ en cada estado de actividad , con el objeto de definir por completo la distribución exponencial que los modela, o incluso probar algún otro tipo de función estadística (de tipo sub-exponencial [8] como la de Pareto).

Con estos datos el modelo a nivel de escena quedaría completamente definido.

3.2 Modelo a nivel de GOP.

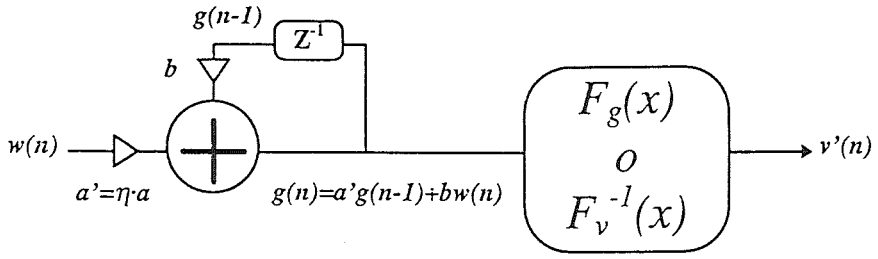
Una vez que se ha dividido la señal $s[n]$ en escenas, agrupamos los GOPs pertenecientes al mismo grado de actividad, de modo que se obtendrán N vectores (v_i con $i \in \{1,...N\}$). En estos vectores v_i se han anulado las dependencias a largo plazo y por tanto pueden ser simulados mediante los modelos convencionales ya comentados. En este artículo se utiliza y propone una mejora del modelo PAR [3] o autorregresivo proyectado, que se ha representado en la figura 3.

En el modelo AR o autorregresivo de primer orden se genera una señal $g[n]$ a partir de un ruido gaussiano blanco $w[n]$ y una muestra de la propia señal retrasada $g[n-k]$:

$$g[n] = a \cdot g[n-k] + b \cdot w[n]$$

Los parámetros a y b, son calculados de tal modo [1] que la señal de salida gaussiana $g[n]$ ajuste la correlación de la secuencia $v[n]$ a simular para el retraso k elegido, que normalmente en el caso de vídeo a nivel de GOP debe ser el primero (k=1), así como la media y la varianza de $v[n]$. El problema de este modelo es que la salida es gaussiana y, por tanto, no ajusta la función de densidad de la señal a imitar sino, como se ha dicho, sólo la media y la varianza. Para solucionar esto, el modelo PAR propone proyectar $g[n]$ sobre su propia función de distribución $F_g(x)$ gaussiana de tal manera que se obtendrá una serie uniformemente distribuida entre 0 y 1 con características de correlación idénticas a las de $g[n]$. Proyectando esta señal uniforme sobre la inversa de la función de distribución $F_v(x)$ de la señal

Figura 3. Esquema del modelo PAR modificado



a imitar, se obtendrá una señal v'[n] que ajustará perfectamente $F_v(x)$. La operación total sobre el ruido gaussiano es la equivalente a la composición de funciones F_g o F_v^{-1}:

$$v'[n] = \left(F_g \; o \; F_v^{-1}\right)(g[n])$$

El problema de esta proyección sobre una función en principio no lineal (F_V^{-1}) es que la aproximación de la correlación que realizaba g[n] se ve debilitada. Para paliar este efecto se propone modificar el modelo mediante un reajuste del parámetro a del modelo AR. Multiplicando a por un factor η, ligeramente mayor a uno, reforzamos la correlación de la señal $g[n]$ en el punto de ajuste, de tal forma que, al ser proyectada, la deformación de la correlación se verá compensada. En concreto, proponemos elegir η como el cociente entre la correlación en el punto a ajustar ($R_g[1]$), que es la que presenta la salida $g[n]$ del filtro AR , y la que presentaría la señal de salida si no se efectuase esta corrección ($R_{v'}[1]$):

$$\eta = \frac{R_g[1]}{R_{v'}[1]}$$

3.2 Modelo a nivel de fotograma.

Tras generar la secuencia $s'[n]$ de GOPs, si se desea aproximar la señal real a una escala de tiempo menor, es preciso dividir la información de cada GOP en la serie de fotogramas correspondiente $p'[j]$, teniendo en cuenta que la codificación MPEG impone a esta serie unas características muy marcadas en la medida en que existen tres tipos distintos de fotogramas cuya ordenación dentro del GOP no es aleatoria.

Para la división del GOP en fotogramas se ha de seguir, evidentemente, el mismo patrón que haya fijado el codificador para la estructura del GOP, esto es, una secuencia determinada de fotogramas I, P y B. El problema radica en distribuir la información (el número de bytes) del GOP (v_{GOP}) entre los G distintos fotogramas que lo componen (1 fotograma I, G_P fotogramas P y G_B fotogramas B), de manera que se aproxime lo más fielmente las características de la muestra real. Para ello, proponemos tres métodos que serán comparados en la fase de simulación y pruebas.

Método 1. Reparto fijo y proporcional entre los tres tipos de fotogramas: la información del GOP es dividida en proporciones fijas entre los fotogramas (p'_I, p'_P y p'_B):

$$p'_I = c_I s' \; ; \quad p'_P = c_P s' \; ; \quad p'_B = c_B s'$$

donde c_I, c_P y c_B son los coeficientes que establecen el peso de cada fotograma dentro de cada GOP y se calculan, previamente, de la propia señal real $p[j]$, como la relación entre el tráfico global de cada tipo y el tráfico total:

$$c_I = \frac{\sum p_I}{\sum p} \; ; \quad c_P = \frac{1}{G_P}\frac{\sum p_P}{\sum p} \; ; \quad c_B = \frac{1}{G_B}\frac{\sum p_B}{\sum p}$$

donde p_I, p_P y p_B representan los fotogramas I, P y B de la señal real respectivamente. Lógicamente:

$$c_I + G_P c_P + G_B c_B = 1$$

Método 2. Para evitar la rigidez del primer método, en el que siempre se mantiene para cada fotograma el peso de los fotogramas I frente a los P y los B, se propone una variante en la que se utilizan los histogramas de los tres tipos de fotogramas para variar dinámicamente dichos pesos para cada GOP. El histograma de una señal, normalizado por el número de muestras registradas, constituye un estimador discretizado de la función de densidad estadística de dicha señal. Integrando esta función de densidad se obtiene la función de distribución cuya inversa puede ser utilizada para generar señales a partir de un ruido uniformemente distribuido (u) entre 0 y 1, tal y como se explicó en la sección 3.1. Así pues, tras calcular los histogramas de los tres tipos de fotogramas y, por tanto, sus funciones de distribución F_I, F_P y F_B (respectivamente para los de tipo I, P y B) y sus funciones inversas, se genera, para cada GOP, una muestra de tipo I (p'_I), G_P muestras ($p'_P(i)$) de tipo P y G_B muestras ($p'_B(i)$) de tipo B. Dichas generaciones se consiguen mediante la proyección del ruido uniforme sobre las funciones

inversas tabuladas ($F^{-1}_I(u)$, $F^{-1}_P(u)$, y $F^{-1}_B(u)$). Basándose en estas muestras se podría recalcular dinámicamente el valor de los pesos c_I, $c_P(i)$ (con $I \in [1, G_P]$) y $c_B(i)$ (con $I \in [1, G_B]$), determinando el porcentaje, sobre el total de los G fotogramas generados, que supone cada fotograma:

$$c_I = \frac{p'_I}{p'_I + \sum_i^{G_P} p'_P(i) + \sum_i^{G_B} p'_B(i)}$$

$$c_P(i) = \frac{p'_P(i)}{p'_I + \sum_i^{G_P} p'_P(i) + \sum_i^{G_B} p'_B(i)}$$

$$c_B(i) = \frac{p'_B(i)}{p'_I + \sum_i^{G_P} p'_P(i) + \sum_i^{G_B} p'_B(i)}$$

Método 3. El tercer método es una variante del segundo en la que se trata de incluir la correlación que existe entre los tráficos de los diversos tráficos dentro de cada GOP. En el método anterior el peso de cada fotograma es calculado independientemente de los demás, sin embargo, en las series MPEG reales se puede comprobar una fuerte correlación entre los distintos tipos de fotogramas, especialmente acusada en el caso de la correlación entre los fotogramas P y B (que resulta, de la fuerte dependencia existente entre la cantidad de información de ambos tipos de fotogramas y el grado de movimiento de la secuencia). Para aproximar esta característica se propone utilizar el método anterior pero esta vez generando el tráfico de los fotogramas P en función del generado por los I, y el de los B en función del generado a su vez por los P. Con este propósito se propone utilizar un filtro autorregresivo de primer orden en el que la relación entre las tramas viene marcada por los coeficientes de correlación cruzada entre los fotogramas I y la suma de los P para cada GOP (R_{IP}), así como entre la suma de los P y la de los B (R_{PB}).

$$p'_I = F_I^{-1}(u)$$

$$p'_P = \frac{1}{G_P}\{a_{IP} \cdot p'_I + b_P \cdot w\}$$

$$a_{IP} = R_{IP}(0)$$

$$p'_B = \frac{1}{G_B}\{a_{BP} \cdot G_P \cdot p'_I + b_B \cdot w\}$$

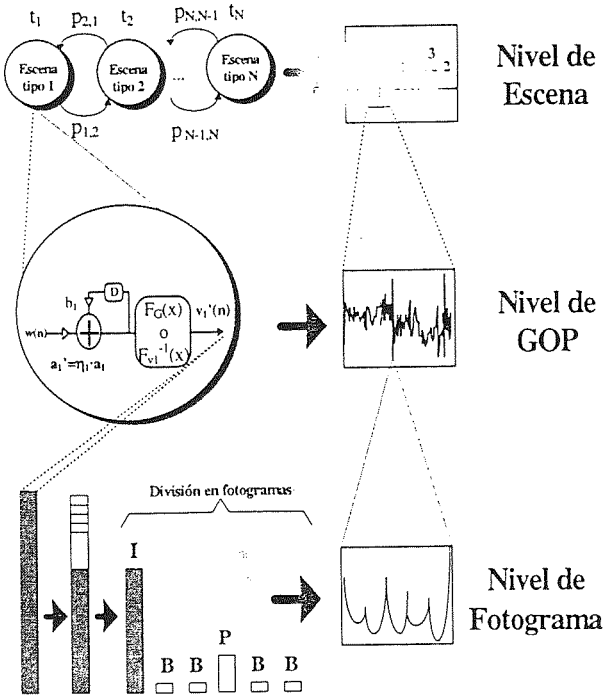
$$a_{BP} = R_{BP}(0)$$

donde w se corresponde con un ruido blanco gaussiano de media 0 y varianza 1, y los coeficientes b_P y b_B son calculados de manera similar a los del modelo PAR explicado en la sección 3.2.

Tras generar el tráfico correspondiente a cada fotograma el ajuste de los pesos se realiza de manera idéntica a la del método anterior.

El modelo global, incluyendo las tres escalas de tiempo, ha sido esquematizado en la figura 4:

Figura 4. Diagrama completo del modelo

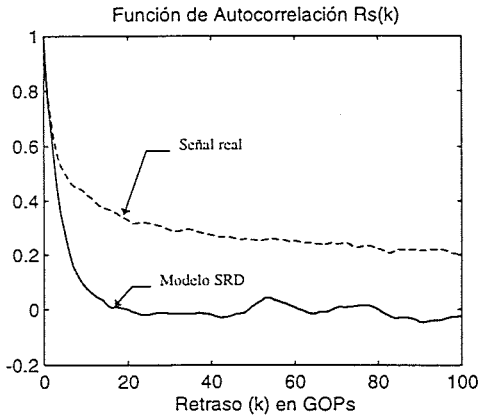


4. Simulación y Resultados

Para probar el esquema propuesto se utilizó la película completa (en torno a dos horas de duración) "La Guerra de las Galaxias" codificada según la norma MPEG-I con una resolución de 240x352 pixels para la luminancia y 120x176 para la crominancia, 24 fotogramas por segundo, y una estructura de GOP de 12 fotogramas en el orden *IBBPBBPBBPBB*. Se eligió dicha película, que cuenta con 14511 GOPs, por la cantidad de modelos existentes en la bibliografía que la utilizan como elemento de prueba [6] [9] [11], ya que presenta fuertes dependencias a largo plazo debido a la presencia de periodos muy acusados de nivel de actividad tanto alto como bajo.

La existencia de las citadas LRD son constatables de diversas maneras, entre las que se pueden citar, por su inmediatez, el análisis de la autocorrelación y de la varianza de la señal agregada. En la figura 5, por ejemplo, se ha representado la correlación de la señal real frente a la de un modelo SRD (que sólo modela dependencias a corto plazo). Se puede observar el decaimiento rápido (exponencial) del modelo frente

Figura 5. Autocorrelaciones de la señal real y de un modelo SRD

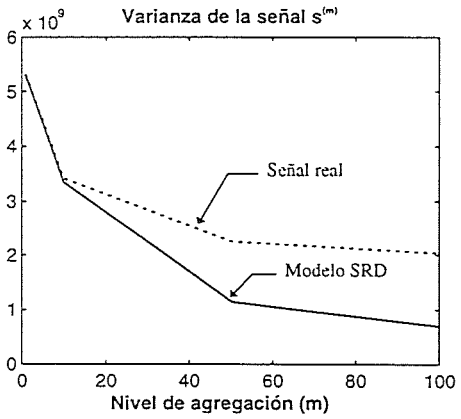


a la lenta evolución (subexponencial) de la autocorrelación de la señal real.

La figura 6, en cambio, muestra la varianza de la señal agregada $s^{(m)}$, definida como sigue, en función del nivel de agregación m.

$$s^{(m)}[i] = \frac{1}{m} \sum_{j=1}^{m} s(m \cdot (i-1) + j)$$

Figura 6. Varianza de la señal agregada



En un primer experimento, para validar la hipótesis de la existencia de diferentes escenas como causa principal de las dependencias a largo plazo, se filtró la señal con una ventana de 1500 GOPs para posteriormente dividirla en tres vectores (N=3) correspondientes a tres estados de actividad distintos. En las figuras 7 y 8 se han representado tanto la función de densidad estadística como la autocorrelación de dichos vectores. De esta última figura se puede observar cómo, al aislar los diferentes niveles de actividad, las LRD desaparecen por cuanto las autocorrelaciones rápidamente decaen hasta hacerse cero. De este modo queda justificado el incluir los cambios de escenas en el modelo global, teniendo en cuenta que estos vectores, que sólo presentan SRD, ya pueden ser caracterizados

Figura 7. Función de densidad estadística de las señales correspondientes a cada tipo de escena

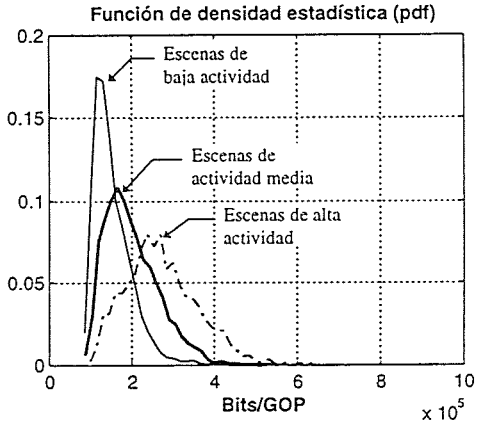
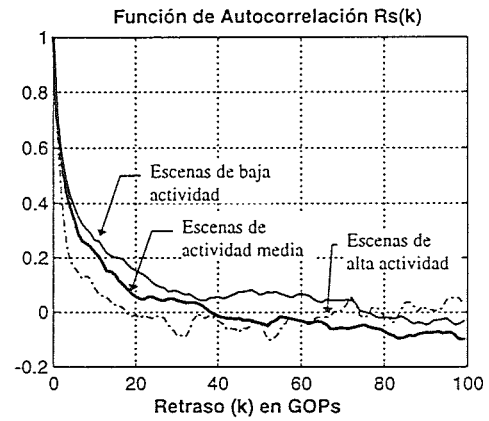


Figura 8. Autocorrelación de la señales correspondientes a cada tipo de escena



por modelos SRD como el PAR modificado propuesto.

Por lo que respecta al modelado en sí del nivel de escena, éste se llevaría a cabo, en este caso en el que se han distinguido tres niveles de actividad, mediante una matriz de conmutación 3 x 3 mientras que el tiempo de permanencia en cada escena se modelaría mediante una distribución exponencial. El ajuste de dicha distribución sobre la distribución real de las duraciones de escena ha sido ejemplificado en la figura 9 para el caso de las escenas de bajo nivel de actividad.

El resultado de establecer un modelo con cambios de escena, desde el punto de vista de la autocorrelación de la señal total (a nivel de GOPs), se presenta en la figura 10, en la que se puede contemplar cómo la autocorrelación del modelo, al incluir cierto modelado de las LRD, decae de una manera lenta y más similar a la de la señal real. Por otro lado y tal y como denota la figura 11, el ajuste de la función de densidad, que realizan los distintos modelos PAR para cada grado de actividad, no se ve afectado por el modelado de los cambios de escena.

Figura 9. Ajuste de la distribución exponencial de la duración de las escenas

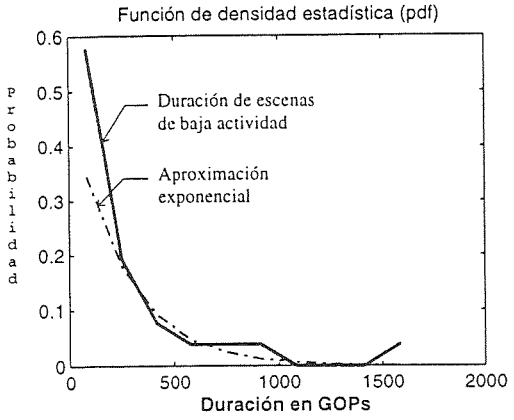


Figura 10. Autocorrelación del modelo y de la señal real

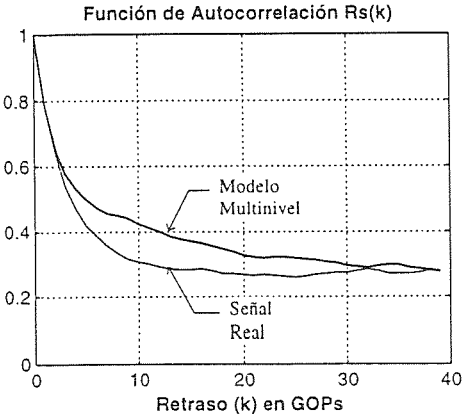
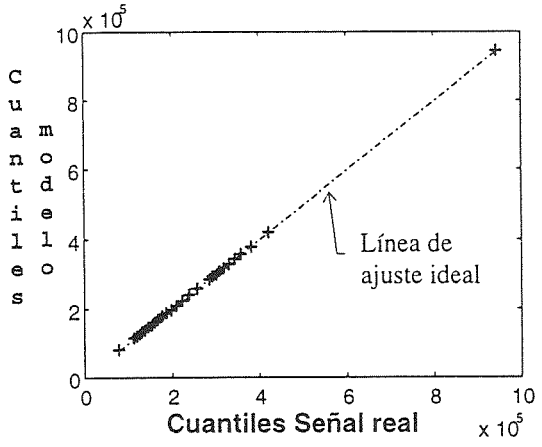


Figura 11 . Representación por cuantiles del ajuste de la distribución de la señal real



Una vez que se ha comprobado que el modelo ajusta razonablemente las características estadísticas de la señal real, hemos de comparar el comportamiento en una cola del modelo y de la señal real, ya que, como ya se dijo, es esta la "prueba de fuego" para validar cualquier modelo que trate de imitar cierto tipo de tráfico. Por ello a partir de ahora analizaremos la aptitud del modelo para imitar la señal real cuando es multiplexado sobre un servidor de tráfico de tasa de servicio fija con cierto buffer o cola a su entrada.

Una de las primeros parámetros a definir dentro del modelo es el tamaño de la ventana con la que se filtrará el tráfico para detectar componentes de baja frecuencia (LRD) en la señal. El tamaño de dicha ventana puede llegar a ser un factor crítico a la hora de representar adecuadamente el comportamiento a largo plazo de la señal [12]. En la figura 12, por ejemplo, se compara el comportamiento a nivel de pérdidas en una cola, para una ocupación del servidor del 47%, del tráfico real y el generado por el modelo, considerando distintos tamaños de ventana y dos tipos de escenas. Se puede apreciar que para tamaños excesivamente pequeños (W=50 y 250 GOPs) de la ventana el modelo no caracteriza bien las LRD de manera que infravalora las pérdidas de la señal real. Por el contrario, valores excesivamente grandes de la misma (W=5000 GOPs) no mejoran substancialmente el comportamiento del modelo además de que restan representatividad a la señal filtrada, en la medida en que ésta es finita y su tamaño empieza a hacerse comparable al de la propia ventana.

Otro aspecto de nuestro modelado que podría evaluarse es la mejora que introduce el reajuste del modelo PAR (lo que hemos llamado PAR modificado) en relación con el comportamiento en una cola. En la figura 13 se comparan , tomando una ventana de 1500 GOPs y dos tipos distintos de escenas, los resultados en la cola del modelo cuando se utilizan tanto procesos PAR como PAR modificado a la hora de representar el tráfico en cada escena. El mejor ajuste de la correlación del modelo PAR modificado redunda en predicciones de las pérdidas más pesimistas y más cercanas a la realidad que las que realiza el modelo PAR.

En cuanto al modelado a nivel de fotograma, en las figuras 14, 15 y 16 se ha representado las pérdidas del modelo para una utilización del 47% del canal cuando se utilizan los

Figura 12. Importancia del tamaño de la ventana

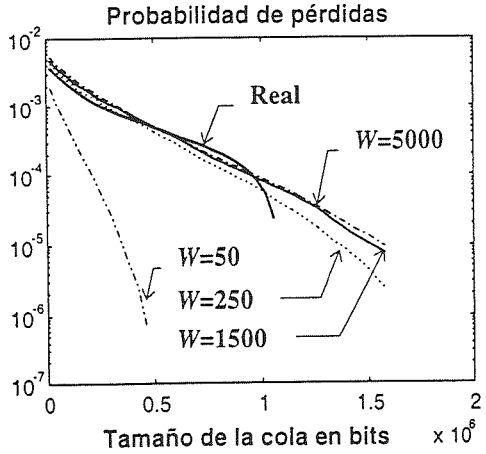
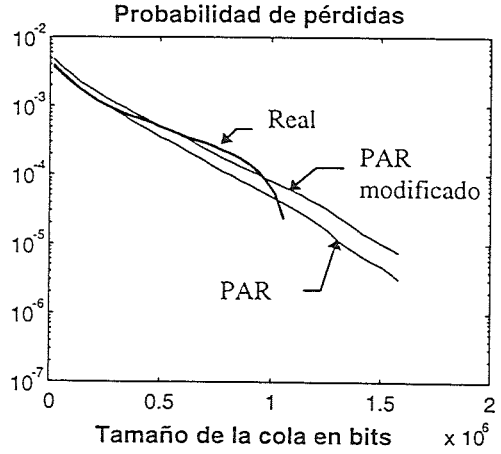


Figura 13. Mejora del modelo PAR modificado



métodos 1, 2 y 3, respectivamente, para dividir el GOP en los distintos fotogramas. De estas figuras se puede colegir que el tercer método introduce cierta mejora frente al primero, al modelar de una manera más flexible la correlación entre los diferentes tipos de fotogramas. El método segundo proporciona siempre un estimador más pesimista puesto que al generar los fotogramas de cada tipo atendiendo únicamente a su propio histograma, presenta mayor posibilidad de que dentro de un GOP haya varias tramas muy densas en posiciones consecutivas, aumentando, en consecuencia, las pérdidas. Esta sobrestimación de las pérdidas se hace, lógicamente, especialmente notable para tamaños de colas más pequeños en tanto que colas mayores amortiguan este carácter más rafagueante de la distribución de los fotogramas, de modo que la curva del método segundo acaba pareciéndose a la de los otros dos. Para dar una idea de la fiabilidad de las simulaciones en estas últimas gráficas, el resultado de la simulación se representó junto con un margen de confianza (en línea discontinua) del 90% según la distribución de t-student.

Por último, para el modelado a nivel de fotograma, también se han incluido (figuras 17 y 18)

Figura 14. Método 1
de distribución de fotogramas.

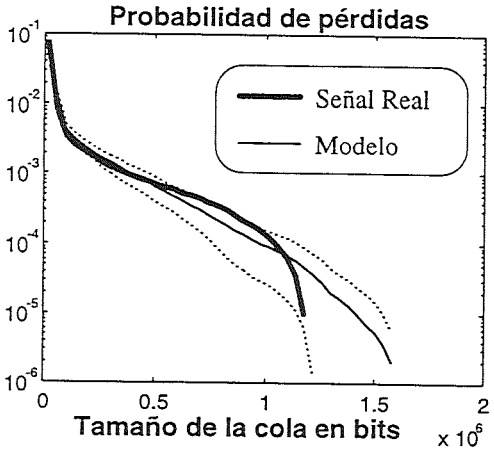


Figura 15. Método 2
de distribución de fotogramas.

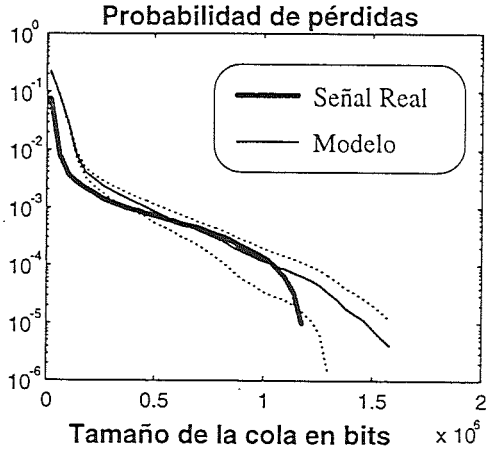
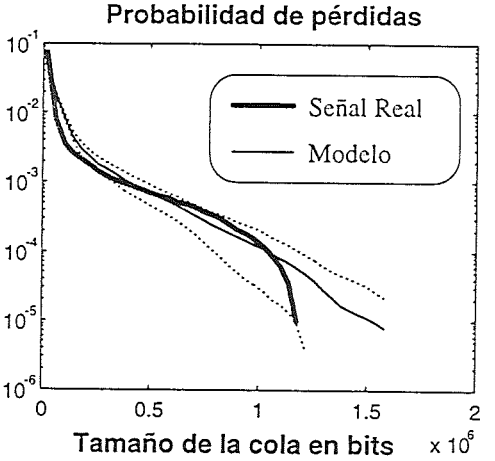


Figura 16. Método 3
de distribución de fotogramas.



gráficas que describen el comportamiento del modelo en cuanto a retardo medio y *jitter* (entendido como la desviación típica del retardo) en función del grado de ocupación del servidor. De dichas gráficas se puede comprobar de nuevo que el método 2 proporciona una estimación pesimista de la realidad mientras que los métodos 1 y 2 ajustan razonablemente bien la señal de vídeo sin que exista una diferencia sustancial entre ellos.

Conclusiones

A lo largo de este trabajo se ha presentado un modelo multinivel que simula las características del tráfico de vídeo con codificación MPEG. El modelo busca tanto representar la naturaleza de la información que se está transmitiendo (películas con cambios de escenas que imponen componentes de baja frecuencia en la señal, grupos de fotograma contiguos que presentan una fuerte correlación) como la conformación propia del tráfico que impone a ciertas escalas de tiempo el propio estándar de codificación (fotogramas con distintas caracterís-

Figura 17. Retardo medio para los tres métodos de división en fotogramas

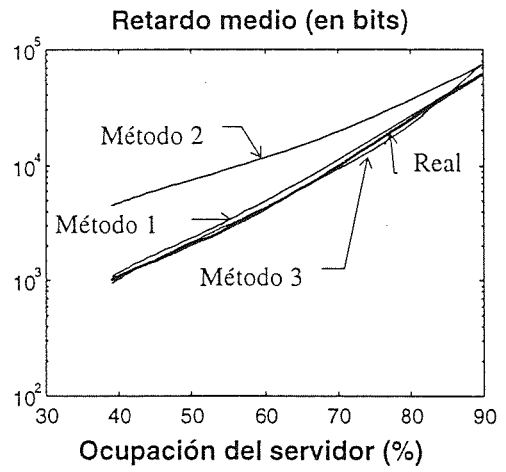
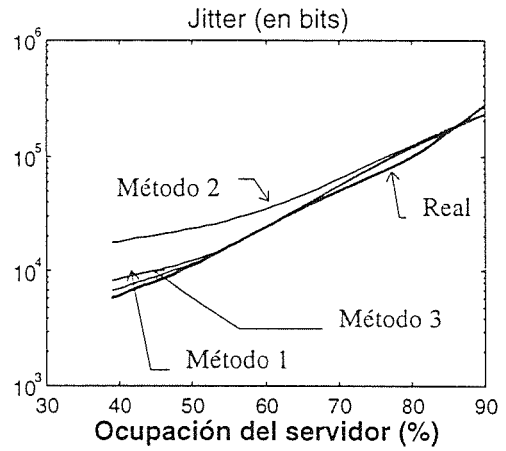


Figura 18. Jitter para los tres métodos de división en fotogramas



ticas). Las dependencias a largo plazo que determina la existencia de escena son modeladas mediante una cadena de Markov, mientras que para el tráfico dentro de cada escena se propone una mejora al proceso PAR o autorregresivo proyectado. En cuanto a la división del tráfico en fotogramas se plantean diversas estrategias que recogen la distribución y las relaciones existentes entre los distintos tipos de fotogramas de la norma MPEG. Finalmente la validez del modelo es comprobada en una cola, tanto a nivel de pérdidas como de retraso medio y jitter.

Agradecimientos

Este trabajo ha sido financiado en parte por la Comisión Interministerial de Ciencia y Tecnología (CICYT), Proyecto N° TIC96-0743. También queremos expresar nuestro agradecimiento a M. Garrett, de Bellcore (USA), por hacer de dominio público el tráfico (bits/fotograma) generado por la codificación MPEG-1 de la película "La Guerra de las Galaxias".

Referencias

[1] Wu, J.L., Chen, Y.W., y Jiang, K.C. "Two Models for Variable Bit Rate MPEG Sources", *IEICE Trans. on Communications*, **E78-B**, 5, 773-745 (1995).

[2] Ohta, N., *Packet Video*, Boston: Artech House (1994).

[3] Melamed, B., y Sengupta, B. "TES Modeling of Video Traffic", *IEICE Trans. on Communications*, **E75-B**, 12, 1292-1300 (1992).

[4] Maglaris B., Anastassious, D., Sen, P., Karlsson, G., y Roberts, J.D. "Performance Models of Statistical Multiplexing In Packet Video Communications", *IEEE Trans. on Communications*, **36**, 7, 834-843 (1988).

[5] Grünenfelder, R., Cosmas, J.P., Manthorpe, S., y Odinma-Okafor, A. "Characterization of Video Codecs as Autorregresive Moving Average Processes and Related Queueing System Performance", *IEEE Journal on Selected Areas in Communications*, **14**, 7, 284-293 (1996).

[6] Conti, M., Gregori, E., y Larson, A. "Study of the Impact of MPEG-1 Correlations on Video-Sources Statistical Multiplexing", *IEEE Journal on Selected Areas in Communications*, **14**, 7, 1455-1471, (1996).

[7] Beran, J., Sherman, R., Taqqu, M.S., y Willinger, W. "Long-Range Dependence in Variable-Bit-Rate Video Traffic", *IEEE Trans. on Communications*, **14**, 2/3/4, 1566-1579 (1996).

[8] Jelenkovic, P.R., Lazar, A.A., y Semret, N. "The effect of Multiple Time Scales and Subexponentiality in MPEG Video Streams on Queueing Behavior", aceptada para el *IEEE Journal on Selected Areas in Communications*, **15**, (1997).

[9] Rose, O. "Statistical properties of MPEG video traffic and their impact on traffic modeling in ATM systems", Report No. 101, University of Wuerzburg, Institute of Computer Science Research Report Series, (1995).

[10] Krunz, M., y Tripathi, S.K. "Modeling Bit Rate Variations in MPEG Sources", Internal Report, Department of Computer Science, University of Maryland at College Park (1996).

[11] Garret, M.W., y Willinger, W. "Analysis, Modeling and Generation of Self-Similar VBR Video Traffic", Proc. ACM Sigcomm'94, London, 269-280 (1994).

[12] Conti, M., Gregori, E., y Larson, A. "Validation and Tuning of an MPEG-1 Video Model", en *ATM Networks: Performance, Modelling and Evaluation*, 2, Ed. D. Kouvatsos, Chapman & Hall (1996).

Videoconferencia sobre LAN

SANTIAGO SÁNCHEZ PALOMINO (santiago.sanchez@svrit.dgtt.cf.jcyl.es)
MIGUEL ANGEL REGIDOR GONZALEZ (migreg@tel.uva.es)
RAFAEL MOMPÓ GÓMEZ (jyr@dvnet.es)
DEPARTAMENTO DE TEORÍA DE LA SEÑAL Y COMUNICACIONES E INGENIERÍA TELEMÁTICA
E.T.S. Ingenieros de Telecomunicación. UNIVERSIDAD DE VALLADOLID
REAL DE BURGOS S.N. VALLADOLID 47011

Abstract:

This is an overview of ITU-T Recommendation H.323, "Visual Telephone Systems and Equipment for Local Area Networks (LAN) which Provide a Non-Guaranteed Quality of Service", which covers the technical requirements for narrow-band visual telephone services in those situations where the transmission path includes one or more packet-switched networks providing a non-guaranteed Quality of Service (QoS).

1. Introducción

El uso de sistemas de videoconferencia sobre medios telefónicos en entornos empresariales es cada día más habitual, sobre todo a nivel internacional o entre ciudades bastante distantes. Esto se debe principalmente a la rentabilidad que se consigue al usar estos sistemas de comunicación para mantener reuniones o cursos, en vez de tener que desplazarse a la ciudad donde tenían lugar, como ocurría anteriormente.

El abaratamiento y desarrollo de estos sistemas de comunicación sobre líneas telefónicas, tanto analógicas como digitales, fue debido sobre todo a la existencia de una serie de estándares, la serie H.32x, que establece todos los requisitos necesarios para asegurar el correcto funcionamiento y la interoperabilidad de sistemas provenientes de distintos fabricantes.

Actualmente, las mismas empresas que ya trabajan con los sistemas de videoconferencia sobre líneas telefónicas, demandan sistemas que puedan funcionar sobre las redes de área local o redes corporativas que ya tiene instaladas. Esta necesidad, justifica el desarrollo de normativa que garantice el correcto funcionamiento de estos sistemas entre sí, y la interoperabilidad entre los ya desarrollados anteriormente para otro tipo de medios.

2. Normativa existente

Actualmente, la familia de normas ITU-T dedicadas a los sistemas de videoconferencia e la H.32x, que comprende los sistemas sobre RDSI Banda Estrecha, RDSI Banda Ancha, telefonía analógica (RTC) y redes locales con garantía de calidad de servicio (Iso Ethernet) o sin ella.

Este grupo normativo, se puede observar en la tabla1, donde además se han incluido todos los estándares que a su vez tiene que cumplirse por las diferentes partes de los sistemas de vídeo, audio, datos, multiplexación, control y señalización en cada una de las normas principales.

Tabla 1. Normativa existente sobre videoconferencia

	H.320	**H.321**	**H.322**	**H.323**	**H.324**
Red	RDSI Banda Estrecha	RDSI Banda Ancha (ATM)	Iso Ethernet	LAN	RTC
Vídeo	H.261	H.261	H.261	H.261 H.263	H.261 H.263
Audio	G.711 G.722 G.728	G.711 G.722 G.728	G.711 G.722 G.728	G.711 G.722 G.728 G.723.1 G.729	G.723.1 G.729
Datos	T.120	T.120	T.120	T.120	T.120
Multiplexación	H.221	H.221	H.221	H.225.0	H.223
Control	H.242	H.242	H.242	H.245	H.245
Señalización	Q.931	Q.931	Q.931	H.225.0 (Q.931)	

El H.261 es un sistema de codificación de vídeo para velocidades de transmisión de p x 64 kbit/s con 'p' entre 1 y 30 (desde un canal 'B' RDSI hasta un acceso primario RDSI). Básicamente, su funcionamiento consiste en dividir la imagen de entrada en varios niveles, donde el nivel inferior está compuesto por pequeños bloques a los cuales aplica un algoritmo predictivo, una codificación por transformada discreta de coseno y una cuantificación variable. Posteriormente se aplica a todo ello una codificación Huffman optimizada, consiguiendo con todo este proceso disminuir drásticamente la cantidad de información a transmitir y con ello la tasas binarias.

Los formatos de la imagen de entrada permitidos son: el CIF (Formato Intermedio Común), que está compuesto por una matriz de luminancia de 288 líneas y 352 pixels por línea, y dos matrices de crominancia compuesta de 144 líneas y 176 pixels por línea, y el QCIF (Quarter CIF), que está formado por la mitad de líneas y de pixeles tanto de luminancia como de crominancia del formato CIF.

La recomendación H.263, es básicamente una versión mejorada de la H.261, donde se realiza una estimación de movimiento de ½ pixel, y donde se amplia el formato de imagen de entrada a Sub-QCIF, QCIF, CIF, 4CIF y 16 CIF. La tabla 2 resume los formatos de imagen soportados por estos estándares.

Los terminales H.323 pueden trabajar de forma asimétrica en lo que a vídeo se refiere. Es decir, son capaces de transmitir un formato de vídeo de los anteriores, y de recibir otro diferente.

Tabla 2. Formato de imágenes soportadas

Formato	Tamaño	H.261	H.263
Sub-QCIF	128x96	no soportado	requerido
QCIF	176x144	requerido	requerido
CIF	352x288	opcional	opcional
4CIF	702x576	no soportado	opcional
16CIF	1408x1152	no soportado	opcional

3.1.2 Audio codec

El codec de audio es un elemento obligatorio en los terminales, teniendo que ser compatibles todos ellos con al menos la recomendación G.711[1]. Así mismo, pueden ser opcionalmente capaces de codificar o decodificar voz usando las recomendaciones G.722[2], G.728[4], G.729[5] y G.723.1[3].

La recomendación G.711, consiste básicamente en una codificación por pulsos (PCM) de la señal de hasta 4 kHz, típicamente de voz, a la que se realiza una cuantificación no uniforme del tipo μ-law o A-law, consiguiendo una tasa binaria de 64 kbit/s.

La recomendación G.722, consigue mejor calidad que la anterior, ya que el canal se divide en dos sub-bandas a las que se realiza una codificación ADPCM (Adaptive Differential Pulse Code Modulation) independientemente. Con esto se consigue trabajar con señales de entrada de hasta 7 kHz con 48, 56 ó 64 kbit/s.

La recomendación G.728, trabaja con señales de hasta 3.4 kHz a las que aplica una cuantificación LD-CELP (low delay code excited linear prediction), consiguiendo una tasa binaria de 16 kbit/s a la salida.

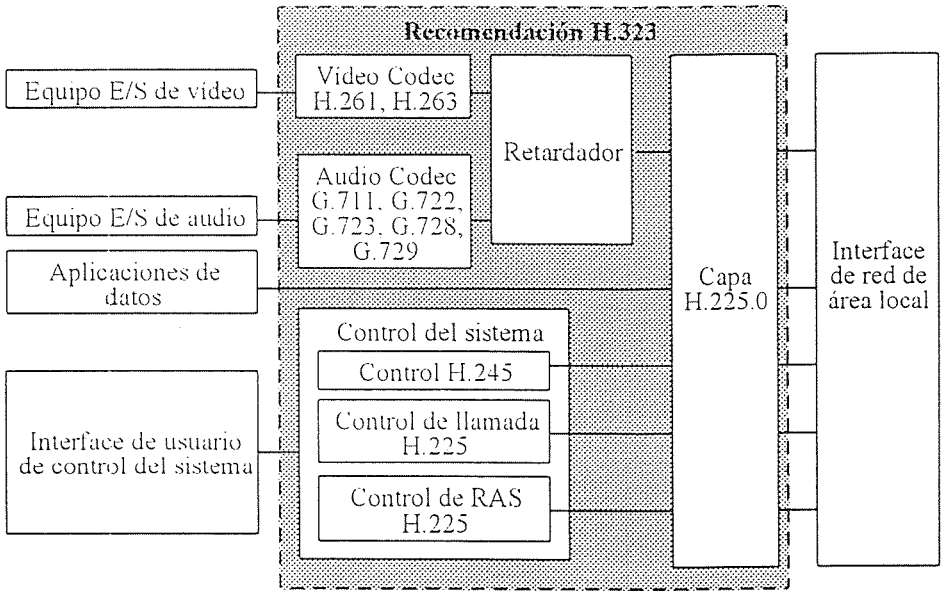


Figura 2. Diagrama de bloques del terminal H.323

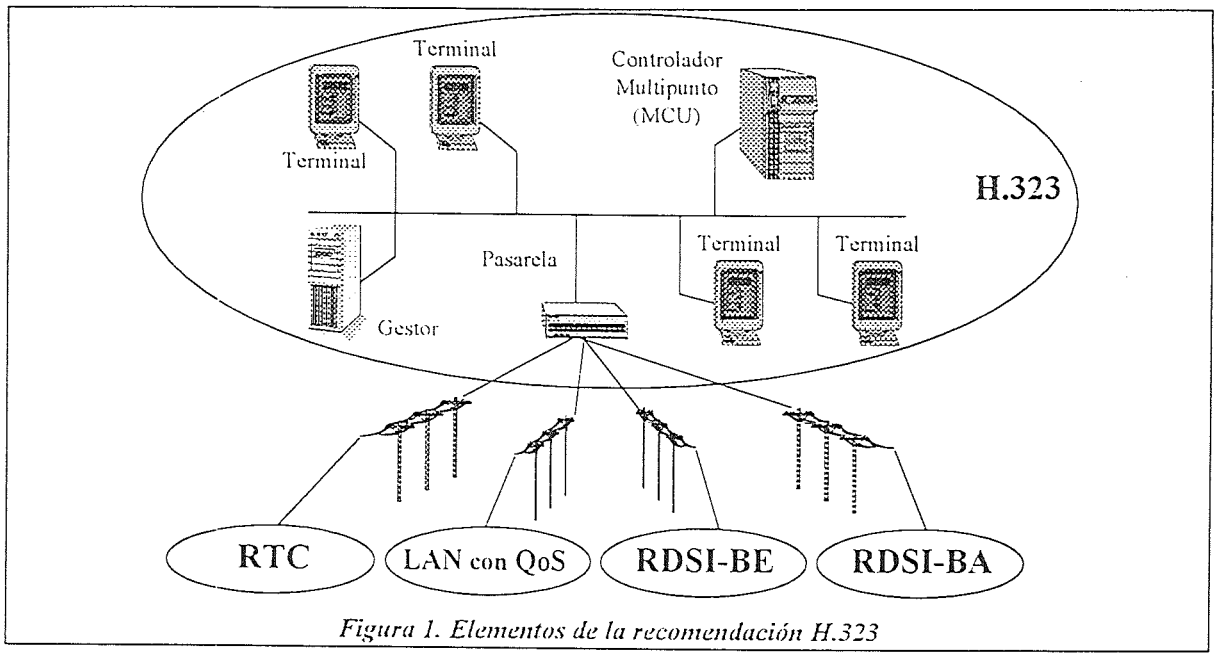


Figura 1. Elementos de la recomendación H.323

3. Recomendación ITU-T H.323

La recomendación H.323[15] cubre los requerimientos técnicos para la prestación de servicios audiovisuales en redes de área local (LAN) sin garantía de calidad de servicio (QoS). Dichos servicios pueden ser provistos a varios niveles: sólo voz, voz y vídeo, voz y datos, o los tres simultáneamente (voz, vídeo y datos).

Ejemplos de redes locales que no garantizan la calidad de servicio son las siguientes:

- Ethernet (IEEE 802.3).
- Fast Ethernet (IEEE 802.10).
- FDDI (modo sin garantía de calidad de servicio).
- Token Ring (IEEE 802.5).

Hay que destacar la interoperabilidad conseguida en los terminales H.323 que permiten la comunicación, además de con sus homólogos, con una amplia variedad de terminales ya existentes, como son: terminales H.321[13] y H.310[11] en RDSI de banda ancha, terminales H.320[12] en RDSI de banda estrecha, terminales H.322[14] en redes de área local con garantía de calidad de servicio y terminales H.324[16] y V.70 de la red telefónica conmutada (RTC). Estos elementos se muestran en la fig. 1.

Los componentes contemplados en el estándar H.323 son: terminales, pasarelas, gestores y unidades de control multipunto (MCU), como puede verse en la fig. 1. Así mismo, en la figura 3 puede observarse la pila de los protocolos utilizados.

3.1 Terminales

Los terminales H.323, proporcionan comunicación bidireccional de audio, vídeo y datos entre los extremos de la comunicación.

En la fig. 2 puede observarse el diagrama de bloques de los terminales H.323, formados por un codificador / decodificador (Codec) de vídeo (opcional), un codec de audio, una unidad de control del sistema, una capa H.225.0[6] y un interface a la red de área local.

Quedan fuera de las especificaciones de este estándar todos los equipos periféricos como son: los equipos de vídeo (cámaras y monitores), equipos de audio (micrófonos, altavoces, mezcladores, etc.), aplicaciones de datos y el interface de usuario de control del sistema.

El caso del interface con la LAN, es especial, ya que aunque queda fuera de las especificaciones de este estándar, debe proveer los servicios descritos en la recomendación H.225.0.

3.1.1 Vídeo codec

El vídeo codec es un elemento opcional en los terminales. Es decir, pueden existir terminales únicamente de audio con posibilidades de comunicación de datos, pero sin posibilidad de visualizar imágenes.

En el caso de que exista, este elemento deberá ser acorde con el estándar H.261[8], para posibilitar la comunicación de vídeo con cualquier terminal de los indicados en la tabla 1, y opcionalmente podrá también ser compatible con el H.263[10].

Los terminales pueden opcionalmente, soportar varios canales de audio al mismo tiempo (para traducción simultánea, multiconferencia, etc.), dependiendo principalmente de su capacidad de proceso y de la capacidad y del retardo del canal.

3.1.3 Retardador

Este elemento se encarga de sincronizar las tramas de audio y vídeo en el receptor, añadiendo un retardo extra á las que primero lleguen. Este bloque sólo existe en recepción, ya que la transmisión se realiza a tiempo real sin retrasos intermedios.

3.1.4 Canal de datos

El terminal H.323 puede soportar uno o varios canales de datos unidireccionales o bidireccionales opcionalmente.

El estándar adoptado para el intercambio de datos entre terminales H.32x es el T.120[17].

3.1.5 Control H.245

Para implementar las funciones de control, se establece un canal de control entre los puntos finales de la comunicación, según la recomendación H.245[7].

Este canal de control, dispone de cuatro tipos de mensajes con los que se desarrolla todo el diálogo de control: *Peticiones, Respuestas, Comandos e Indicaciones*. Los mensajes de *petición*, requieren una acción específica por el receptor, incluyendo una respuesta inmediata, siendo por tanto los mensajes de *respuesta* para responder a mensajes de *petición*. Los mensajes de *comandos* requieren una acción específica, sin necesidad de respuesta, y los mensajes de *indicación* son meramente informativos, no precisando acción ni respuesta.

Por medio de estos mensajes, los terminales realizan principalmente las siguientes tareas:

- Negociación de las capacidades de intercambio, indicando los dos extremos qué tipo de formatos de vídeo, audio y datos admiten, para acordar posteriormente el que van a utilizar.

- Señalizaciones de los canales lógicos: la información de vídeo, audio, datos y control, se transmite por medio de diferentes canales lógicos, los cuales son numerados de forma única e independiente en cada sentido de la transmisión. Es tarea del canal de control, el indicar la apertura o cierre de estos canales, así como el asociar aquellos canales que sean bidireccionales.

- Asignación del maestro y del esclavo de la comunicación. Este es un sistema para resolver casos de conflicto, considerando siempre que predomina el maestro.

3.1.6 Control RAS

Este es un canal especial que se establece entre los terminales y el gestor de videoconferencia, únicamente en el caso de que exista este último. En dicho supuesto, se utilizan mensajes H.225.0 para realizar registros, admisiones, cambios de ancho de banda, estado y procedimientos de desconexión entre los terminales y el gestor.

3.1.7 Control de llamada

Las funciones de control de llamada, usan un canal independiente del canal RAS y del de control H.245, el canal de señalización de llamada, el cual establece una conexión entre dos extremos H.323 usando los mensajes de control de llamada H.225.0.

<table>
<tr><td>Aplicac. de audio</td><td>Aplicac. de vídeo</td><td colspan="4">Control y gestión del terminal</td><td>Aplicac. de datos</td></tr>
<tr><td rowspan="2">G.711
G.722
G.723
G.728
G.729</td><td rowspan="2">H.261
H.263</td><td rowspan="3">RTCP</td><td rowspan="3">H.225.0
Canal
RAS</td><td>H.225.0
Canal de
Señalización
de llamada</td><td>H.245
Canal de
Control</td><td rowspan="2">T.124</td></tr>
<tr><td colspan="2">RTP</td><td colspan="2">X.224 Clase 0</td><td>T.125</td></tr>
<tr><td colspan="4">Transporte no fiable (UDP)</td><td colspan="2">Transporte fiable (TCP)</td><td rowspan="3">T.123</td></tr>
<tr><td colspan="6">Nivel de red (IP)</td></tr>
<tr><td colspan="6">Nivel de enlace (IEEE 802.3)</td></tr>
<tr><td colspan="7">Nivel física (IEEE 802.3)</td></tr>
</table>

Figura 3. Pila de protocolos

3.1.8 Capa H.225.0

Esta capa se encarga de formatear todos los canales lógicos de vídeo, audio, datos e información de control que se establecen según los procedimientos de la recomendación H.245.

El envío de audio y vídeo se realiza usando RPT independientes por medio de canales no fiables para minimizar el retardo y poder conseguir diferentes calidades en servicios diferentes.

3.2 Pasarela

La pasarela es el elemento encargado de realizar la traducción de los formatos utilizados por los terminales de este estándar, a los demás estándares posibles del resto de terminales posibles, es decir, de H.323 a:

- Terminales H.320 en RDSI Banda Estrecha.
- Terminales H.324 en RTC.
- Terminales H.322 en LAN con calidad de servicio garantizado.
- Terminales H.321 y H.310 en RDSI Banda Ancha.

Esta transformación se realiza a nivel de señalización de llamada, mensajes del canal de control, técnicas de multiplexación y codificación de audio y vídeo. Para ello, la pasarela tiene las características de un terminal H.323 en el lado conectado a la LAN, y del tipo de terminal concreto conectado en el otro extremo (de los enumerados anteriormente).

La pasarela es un elemento no necesario en el caso de realizar comunicaciones directas entre terminales H.323 situados en la misma red, pero es totalmente necesario en el caso de querer realizar comunicaciones con terminales de otro tipo de redes.

3.3 Gestor

El gestor de videoconferencia es un elemento opcional dentro de la recomendación. Su función consiste en dotar de servicios de control de llamadas a los extremos finales de la comunicación. Para ello, los terminales piden permiso al gesto para establecer una videoconferencia, siendo éste el encargado de decidir si lo concede (en función de diversos parámetros), y de reservar el ancho de banda necesario para la comunicación en caso afirmativo.

Cuando existe, proporciona como mínimo los siguientes servicios:

- Traducción de direcciones de 'alias' a direcciones de transporte.
- Control de admisión, gestionando el acceso de los terminales en función de diversos criterios como por ejemplo, autorización, ancho de banda disponible, etc.
- Control de ancho de banda, verificando en todo momento el disponible en el canal y gestionando el necesario para las comunicaciones.

Además, puede proporcionar otras funciones opcionales, como:

- Autorización de llamada, rechazando llamadas de determinados terminales o durante ciertos periodos de tiempo.
- Gestión de llamada, manteniendo listas de las comunicaciones en curso, permitiendo interrumpirlas en caso de necesidad.
- Funciones opcionales que se posponen para mayor estudio dentro del estándar: señalización de control de llamada, la estructura de datos de información de gestión del gestor, la reserva de ancho de banda para terminales que no soportan esta función y los servicios de directorio.

3.4 Unidad de Control Multipunto

La Unidad de Control Multipunto (MCU) es la encargada de establecer la interconexión de más de dos terminales por medio de una multiconferencia. Esta unidad está formada por un

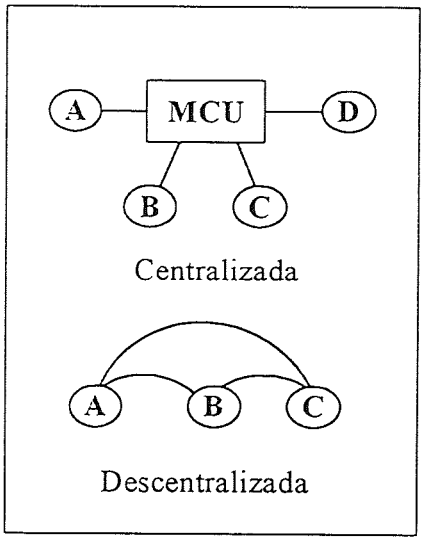


Figura 4. Configuraciones multipunto

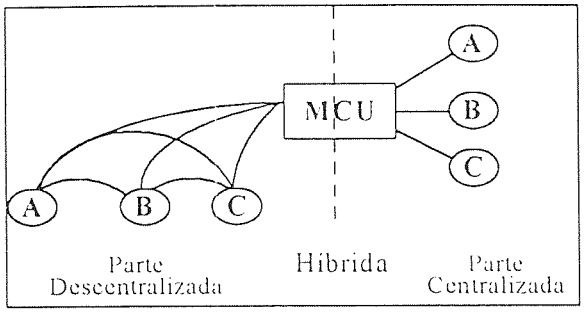


Figura 5. Configuración multipunto híbrida

Controlador Multipunto (MC) y cero o más Procesadores Multipunto (MP). La función de control multipunto podría ser distribuida entre varias MCU pero esta configuración se deja para posteriores estudios dentro de la recomendación.

El MC proporciona las funciones de control para soportar la comunicación entre tres o más terminales en una multiconferencia, encargándose entre otras cosas de determinar el Modo de Comunicación Seleccionada (SCM) para la comunicación en función de las capacidades de transmisión de cada uno de los integrantes de la misma. El MC puede estar localizado en la pasarela, en el gestor y en los terminales, además de en la MCU.

El MP recibe los canales de vídeo, audio o datos de los terminales, los procesa y los devuelve otra vez a los terminales. Estos procesadores pueden ser genéricos o específicos, encargándose en este último caso de procesar únicamente el vídeo, el audio o los datos. En el caso de los MP de vídeo, estos envían de vuelta a los terminales la imagen del terminal que esté transmitiendo voz en ese momento, o una imagen conteniendo a todos los participantes de la videoconferencia en una disposición de cuadrícula. De igual forma, los MP de audio pueden seleccionar la voz principal, o realizar una mezcla de todas las voces que se encuentren transmitiendo en ese momento. El MP puede estar localizado en la pasarela y en el gestor además de en la MCU.

Las posibilidades contempladas en el estándar, con que se puede llevar a cabo la videoconferencia multipunto son: de forma centralizada, de forma descentralizada y de forma híbrida.

3.4.1 Multipunto centralizada

Este modo de operación es obligatorio para todos los terminales. En él, los terminales comunican punto a punto el canal de control con el MC de la MCU, y los canales de audio, vídeo y datos con el MP de la MCU como puede verse en la fig. 4.

El MC se encarga de las funciones de control multipunto H.245, y el MP realiza la conmutación o mezcla de vídeo, la mezcla de audio y la multidistribución de datos T.120, transmitiendo los canales resultantes de vuelta a los terminales.

3.4.2 Multipunto descentralizada

En el caso de que los terminales dispongan de esta capacidad opcional, éstos comunican de forma punto a punto el canal de control H.245 con un MC del sistema (presente en una MCU, en una pasarela, en un gestor o en un terminal) y opcionalmente, los canales de datos con un MP (presente en una MCU, en una pasarela o en un gestor) como puede verse en la fig. 4.

En este caso, los terminales transmiten sus canales de vídeo y audio en multidistribución a los demás. Los terminales reciben todos los canales de vídeo multidistribuidos, mostrando uno o varios al usuario, y reciben todos los canales de audio multidistribuidos, que son mezclados a la salida.

3.4.3 Multipunto híbrida

En el caso de que los terminales dispongan de esta capacidad opcional, éstos comunican de forma punto a punto el canal de control H.245 con el MC de la MCU, y opcionalmente los canales de datos con un MP. Esta configuración se muestra en la fig. 5.

En este caso, se plantean dos posibles configuraciones:

- Multipunto híbrida con audio centralizado: se realiza multidistribución de vídeo y centralización de audio en el MP de la MCU. El MP realiza la mezcla de todos los canales de audio enviando el resultado a los terminales.

- Multipunto híbrida con vídeo centralizado: se realiza multidistribución de audio y centralización de vídeo en el MP de la MCU. El MP realiza la conmutación o mezcla de los canales de vídeo, enviando el resultado a los terminales.

4. Conclusiones

El cumplimiento de la recomendación H.323, asegura la interoperabilidad de los terminales para LAN de diversos fabricantes, y de todos ellos con el resto de terminales disponibles para otros medios de transmisión.

Así mismo, el desarrollo de este estándar impulsa los sistemas de videoconferencia sobre

redes sin garantía de calidad de servicio, anticipando el futuro de este tipo de comunicaciones tanto sobre LAN como sobre la red Internet en cuanto ésta aumente su ancho de banda disponible, o la tecnología evolucione lo suficiente, para aumentar los grados de compresión de la información audiovisual.

Bibliografía:

[1] Recomendación ITU-T G.711, "Pulse code modulation (PCM) of voice frequencies", 1988.

[2] Recomendación ITU-T G.722, "7 kHz audio-coding within 64 kbit/s", 1988.

[3] Recomendación ITU-T G.723.1, "Dual rate speech coder for multimedia communication transmitting at 5.3 & 6.3 kbit/s", 1996.

[4] Recomendación ITU-T G.728, "Coding of speech at 16 kbit/s using low-delay code excited linear prediction", 1992.

[5] Recomendación ITU-T G.729, "Speech codec for multimedia telecommunications transmitting at 8/13 kbit/s", 1995.

[6] Recomendación ITU-T H.225.0, "Media stream packetization and synchronization on non-guaranteed quality of service LANS", 1996.

[7] Recomendación ITU-T H.245, "Control protocol for multimedia communication", 1996.

[8] Recomendación ITU-T H.261, "Video codec for audio-visual services at px64 kbit/s", 1993.

[9] Recomendación ITU-T H.262, "Generic coding of moving pictures and associated audio: video", 1995.

[10] Recomendación ITU-T H.263, "Video coding for low bitrate communication", 1996.

[11] Recomendación ITU-T H.310, "Broadband audio-visual communications systems and terminal equipment", 1996.

[12] Recomendación ITU-T H.320, "Narrow-band ISDN visual telephone systems and terminal equipment", 1995.

[13] Recomendación ITU-T H.321, "Adaptation of H.320 visual telephone terminals to B-ISDN environments", 1995.

[14] Recomendación ITU-T H.322, "Visual telephone systems and terminal equipment for local area networks which provide a guaranteed quality of service", 1995.

[15] Recomendación ITU-T H.323, "Visual telephone systems and equipment for local area networks which provide a non-guaranteed quality of service", 1996.

[16] Recomendación ITU-T H.324, "Terminal for low bitrate multimedia communication", 1995.

[17] Recomendación ITU-T T.120, "Transmission protocols for multimedia data - under development".

Conformación Predictiva de Tráfico de Vídeo VBR MPEG a partir de su Caracterización como Proceso ARIMA*

Luis J. de la Cruz, Juan J. Alins, Esteve Pallarès, Marcos González y Jorge Mata
DEPARTAMENTO DE MATEMÁTICA APLICADA Y TELEMÁTICA
UNIVERSIDAD POLITÉCNICA DE CATALUÑA
C/ JORDI GIRONA, 1. 08034 BARCELONA
Correo electrónico: ljcruz@mat.upc.es / jmata@mat.upc.es

Abstract:

The use of smoothing techniques to remove the periodic fluctuations of the bit rate generated by the codification modes of the MPEG algorithm is very suitable in video transmission. In this way, the multiplexing gain is maximized and the resource allocation is reduced in ATM Networks. The traffic smoothing can be achieved storing the cells in a buffer. This buffer is allocated between the coder and the user-interface. To reduce the delay introduced in the storage process a new technique to forecast the VBR MPEG traffic is presented. This technique is based on the characterization of bits per frame generated by the MPEG coder as an ARIMA process. In this study the invariance of the ARIMA coefficients is verified for all coded sequences. In addition, these coefficients are invariant also in front of the changes of the selected image quality in the coder. This characterization allows to propose a new traffic shaper scheme when forecast techniques are applied.

1. Introducción

Los algoritmos utilizados en la actualidad para la compresión de tráfico de vídeo, como por ejemplo el MPEG [1], provocan fluctuaciones periódicas de la tasa generada a la salida del codificador. Estas fluctuaciones se producen como resultado de los distintos modos de codificación empleados por el algoritmo en los cuadros de la secuencia de vídeo. Al transmitir sobre redes ATM [2], la ubicación de recursos se puede ver afectada negativamente, debido a que la periodicidad comentada provoca un descenso en la ganancia de multiplexación estadística. Para evitar estos efectos, se recurre al uso de técnicas de suavizado. La más simple de dichas técnicas consiste en el almacenamiento de la información a transmitir en un buffer durante un intervalo, enviándola posteriormente a la red a la tasa media obtenida durante dicho intervalo. Sin embargo, estas técnicas de almacenamiento añaden un retardo a la transmisión que no es admisible en el caso de servicios interactivos. En este artículo se propone y analiza un nuevo método que reduce el retardo introducido por el almacenamiento mediante el uso de técnicas de predicción de tasa binaria. Dicha predicción se lleva a cabo en base a la caracterización del tráfico de vídeo generado por un codificador MPEG mediante un modelo autoregresivo e integrativo de media móvil (ARIMA). La identificación del modelo ARIMA se ha llevado a cabo empleando tres secuencias patrón de larga duración (34000, 51000 y 174000 cuadros) codificadas en MPEG-I. Como resultado se ha observado la invarianza temporal de los coeficientes del predictor, junto con su insensibilidad a los cambios en la calidad de imagen seleccionada en el codificador. A partir de estos resultados se propone un nuevo esquema de conformación de tráfico y se analizan las ventajas e inconvenientes respecto a los clásicos sistemas de almacenamiento.

El resto del artículo está organizado como sigue. En el apartado siguiente se revisan las características principales del tráfico MPEG VBR. A continuación se introducen los procesos ARIMA para caracterizar la tasa binaria generada por cuadro por este tipo de codificación. Posteriormente, en el apartado 4, se hace uso de dicha caracterización para realizar un predictor de tasa, cuyo correcto funcionamiento es comprobado para las tres secuencias bajo estudio. En el apartado 5, se aborda el problema de la conformación del tráfico comentado, comparando los clásicos sistemas de almacenamiento con un nuevo esquema basado en el predictor. Como se verá, la principal aportación del nuevo esquema es la reducción del retardo introducido por la conformación, dado que no se basa en el almacenamiento de muestras pasadas sino en la predicción de muestras futuras. Finalmente, se exponen las conclusiones y principales aportaciones derivadas de este trabajo.

2. Tráfico MPEG VBR

Dentro de la variedad de técnicas de codificación, el algoritmo de codificación MPEG (Motion Picture Expert Group) se ha revelado como el más adecuado para la transmisión y almacenamiento de secuencias de vídeo [3]. El algoritmo de codificación MPEG, para vídeo digital, emplea conjuntamente la técnica de compresión transformada, a través de la transformada coseno discreta (DCT), y la técnica de compensación de movimiento. Este análisis se ha llevado a cabo a través de estudios estadísticos de secuencias de vídeo de larga duración. El algoritmo de codificación de vídeo MPEG inicialmente fue desarrollado para aplicaciones de almacenamiento y recuperación de vídeo comprimido a tasas binarias reducidas, del orden de 1'5 Mbps. Sin embargo, el

*Este trabajo está soportado por el Proyecto SIGLA [CICYT, proyecto TEL 96-1452], dentro del Plan Nacional de I+D

algoritmo de codificación se ha mostrado muy adecuado para la transmisión de vídeo sobre redes de comunicaciones, dado que reduce sustancialmente la tasa binaria en transmisión y, por tanto, los recursos necesarios de la red. Dentro de las aplicaciones más comunes donde es empleado este mecanismo de codificación caben destacar: videocorreo electrónico, videotelefonía, videoconferencia, vídeo juegos, documentos multimedia, distribución de vídeo con calidad VCR (Video Casette Recorder), vídeo bajo petición y televisión digital de alta definición.

El algoritmo MPEG se adecua perfectamente para transferir vídeo comprimido sobre redes locales, metropolitanas y de área extensa. Esta adaptación se obtiene configurando el modo de funcionamiento de la codificación de forma que, dependiendo del tipo de red, se puede obtener una tasa binaria constante o variable [4] [5]. La diferencia entre la transmisión con tasa binaria constante o variable se manifiesta en que la calidad de la secuencia varía dependiendo de la complejidad y actividad de las imágenes o se mantiene constante. La estandarización del algoritmo de codificación MPEG-I, para el almacenamiento y distribución de vídeo comprimido en el rango de 1 a 2 Mbps con calidad de VCR se realizó en 1991 [6]. Posteriormente, se ampliaron las aplicaciones del algoritmo de codificación, de forma que se puede emplear para la distribución de vídeo digital con resolución de televisión y televisión de alta definición. A su vez, se incorporaron un conjunto de mejoras para permitir la compatibilidad en la presentación para diferentes resoluciones, disminuir los efectos de las pérdidas en transmisión y permitir la codificación de señales entrelazadas. Esta nueva versión MPEG-II, admite tasas que pueden ir desde 1 a 83 Mbps [7].

En la recomendación MPEG-I, se aconseja como formato de entrada el SIF ya que tanto para señal proveniente de PAL o NTSC se obtienen tasas binarias entre 1 y 2 Mbps. La secuencia SIF se estructura, como se muestra en la Fig.1, en cuatro niveles de codificación: cuadro, tira o *slice*, macrobloque y bloque. El cuadro es la unidad básica de presentación cuyo número de *pels* (pixels de 8 bits) depende de la resolución. La imagen se estructura en zonas o bloques de 8 x 8 pels donde se aplica la DCT. Los coeficientes proporcionados son cuantificados con un paso Q. La variación de dicho parámetro provoca la variación de la calidad de la imagen suministrada por el codificador. Las componentes de croma del cuadro se submuestrean en una relación 4:2:0 respecto a la componente de luminancia. La agrupación de 4 bloques de luminancia y uno por cada componente de croma se denomina macrobloque. El macrobloque es la unidad básica donde se aplica la técnica de compensación de movimiento. Un conjunto de macrobloques consecutivos horizontalmente se denomina tira o *slice*. La tira es el elemento mínimo donde se puede resincronizar la decodificación en el caso de pérdidas de información. El número de macrobloques consecutivos que forman una tira es seleccionable en el proceso de codificación. En este trabajo se ha considerado la tira como el conjunto de macrobloques que contienen los pixels de 16 líneas consecutivas, es decir, los macrobloques con la misma posición vertical en un cuadro. El presente estudio se ha realizado empleando la resolución estandarizada en la recomendación MPEG-I de 352 por 288 pels, con submuestreo de las componentes de croma tanto vertical como horizontalmente, y 25 imágenes por segundo. Las secuencias analizadas se han obtenido a través de la digitalización de la señal de vídeo PAL correspondiente a distintos discos láser.

Los cuadros de una secuencia SIF pueden codificarse en tres modos diferentes:

i) *Intra* (I): son los cuadros codificados empleando únicamente predicción espacial.

ii) *Predictivo* (P): son los cuadros codificados con predicción temporal hacia atrás, usando como referencia el anterior cuadro I o P, y con predicción espacial

iii) *Predictivo bidireccional* (B): son los cuadros codificados con compensación de movimiento, empleando como referencias la pasada o futura I o P. La compensación de movimiento se puede realizar sobre los macrobloques de una de las referencias o sobre una semisuma de un macrobloque de cada una ellas. También se aplica predicción espacial en los bloques del macrobloque diferencial obtenido.

El almacenamiento o transmisión de las imágenes de una secuencia se hace de forma que el decodificador pueda procesar la información lo antes posible. Para ello, en el almacenamiento o transmisión, las imágenes de referencia preceden a aquellas que las necesitan para ser decodificadas. Este efecto produce en aplicaciones en tiempo real un retardo de reordenación, dado que el orden de decodificación de los cuadros es distinto al de su presentación. A su vez, el codificador también introduce un retardo de proceso dado que necesita imágenes que temporalmente son posteriores para codificar otras que las preceden. Por ello, no es aconsejable en este tipo de aplicaciones que el número de imágenes B consecutivas sea superior a 3 [8].

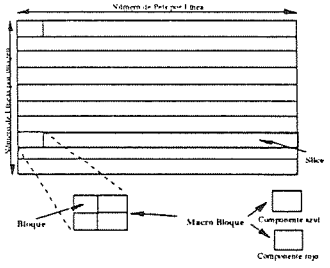


Fig.1. Estructura de un cuadro SIF

La secuencia de imágenes transmitida también se estructura en dos niveles, ilustrados en la Fig.2:

i) *Grupo de imágenes* (Group of Pictures, GoP), compuesto por una imagen I y las imágenes B y P que directa o indirectamente la han utilizado como referencia.

ii) *Subgrupo de imágenes* (Subgroup of Pictures, SGoP) compuesto por una imagen de referencia I o P y las imágenes B que emplearon la imagen I o P como segunda referencia en su proceso de codificación.

Esta estructura periódica da lugar a una entrega de información a la red de la forma que se muestra en la Fig. 3. En ella se observan perfectamente los picos periódicos de tasa, correspondientes a las imágenes I, que afectarán negativamente a la ubicación de los recursos de la red.

3. Caracterización del tráfico de vídeo MPEG VBR como proceso ARIMA

La confección de modelos de tráfico de los servicios soportados por las redes ATM es un objetivo necesario para el dimensionado de los componentes de estas redes y para la evaluación de las prestaciones de los dispositivos que las componen. En particular, el modelado del tráfico de vídeo digital codificado es especialmente importante, ya que los servicios de vídeo bajo petición y distribución de vídeo, además de tener una demanda en continuo crecimiento, son los mayores consumidores de ancho de banda de la red.

A través de los modelos de tráfico se pueden hallar unos descriptores de tráfico adecuados que caractericen un servicio, con lo que se facilitan las labores de gestión de redes ATM. La aportación de modelos de fuentes de tráfico permite, entre otros:

i) Dimensionar la red para soportar una carga de tráficos heterogéneos simultáneos.

ii) Evaluar las prestaciones de los dispositivos o del comportamiento de la red extremo a extremo.

iii) Establecer criterios control de admisión de nuevas llamadas con un nivel de calidad de servicio especificado.

iv) Determinar un control de congestión preventivo que monitorice el comportamiento de la conexión de forma que se respete el contrato usuario-red.

v) Definir las funciones reguladoras del control reactivo y los umbrales de actuación de éste.

vi) Predecir el comportamiento del tráfico, simple o multiplexado, para aumentar el grado de servicio ofrecido y la explotación de los recursos.

Un modelo de tráfico debe capturar los comportamientos del tráfico generado por el servicio que son significativos a la hora de desarrollar las funciones especificadas anteriormente. La bondad del ajuste de un modelo debe ser evaluada en tanto en cuanto capture estos comportamientos. La gran mayoría de los modelos propuestos intentan caracterizar el comportamiento de uno o varios de los parámetros relacionados con la tasa de generación (λ) [9].

Los modelos pueden ajustar distintos parámetros para capturar el comportamiento del tráfico generado por los servicios en diferentes niveles temporales. Cabe distinguir tres niveles temporales [10]:

i) Nivel de llamada o duración de la conexión.

ii) Nivel de ráfaga o variación de la actividad de la conexión.

iii) Nivel de celda relacionado con el tiempo entre llegadas de celdas.

El nivel de llamada ha sido caracterizado, en general, por procesos de Markov y se ha comprobado que este modelo es válido para el dimensionado de redes [11]. El nivel de ráfaga es el más analizado por su impacto en la ubicación de recursos y la calidad de servicio y, por tanto, su caracterización es fundamental. El nivel de celda es considerado en algunos estudios, con el fin de establecer un análisis detallado del comportamiento de los dispositivos. El tamaño de las colas de almacenamiento en los nodos y multiplexores reduce sustancialmente el interés del estudio de este nivel en la ubicación de recursos. El estudio del nivel de celda contribuye al análisis de la variación del retardo entre celdas consecutivas, al diseño de estrategias de sincronización (p.e., los parámetros de los PLL digitales) y al dimensionado de los buffers de contención de multiplexores y conmutadores.

Los diferentes modelos propuestos en la literatura se pueden clasificar, o bien por los parámetros de tráfico que ajustan, o bien por el nivel temporal donde son aplicados. A su vez, admiten clasificaciones en modelos continuos o discretos, atendiendo al tipo de síntesis realizada.

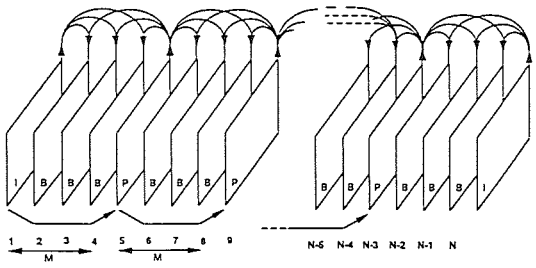


Fig. 2. Estructuración periódica de los modos de codificación de los cuadros en un GoP de N cuadros con M cuadros por SGoP

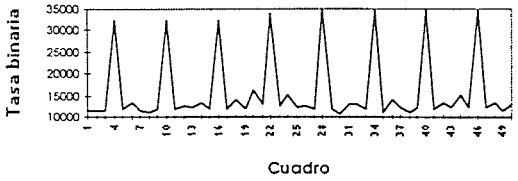


Fig. 3. Bits por cuadro generados por un codificador MPEG con el conjunto de parámetros (Q=6, M=2, N=6)

Las series temporales generadas por los modelos de tráfico son eventos que pretenden definir la tasa instantánea de generación o la tasa media de generación en un intervalo dado. Los procesos de generación de tasa media en intervalos de duración dados proporcionan como eventos el volumen de información a transferir en un intervalo, mientras que los procesos de generación de llegadas hacen hincapié en como se producen las trasferencias de información indicando el tiempo entre dos llegadas consecutivas de paquetes de información. En el caso de modelos en tiempo discreto los valores generados son números enteros positivos. Los modelos en tiempo continuo operan con números reales, aunque, posteriormente pueden ser truncados cuando se emplean en simulaciones. Dentro de los trabajos presentados en la literatura se han desarrollado también modelos compuestos. Estos modelos conjugan la generación de tasas en intervalos dados y tasas instantáneas. Se basan en desarrollar un proceso que sintetice el tiempo entre llegadas y otro proceso que determine el número de llegadas en ese instante. Se puede denominar procesos de llegada en grupo (Batch Arrival Processes, BAP).

3.1. Procesos ARIMA

Estos procesos han sido ampliamente estudiados en la literatura y en su forma más general se denominan procesos autoregresivos, integrativos de media móvil (autoregresive integrative moving average, ARIMA) [12][13]. Los modelos autoregresivos se emplean en el contexto de fuentes de tráfico sintéticas o en predicción de tráfico para la generación de tasas medias en intervalos de duración fija [14][15]. Los modelos ARIMA(p,d,q) se descomponen en una parte autoregresiva de orden p, una parte integrativa de orden d y una parte de media móvil de orden q. La parte autoregresiva refleja la dependencia entre la generación actual y las pasadas p generaciones. Así, para un proceso AR(p) los valores generados en una serie temporal $Y=(y_0,y_1,..,y_n)$ se obtienen de los p valores pasados y un factor independiente de la serie temporal, modelable como un proceso de valores idénticamente distribuidos e independientes entre sí $W=(w_0,w_1,..w_n)$. Habitualmente, los valores de la serie W se sintetizan a partir de la realización de una variable aleatoria gaussiana con una media y una desviación típica relacionadas directamente con los correspondientes momentos del proceso AR a generar. De forma que:

$$y(n)=a_1y(n-1)+a_2y(n-2)+...+a_py(n-p)+w(n) \quad (1)$$

donde los términos a_i son coeficientes constantes.

La parte MA(q) del proceso refleja la dependencia en la generación de los valores pasados del proceso independiente que contribuye en el valor obtenido. Así un proceso MA (q) podría expresarse como :

$$x(n)=b_0w(n)+b_1w(n-1)+b_2w(n-2)+...+b_qw(n-q) \quad (2)$$

donde los términos b_i son coeficientes constantes.

La contribución integrativa, pretende modelar la no estacionariedad de los momentos del proceso estocástico. Si bien podría considerase dentro de la parte AR por su formulación, su síntesis depende de factores distintos de la parte autoregresiva. Así, la parte integrativa también muestra la dependencia con valores pasados de la realización pero depende de los momentos del proceso no estacionario más que de la relación temporal de las generaciones.

El orden d de la parte integrativa queda fijado por el orden del momento del proceso estocástico no estacionario.

En general, se puede expresar la dependencia:

$$z(n)=c_1z(n-1)+c_2z(n-2)+...+c_dz(n-d)+w(n) \quad (3)$$

donde:

$$c_i=\binom{d}{i}(-1)^{i+1}, \quad i\in\{1,2,\cdots,d\} \quad (4)$$

Como caso de aplicación, un proceso cuya media no es estacionaria, pero sí sus momentos de orden superior, tendría una parte integrativa de orden 1. Los procesos integrativos de orden 1 reciben el nombre de marcha aleatoria o "random walk" y no están acotados.

La interpretación de un proceso ARIMA(p,d,q) puede ser realizada definiendo el operador de retardo z^{-1} [16]. De forma que, la expresión general de un proceso ARIMA(p,d,q) quedaría expresada por su transformada Z como:

$$Y(z)=\frac{B(z)}{A(z)C(z)}W(z) \quad (5)$$

donde:

$$Y(z)=Z(y)=\sum_{n=0}^{\infty}y_nz^{-n}; \quad A(z)=Z(a)=\sum_{n=0}^{p}a_nz^{-n}$$
$$W(z)=Z(w)=\sum_{n=0}^{\infty}w_nz^{-n}; \quad B(z)=Z(b)=\sum_{n=0}^{q}b_nz^{-n}$$
$$C(z)=Z(c)=\sum_{n=0}^{d}c_nz^{-n} \quad (6)$$

Interpretando esta expresión como la relación entre entrada y salida de un filtro digital con excitación wn y cuya salida es yn en un instante dado, podríamos definir la función de transferencia del filtro H(z) como:

$$H(z)=\frac{Y(z)}{W(z)}=\frac{B(z)}{A(z)C(z)} \quad (7)$$

Obsérvese que las raíces del polinomio B(z) se corresponden con los ceros del filtro y los ceros de A(z) y C(z) con los polos. Según la definición de los valores c_i realizada en la expresión (4), el orden integrativo define la multiplicidad del polo en z = 1, el cual genera la inestabilidad de la respuesta impulsional. El resto de polos z_k obtenidos a partir

de A(z), con $k \in \{1,2,...,p\}$, se encontrarán en el círculo unidad del plano Z, es decir, cumplirán $|z_k| < 1$. De forma esquemática se puede representar el modelo ARIMA como se muestra en la Fig. 4.

La serie temporal w, en general, se denomina serie residual. Se suele considerar como la parte impredictible de la siguiente generación a partir de los valores anteriores. Estos procesos estocásticos son incorrelados y su distribución suele ser gaussiana.

3.2 Caracterización del tráfico MPEG

En este trabajo se ha desarrollado un nuevo predictor basado en la síntesis de un modelo ARIMA del tráfico VBR MPEG. Este predictor se basa en la caracterización previa de dicho tipo de tráfico como proceso ARIMA. Se han utilizado tres secuencias para elaborar y evaluar el modelo, "Live in Central Parc" del grupo musical América, "Jurasic Park" y "Geografía de Catalunya", de 34000, 174000 y 51000 imágenes respectivamente. Estas secuencias han sido codificadas con parámetros (Q=9, M=2, N=6), (Q=6, M=2, N=6) y (Q=6, M=2, N=6). Las dos primeras secuencias presentan las características típicas de actividad y complejidad, mientras que la tercera se caracteriza por la complejidad y corta duración de las escenas. Asimismo, se han contrastado los resultados obtenidos con la codificación de la secuencia "Live in Central Park" con parámetros (Q=9,M=2,N=4).

Para elaborar el modelo ARIMA se procede inicialmente a la determinación de la parte integrativa. La dependencia a largo término provoca que la tasa media de grupos de imágenes varíe suavemente. Esta variación llega a alcanzar niveles máximo y mínimos muy distantes. Sin embargo, la varianza se mantiene casi constante. Esto nos permite concluir que la parte integrativa del modelo debe ser de orden 1.

Para determinar cuál es el valor de las partes AR y MA será necesario extraer la parte integrativa. La serie temporal resultante s(n) será la salida de un filtro FIR, cuya función de transferencia es $(1-z^{-1})$, excitado con la serie temporal generada por el codificador. Se puede comprobar que la serie temporal s(n), es un proceso estocástico de media 0 y con coeficientes de autocorrelación invariantes. Este análisis estadístico se ha realizado con las tres secuencias utilizando bloques de al menos 15000 cuadros y con retardos de autocorrelación de hasta 100 unidades de s(n). La función de distribución de probabilidad se ajusta perfectamente a una distribución gaussiana en todos los casos. La única diferencia observada en las series temporales es la desviación estándar. Esta discrepancia está relacionada con la variabilidad y complejidad de las secuencias.

La serie temporal s(n), presenta un comportamiento estacional de período N=4 o N=6 según el parámetro de codificación MPEG elegido. La estacionalidad se reflejada en la función de autocovarianza de la serie temporal, como se ilustra en la Fig. 5. Utilizando los picos de la función de autocovarianza que aparecen en múltiplos de N se ha sintetizado la parte AR. Para la estimación de los coeficientes de la parte AR se ha empleado la técnica de mínimos cuadrados sobre la ecuaciones de Yule-Walker modificadas. El orden del modelo estacional obtenido es 2, por lo que la función de transferencia de la parte AR será:

$$A(z) = (1-0.6950\, z^{-6}-0.3\, z^{-12}) \qquad (8)$$

Para determinar la parte MA se debe extraer la parte AR de la serie s(n). De la misma forma que para la parte integrativa, se aplica como entrada a un filtro FIR la secuencia s(n) y se obtiene a la salida la serie a modelar x(n). La función de transferencia del filtro será el denominador de la función de transferencia del filtro AR. A través de la función de autocorrelación parcial de la serie x(n) y utilizando la técnica de mínimos cuadrados se obtiene un filtro MA de orden 13, cuya función de transferencia es:

$$\begin{aligned} B(z) = {} & 1 - 0.7618\, z^{-1} + 0.1136\, z^{-2} - 0.1676\, z^{-3} + \\ & + 0.0195\, z^{-4} - 0.0451\, z^{-5} - 0.1691\, z^{-6} + 0.0386\, z^{-7} + \\ & + 0.0397\, z^{-8} - 0.0268\, z^{-9} + 0.0523\, z^{-10} - 0.0371\, z^{-11} - \\ & - 0.1910\, z^{-12} + 0.1351\, z^{-13} \end{aligned} \qquad (9)$$

Finalmente, el modelo ARIMA tendrá como función de transferencia:

$$H(z)=B(z)/[A(z)(1-z^{-1})] \qquad (10)$$

Para evaluar el comportamiento del modelo sintetizado se ha realizado un análisis de los errores de predicción para todas las secuencias. En la Fig. 6 se presenta la autocorrelación de los residuos y los intervalos de confianza para el 99%.

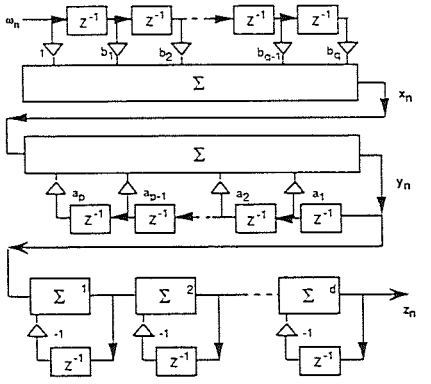


Fig. 4. Esquema de un filtro ARIMA(p,d,q)

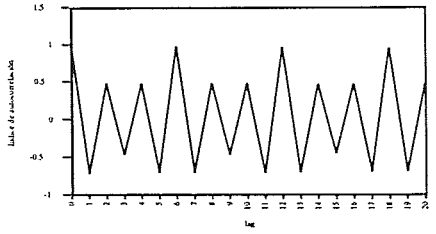


Fig. 5. Indice de autocorrelación de la serie temporal "Live in Central Park" una vez extraída la parte integrativa codificada con (Q=9, M=2, N=6)

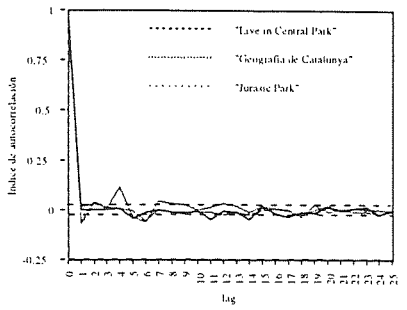


Fig. 6. Indice de autocorrelación de las series residuales

4. Predicción de la tasa de salida del codificador MPEG VBR

A partir de la caracterización obtenida en el apartado anterior para la tasa generada por cuadro por el codificador de vídeo, en esta sección se plantea la posibilidad de llevar a cabo una predicción de dicha tasa en función de sus valores anteriores. Como ya se ha comentado, dicha predicción será utilizada en la conformación del tráfico previa a su entrega a la red.

Recordando que la componente integrativa del modelo obtenido es de orden 1, podemos expresar dicho modelo de la siguiente manera:

$$y(n) = b_0 w(n) + b_1 w(n-1) + \cdots + b_q w(n-q) + \\ + a'_1 y(n-1) + a'_2 y(n-2) + \cdots + a'_{p+1} y(n-p-1) \quad (11)$$

donde los coeficientes a'_i se obtienen de la convolución de los anteriores coeficientes a_i y c_i.

Partiendo de (12), la predicción de la muestra (n+1) a partir de la muestra n y anteriores sería de la forma:

$$\hat{y}(n+1) = b_0 \hat{w}(n+1) + b_1 w(n) + \cdots + b_q w(n-q+1) + \\ + a'_1 y(n) + a'_2 y(n-1) + \cdots + a'_{p+1} y(n-p) \quad (12)$$

Sin embargo, los valores de la serie w(n) en este contexto de predicción son desconocidos, ya que el predictor trabajará únicamente con los valores anteriores de la serie y(n). Además, el valor $\hat{w}(n+1)$ es un valor futuro. La mejor predicción que se puede hacer para él es tomar el valor medio de la serie, el cual según lo visto en el apartado anterior es igual a cero. Por otra parte, recordando que:

$$y(n) = b_0 w(n) + b_1 w(n-1) + \cdots + b_q w(n-q) + \\ + a'_1 y(n-1) + a'_2 y(n-2) + \cdots + a'_{p+1} y(n-p-1) \quad (13)$$

y que, por tanto:

$$\hat{y}(n) = b_0 \hat{w}(n) + b_1 w(n-1) + \cdots + b_q w(n-q) + \\ + a'_1 y(n-1) + a'_2 y(n-2) + \cdots + a'_{p+1} y(n-p-1) \quad (14)$$

se tiene que, restando (15) de (14):

$$y(n) - \hat{y}(n) = b_0 \left(w(n) - \hat{w}(n) \right) \quad (15)$$

Tomando, al igual que en el caso anterior, la esperanza de w(n) como la mejor predicción posible para $\hat{w}(n)$, y recordando que esta es igual a cero, se obtiene:

$$y(n) - \hat{y}(n) = b_0 w(n) \quad (16)$$

de donde:

$$w(n) = \frac{y(n) - \hat{y}(n)}{b_0} \quad (17)$$

De esta forma, se obtiene la siguiente predicción, en función de las muestras reales y predecidas anteriores:

$$\hat{y}(n+1) = \frac{b_1}{b_0}\left(y(n) - \hat{y}(n) \right) + \cdots + \\ + \frac{b_q}{b_0}\left(y(n-q+1) - \hat{y}(n-q+1) \right) + \\ + a'_1 y(n) + a'_2 y(n-1) + \cdots + \\ + a'_{p+1} y(n-p) \quad (18)$$

Dicha predicción es representable gráficamente como se observa en la Fig 7. De esta forma, el predictor nos da la mejor estimación posible para la muestra (n+1) en función de las n anteriores. También es posible, utilizando dichas n muestras, la estimación de las salidas (n+2), (n+3), etc. Es decir, podemos utilizar el mismo predictor y las mismas n muestras para la obtención de las predicciones "1 adelante", "2 adelante", etc. Para la obtención de la predicción de la muestra (n+j) no hay más que inyectar al predictor la estimación de la muestra (n+j-1).

La validez del predictor ha sido comprobada para las tres secuencias de estudio. En la Fig. 8. se muestra la estimación de una serie de 40 cuadros de la película "Jurasic Parc". Estos 40 cuadros pertenecen a un cambio brusco de escena en la película, es decir, a uno de los momentos en los cuales la predicción se aleja más de la serie original. El principio corresponde a una escena con poco movimiento, obteniéndose una tasa media de aproximadamente 50000 bits por cuadro. Tras el salto, se pasa a una escena más compleja espacialmente y con mucho movimiento, lo cual provoca tasas alrededor de los 200000 bits por cuadro. Se puede observar como la predicción es muy precisa durante lo que podríamos llamar régimen permanente, y como se adapta rápidamente a los cambios de escena.

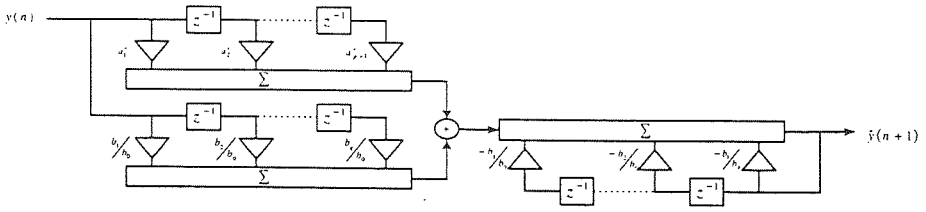


Fig. 7. Predictor ARIMA.

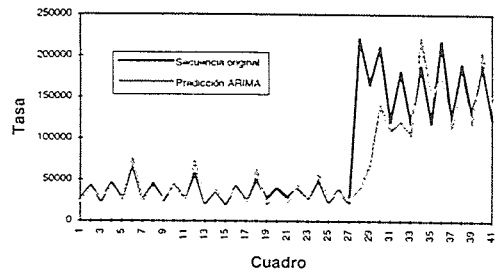


Fig. 8. Predicción "1 adelante" en cambio de escena.

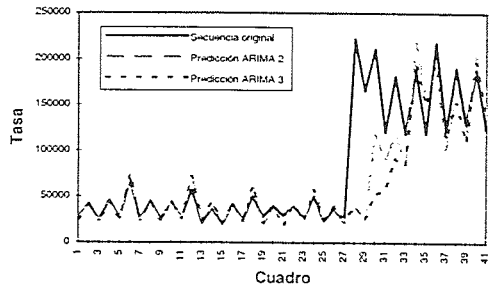


Fig 9. Predicción "2 adelante" y "3 adelante"

Como se ha mencionado anteriormente, el predictor se puede utilizar también para la obtención de muestras "j adelante". En la Fig. 9 se muestran, para el mismo intervalo de la serie anterior, las predicciones "2 adelante" y "3 adelante".

5. Conformación del tráfico MPEG VBR

Para minimizar la variabilidad de la tasa generada es aconsejable la conformación del tráfico generado por el codificador. El suavizado se puede llevar a cabo a través del almacenamiento de las imágenes y su posterior transmisión, una vez determinada la tasa media requerida. Este mecanismo sólo puede ser empleado en servicios que puedan aceptar un retardo de transmisión superior al tiempo necesario para almacenar el grupo de imágenes a suavizar. Para los servicios que no admitan elevados retardos es aconsejable el empleo de técnicas de predicción que permitan, sin introducir retardos adicionales, transmitir a una tasa estimada. De esta manera, se puede reducir la variabilidad del tráfico VBR MPEG de forma similar al empleo de la suavización, sin necesidad de introducir un retardo adicional. Sin embargo, el empleo de técnicas de predicción puede provocar un error de estimación que incremente la tasa de transmisión de las imágenes notablemente. Este efecto aparecerá cuando se produzca un cambio de escena o en escenas con movimientos de cámara. El incremento de la tasa de generación será más importante cuanto menor sea el retardo máximo aceptable por el servicio. Sin embargo, este efecto puede ser reducido significativamente si el codificador VBR MPEG se diseña para que en los cambios escena la variación de la tasa no sea brusca aunque se reduzca la calidad de la imagen. Este efecto no será apreciable, dado que el sistema visual humano necesita más de 1 segundos para observar un cambio en la secuencia, y no es tan sensible a la calidad de la imagen cuando existe gran actividad en la escena.

En este apartado se presenta un nuevo esquema de conformación de tráfico MPEG VBR, basado en el predictor desarrollado en la sección anterior. Además, sus prestaciones son comparadas con los clásicos sistemas de almacenamiento, los cuales son presentados en primer lugar.

5.1 Métodos clásicos de suavizado

Uno de los sistemas más clásicos de suavizado se basa en el almacenamiento de un número determinado de cuadros, entregándolos después a la red a la tasa media obtenida para todo el grupo. Este tipo de suavizado se ha llamado en otros trabajos suavizado ideal [17], ya que calcula exactamente la tasa media a la que debe entregar la información a la red. Generalmente, el número de cuadros utilizado para este suavizado es N, es decir, el número de cuadros de un GoP. Así, llamando S(n) al número de bits necesarios para codificar el cuadro n, la tasa entregada a la red durante el siguiente GoP será:

$$r = \frac{S(n) + S(n+1) + \cdots S(n+N-1)}{N\tau} \tag{19}$$

donde hemos denominado con τ al tiempo de cuadro.

Por la propia filosofía de este tipo de suavizado, un cuadro en concreto puede sufrir un retardo de almacenamiento previo antes de ser entregado a la red de hasta ($2N\tau$) segundos. En otras palabras, en un sistema funcionando a 30 cuadros por segundo y con 6 cuadros por GoP, una imagen se puede ver retardada hasta 400 mseg. antes de ser entregada a la red. Este retardo, si bien puede ser admisible para un servicio de difusión, como puede ser el vídeo bajo demanda, no lo es para servicios interactivos, como la videoconferencia. En la Fig. 10 se muestra una porción de la secuencia "Jurasic Parc", y su conformación mediante el suavizado ideal. Por otra parte, el retardo sufrido por los cuadros representados en la figura anterior se puede observar en la Fig 11, representados en tiempos de cuadro. Recuérdese que dicha secuencia ha sido codificada con un parámetro N igual a 6.

Otro tipo de conformación posible consiste en tomar como tasa de salida de un cuadro el promedio de los N anteriores, e ir actualizando dicha tasa para cada cuadro. El suavizado llevado a cabo por este método, que podemos llamar de ventana deslizante, se muestra en la Fig. 12, y su retardo en la Fig. 13.

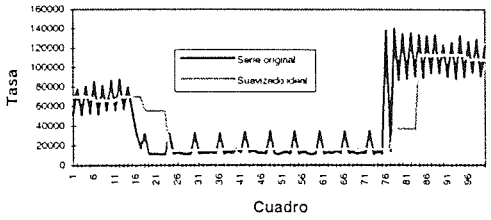


Fig. 10. Suavizado ideal

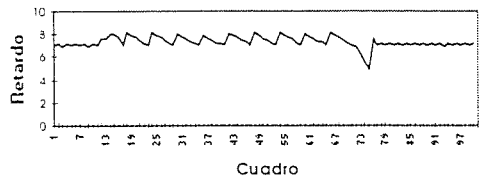


Fig. 11 Retardo por cuadro con suavizado ideal

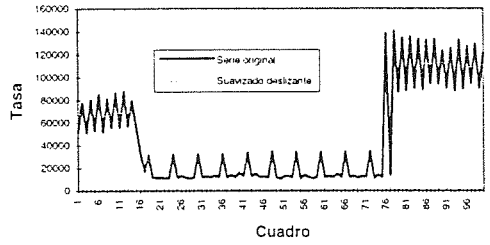


Fig. 12. Suavizado deslizante

Los resultados obtenidos para las tres secuencias bajo estudio para cada uno de los dos modelos de suavizado anteriores se presentan en la tabla 1. Se puede observar como en cada caso se ha reducido de forma considerable tanto el coeficiente cuadrático de variación de la serie de salida,C^2, como su relación de rafagueo, B. En el peor de los casos, la reducción de esta última es del 20%, lo cual incidirá en una considerable mejora a la hora de asignar recursos en una red ATM.

El principal inconveniente de ambos métodos de suavizado es, como se ha comentado anteriormente, el retardo introducido a causa del almacenamiento, que llega a ser de más de 10 tiempos de cuadro en el peor de los casos. Hay que tener en cuenta que la calidad del servicio ofrecida vendrá determinada por el retardo máximo que pueda sufrir un cuadro en concreto de la secuencia que se está transmitiendo.

Por otra parte, en las Figs. 14 y 15 se representa la función de autocorrelación de la secuencia "Jurasic Parc", junto con la de las salidas suavizadas. En ellas se destaca claramente como se han eliminado los picos periódicos de tasa elevada que tanto perjudican la eficiente asignación de recursos.

En definitiva, se puede concluir que para ambos métodos las prestaciones ofrecidas en cuanto a conformación de tráfico son buenas para servicios que no presenten requisitos de retardo. Sin embargo, si dichos requisitos son más restrictivos, los sistemas de almacenamiento pueden hacer bajar la calidad del servicio por debajo de la mínima requerida.

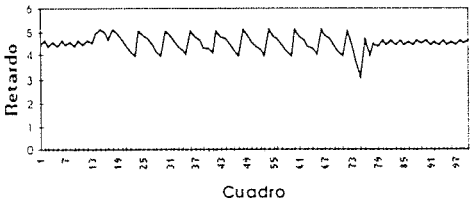


Fig. 13. Retardo por cuadro con suavizado deslizante

Tabla 1. Sistemas de suavizado con almacenamiento

Secuencia	Suavizado	Bits/Cuadro						Retardo		
		Min	Max	Media	σ	C^2	B	Min	Max	Media
América	Original	5856	92942	25537	14818.5	0.34	3.64	1	1	1
	Ideal	0	63769	25537	7763.13	0.09	2.50	4.4	7.6	6.4
	Deslizante	0	63764	25537	7749.63	0.09	2.50	3.6	6.1	4.8
Geografía	Original	4383	297635	59672	46919.5	0.62	4.99	1	1	1
de Cataluña	Ideal	0	240943	59672	38433.9	0.41	4.04	3.5	8.9	6.5
	Deslizante	0	240939	59672	38448.9	0.42	4.04	3.1	6.9	4.8
Jurasic Parc	Original	7307	261901	43924	31556.8	0.52	5.96	1	1	1
	Ideal	0	207708	43924	23164.6	0.27	4.73	4.5	9.6	7.4
	Deslizante	0	209392	43924	23155.8	0.28	4.77	3.0	10.2	5.6

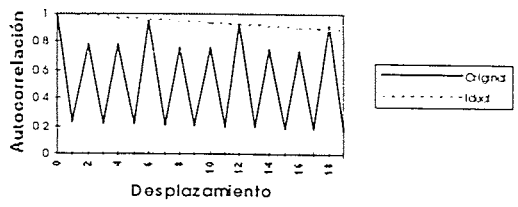


Fig 14. Autocorrelación de la serie conformada mediante el suavizado ideal

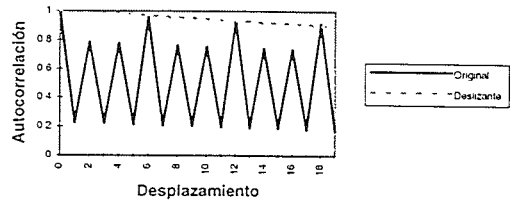


Fig 15. Autocorrelación de la serie conformada mediante el suavizado deslizante

5.2 Conformación predictiva

Con objeto de adaptar la conformación del tráfico a servicios con fuertes restricciones temporales, se introduce en este apartado un nuevo método de conformación basado en la predicción de muestras futuras. En su forma más general, el conformador de tráfico presenta la estructura de la Fig. 16. En ella se puede observar un buffer de almacenamiento, donde se coloca la información correspondiente a los nuevos cuadros, y que será extraída a la velocidad que le indique el controlador. Dicho controlador, en los casos vistos en el apartado anterior, no haría más que promediar las longitudes de los N cuadros anteriores, para obtener la velocidad de extracción del buffer. En el caso del suavizado ideal esta velocidad sería constante durante todo un GoP, mientras que en el caso deslizante se iría actualizando cuadro a cuadro.

Para el suavizado predictivo, el controlador estará formado por un predictor ARIMA y dos registros tal y como se muestra en la Fig. 17. La idea consiste en la utilización de un número determinado de muestras pasadas y otro de muestras futuras predecidas para el cálculo de la tasa de salida. Según el dibujo de la figura, se utilizarían K muestras en total, de las cuales L_1 serían pasadas, incluyendo la actual, y por tanto exactas, y L_2 serían muestras futuras. Es decir, cada vez que la tasa de un cuadro nuevo llega al conformador, se almacena dentro del grupo L_1. Por otra parte, con esta misma tasa como entrada, se hace funcionar al predictor L_2 veces, de forma que se calculan las predicciones "1 adelante", "2 adelante", y así hasta la "L_2 adelante". Promediando esas K muestras, se obtendrá la tasa de salida deseada.

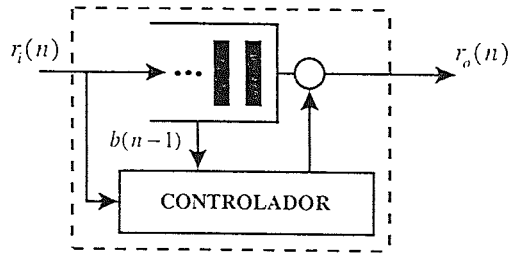


Fig. 16. Esquema general del conformador de tráfico

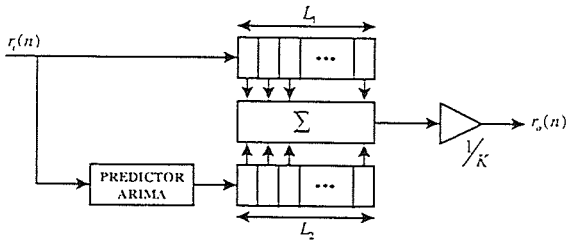


Fig. 17. Controlador para el suavizado predictivo

En particular, en el conformador presentado en este apartado no se hará necesario tomar un número de muestras pasadas L_1 superior a la unidad, lo cual quiere decir que nos bastará con la muestra actual para calcular la tasa de salida. En cuanto al número de muestras predecidas, deberá ser el necesario para completar al menos un GoP. Es decir, como mínimo L_2 deberá ser igual a N-1. Así, se dispondrá de información del tamaño de todos los cuadros I, B y P dentro del GoP, con lo que la tasa será lo más precisa posible. Por otra parte, la posibilidad de aumentar el número de muestras predecidas para disponer de la información de más de un GoP ha sido rechazado de forma intuitiva en otros trabajos previos [17]. En este momento es posible corroborar dicha hipótesis, ya que predecir un segundo GoP daría como resultado la misma tasa de salida. Esto es así debido a que la predicción de las tasas del segundo GoP sería prácticamente igual a la de las tasas del primero.

Además, se incorpora la posibilidad de utilizar como información adicional el contenido del buffer previo a la introducción de la muestra actual. Esta información será necesaria en el caso de que el conformador quiera garantizar una calidad de servicio determinada, en lo que a retardo de almacenamiento máximo de un cuadro se refiere. El modo de funcionamiento será el mismo que el presentado hasta ahora, con una pequeña modificación que garantizará el cumplimiento de la cota máxima de retardo. Dicha modificación consiste en el cálculo de una tasa mínima de salida para cada cuadro n, en función de su longitud S(n), del contenido del buffer en el momento de su llegada b(n-1), y del retardo máximo permitido, D:

$$r_{min}(n) = \frac{S(n) + b(n-1)}{D} \tag{20}$$

El controlador funcionará de la misma forma expuesta anteriormente, pero teniendo en

cuenta que existe una cota mínima de salida para cada cuadro que no haya sido aún completamente extraído del buffer. Por tanto, la tasa de salida será como mínimo el máximo de dichas cotas.

En la Fig. 18 se representa el suavizado llevado a cabo por este tipo de conformador para una fracción de la secuencia "Jurasic Parc", en la cual se produce una bajada brusca de tasa. Se presentan los casos en los que el retardo está limitado a 2 y 3 tiempos de cuadro, junto con la secuencia original. Es de notar que la secuencia original coincide con la que se tendría si se admitiese un retardo como máximo de 1 tiempo de cuadro. En la Fig. 19, se contemplan los casos de retardo máximo igual a 4, 5 y 6 tiempos de cuadro.

En las figuras se puede observar como al aumentar el máximo retardo permitido se consigue mejor suavizado en las zonas de transición. Los retardos sufridos por los distintos cuadros para cada uno de los casos anteriores se representan en la Fig. 20, donde se puede observar como en ningún caso el retardo supera el máximo permitido.

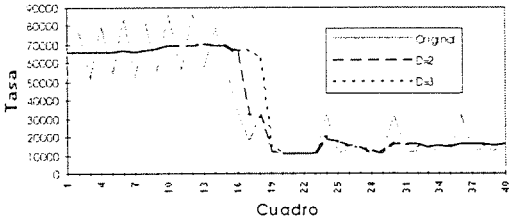


Fig. 18. Suavizado predictivo en bajada de la serie "Jurasic Parc"

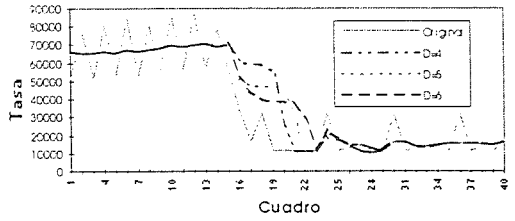


Fig. 19. Suavizado predictivo en bajada de la serie "Jurasic Parc"

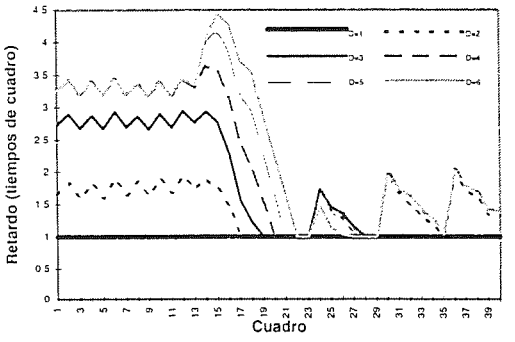


Fig. 20. Retardo en transición de bajada

El comportamiento del conformador en una transición de subida se muestra en la Fig. 21. En esta ocasión tan sólo se representa hasta el caso D=3, ya que a partir de este valor las curvas son coincidentes con valores de D superiores. El retardo puede observarse en la Fig. 22.

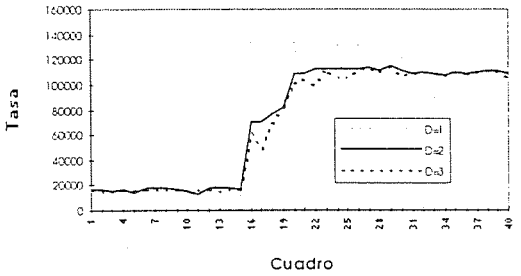


Fig. 21. Suavizado predictivo en subida de la serie "Jurasic Parc"

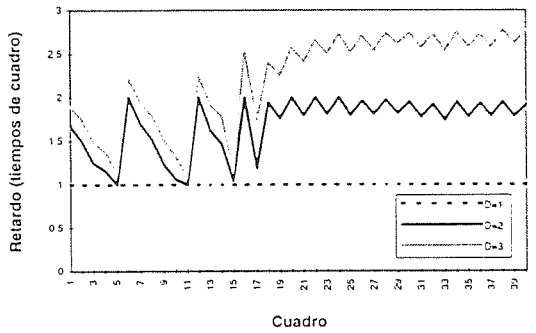


Fig. 22. Retardo en transición de subida

De las gráficas anteriores se deduce que para este tipo de conformador los momentos más críticos son aquellos en los que se producen bajadas bruscas de tasa. Esto es debido a que el predictor ordenará que las tasas de salida sean bajas, pero en el buffer aún quedará un remanente grande de cuando la tasa era elevada. Es decir, aún quedan cuadros, o parte de ellos, en el buffer, que deben ser extraídos a tasas más elevadas. Es entonces cuando entra en funcionamiento el corrector de tasa, ajustándose a las cotas mínimas previamente calculadas.

En la tabla 2 aparecen los resultados estadísticos obtenidos con este conformador, para los casos de estudio, es decir, para retardos máximos de 1, 2, 3, 4, 5 y 6 tiempos de cuadro. Se comprueba como de nuevo se ha conseguido una importante reducción tanto para el coeficiente cuadrático de variación como para la relación de rafagueo. Además, estos valores permanecen prácticamente constantes a partir de D=2.

Para determinar finalmente hasta qué punto es válido el conformador presentado, debemos comprobar la autocorrelación de las series generadas. Recordemos que el principal motivo del suavizado es la eliminación de las llegadas periódicas de tasa elevada. En la Fig. 23 se puede observar dicha función de autocorrelación para los casos D=1, 2 y 3. Como siempre, el caso D=1 (o lo

que es lo mismo, la secuencia original), presenta una fuerte correlación. Para D=2, se reduce sustancialmente, pero aún se pueden observar algunos picos cada (kN+1) desplazamientos. Sin embargo, para D=3 la función queda completamente descorrelada. La diferencia entre estos dos últimos casos se aprecia mejor en la Fig. 24, en la cual se ha suprimido la autocorrelación de la secuencia original con objeto de ganar resolución para los casos D=2 y 3. Las autocorrelaciones obtenidas para D=4, 5 y 6 ni siquiera se reproducen en la figura, ya que presentan prácticamente la misma forma que la obtenida para D=3. De dichas representaciones se deduce inmediatamente que a partir de una restricción de D=3 conseguimos la máxima descorrelación posible de la serie de salida, lo cual llevaría a una multiplexación estadística óptima de fuentes de este estilo. Si los requisitos de retardo fuesen tan estrictos como para no poder permitir un retardo de suavizado mayor de 2 tiempos de cuadro, es posible la utilización del mismo conformador con un aumento de los recursos necesarios para su ubicación en la red.

Las distintas pruebas presentadas para la secuencia "Jurasic Parc", se han realizado también para las otras dos secuencias bajo estudio. Los resultados estadísticos obtenidos para los casos D=3 y D=4 se reflejan en la tabla 3.

Finalmente, se procede a la comparación de los resultados obtenidos con el nuevo conformador, para el caso en el que el retardo máximo permitido es igual a 3 tiempos de cuadro, con los que se obtuvieron para los sistemas clásicos, utilizando la secuencia "Jurasic Parc". En la tabla 4 se puede comprobar como la reducción del coeficiente cuadrático de variación y de la relación de rafagueo es aproximadamente igual para todos los casos. Sin embargo, el conformador predictivo proporciona un retardo mucho menor.

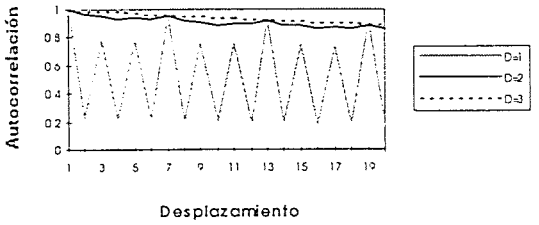


Fig. 23. Autocorrelación de las series de salida del conformador predictor

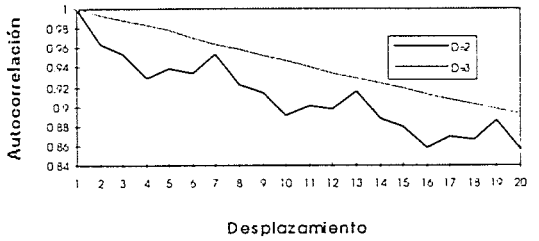


Fig. 24. Autocorrelación de las series de salida del conformador predictor

Tabla 2. Conformación predictiva de la secuencia "Jurasic Parc"

D	Bits/Cuadro						Retardo		
	Min	Max	Media	σ	C^2	B	Min	Max	Media
1	7307	261901	43925	31516.8	0.52	5.96	1	1	1
2	7741	209530	43925	23691.7	0.29	4.77	1	2	1.4
3	9162	209530	43925	23218	0.28	4.77	1	3	2
4	8292	209530	43925	23170.1	0.28	4.77	1	4	2.2
5	8292	209530	43925	23153.3	0.28	4.77	1	5	2.4
6	8292	209530	43925	23145.3	0.28	4.77	1	6	2.6

Tabla 3. Resultados estadísticos para las secuencias "América" y "Geografía de Cataluña"

Secuencia	D	Bits/Cuadro						Retardo		
		Min	Max	Media	σ	C^2	B	Min	Max	Media
América	1	5856	92942	25537	14818.5	0.34	3.64	1	1	1
	3	5615	63534	25537	7786.87	0.09	2.49	1	3	2
	4	5615	63534	25537	7765.38	0.09	2.49	1	4	2.6
Geografía de Cataluña	1	4383	297635	59672	46919.5	0.62	4.99	1	1	1
	3	4660	241697	59672	38630.7	0.43	4.05	1	3	2
	4	4660	241697	59672	38511	0.42	4.05	1	4	2.3

Tabla 4. Comparación de los métodos de suavizado

Suavizado	Bits/Cuadro						Retardo		
	Min	Max	Media	σ	C^2	B	Min	Max	Media
Ninguno	7307	261901	43924	31556.8	0.52	5.96	1	1	1
Ideal	0	207708	43924	23164.6	0.27	4.73	4.5	9.6	7.4
Deslizante	0	209392	43924	23155.8	0.28	4.77	3.0	10.2	5.6
Predictivo	9162	209530	43925	23218	0.28	4.77	1	3	2

En la Fig. 25 se muestra el suavizado proporcionado por los tres métodos en momentos de gran transición, y en la Fig. 26 se presenta el retardo en dichos momentos. De ambas se deduce el mejor comportamiento presentado por el conformador predictivo.

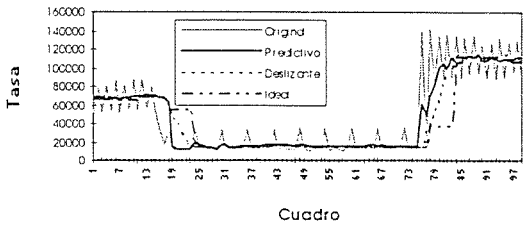


Fig.25. Comparación de suavizados

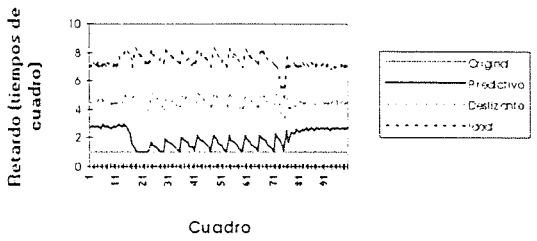


Fig. 26. Comparación de retardos

6. Conclusiones

En este artículo se ha presentado un nuevo conformador para tráfico de vídeo MPEG VBR. Dicho conformador emplea técnicas de predicción, basándose en la caracterización del tráfico real como proceso ARIMA. Las dependencias a largo término son aproximadas mediante la parte integrativa del modelo. Como primera conclusión del trabajo se puede extraer la invarianza de los coeficientes del modelo obtenida para todas las secuencias codificadas. Es de resaltar que la invarianza se mantiene incluso respecto al cambio en la calidad de la imagen solicitada al codificador, mediante la variación del parámetro Q.

Las prestaciones del nuevo conformador de tráfico han sido comparadas con las de los habituales sistemas de suavizado por almacenamiento. A través de dicha comparación, se ha puesto de manifiesto como la calidad del suavizado es tan buena para el conformador predictivo como para los anteriores. Sin embargo, mientras que los suavizadores por almacenamiento presentan retardos inaceptables para servicios interactivos, el conformador predictivo mantiene dicho retardo por debajo de los requisitos establecidos a priori.

Referencias

[1] D. Le Gall. "MPEG: A Video Compression Standard for Multimedia Applications", *Communications of the ACM*, vol **34**, no. 4, pp. 305-313 (April 1991).

[2] M. de Prycker, "Asynchronous Transfer Mode. Solution for Broadband ISDN", Second Edition, Ellis Horwood (1993).

[3] P. Pancha, M. El Zarki, "MPEG Coding For Variable Bit Rate Video Transmission", *IEEE Communications Magazine*, vol. **32**, no. 5, pp. 54-66 (May 1994).

[4] T. Tanaka, S. Okubo, H. Hashimoto and H. Yasuda, "A study on comparisson between VBR and CBR video service in ATM enviroment", *Proceedings of the IEEE ICC'92*, pp. 551-555 (1992).

[5] B. G. Haskell and A. R. Reibman, "Multiplexing of Variable Rate Encoded Streams", *IEEE Transactions on Circuits and Systems for Video Technology*, vol. **4**, no. 4, pp. 417-424 (August 1994).

[6] CODING OF MOVING PICTURES AND ASSOCIATED AUDIO FOR DIGITAL STORAGE MEDIA AT UP TO ABOUT 1.5 Mbps, ISO/IEC 11172-2 (November 25, 1991).

[7] GENERIC CODING OF MOVING PICTURES AND ASSOCIATED AUDIO, Recommendation H.262, ISO/IEC 13818-2 (November 4, 1993).

[8] M. Kawashima, C. Chen, F. Jeng, S. Singhal, "Adaptation of MPEG Video-Coding Algorithm to Network Applications", *IEEE Transactions on Circuits and Systems for Video Technology*, vol. **3**, no. 4, pp. 261-269 (August 1993).

[9] J. Andrade, "Statistical parameter to describe all traffic generated by broadband services", *Comunicaciones de Telefónica I+D*, vol.**4**, no.2, pp. 81-88 (Julio-Diciembre 1993).

[10] O.Gihr and P. Tran-Gia, "A Layered Description of ATM Cell Traffic Streams and Correlation Analysis", *Proceedings of the IEEE INFOCOM'91*, pp.137-142 (1991).

[11] J. Y. Hui, "Switching and Traffic Theory for Integrated Broadband Networks", Kluwer Academic Publishers (1990).

[12] G. E .P. Box, G. M. Jenkins, G. C. Reinsel, "Time Series Analysis", Prentice Hall (1994).

[13] Alparone et al, "Statistical Models for Video Packet Communications", *Digital Signal Processing 91*, Elsevier Science Publishers, pp. 654-659 (1991).

[14] Grunenfelder, Cosmas, Manthrope and Odinma-Okafor,"Characterization of Video Codecs as Autoregressive Moving Average Processes and Related Queueing System Performance", *IEEE Journal on Selected Areas in Communications*, vol. **9**, no. 3, pp. 284-292 (April 1991).

[15] F. Yegenoglu, B. Jabbari and Y. Zhang, "Motion Classified Autoregresive Modeling of Variable Bit Rate Video", *IEEE Transactions on Circuits and Systems for Video Technology*, vol. **3**, no. 1, pp. 42-53 (February 1993).

[16] Proakis, "Digital Communications" New York: McGraw-Hill (1983).

[17] S. S. Lam, S. Chow and D. K. Y. Yau, "A Lossless Smoothing Algorithm for Compressed Video", *IEEE/ACM Transactions on Networking*, vol. **4**, no. 5, pp. 697-708 (October 1996).

Estudio del control de tasa de fuente en servicio ABR para aplicaciones de audio y vídeo

XAVIER HESSELBACH SERRA, SEBASTIÀ SALLENT RIBES
DEPARTAMENTO DE MATEMATICA APLICADA Y TELEMÁTICA
ETSETB, UNIVERSIDAD POLITÉCNICA DE CATALUNYA
C/ JORDI GIRONA, 1 Y 3, MÓDULO C3- CAMPUS NORTE, 08034 BARCELONA
Correo electrónico: {xavierh, sallent}@mat.upc.es

Abstract:

The ATM Broadband ATM network permits the transmission of real-time high quality audio and video. Usually, the offered services from network not match the rate generated profile in the user sources. Therefore, efficiency derived from reserved resources is not optimum. Source bit rate adaptation can improve the global throughput. By means of ABR services, user bit rate can be adapted using a feedback loop. Considering N heterogeneous sources, a multiplexion system and a shaping conformance, an adaptive solution is proposed based on bit rate control on every source. This control is developed according to a prediction system based on the source features in order to evaluate the required resources to minimize the cell transfer delay CTD and cell delay variation CDV. The control system is formulated and analyzed in this paper, and fairness and quality of service studied in a simulation platform.

Keywords: ABR, Adaptation, Control, Prediction, Quality of Service, Real-time, Resources, Video.

1. Introducción

El Modo de Transferencia Asíncrono (ATM) es la tecnología escogida como soporte para las Redes Digitales de Servicios Integrados de Banda Ancha (RDSI-BA). Las pequeñas celdas de información de sólo 53 octetos de información fluyen por la red transportando fragmentos de información hacia sus destinos. El tipo de los datos y los requisitos temporales de entrega condicionan el funcionamiento que la red debe ofrecer a los servicios que va a soportar. ATM define diversos tipos de servicio para los cuales los dispositivos multiplexores y conmutadores deben administrar los recursos de acuerdo a las reservas efectuadas.

Los servicios de tiempo real tales como audio o vídeo, requieren la entrega a tiempo de las celdas para que sean conformes a las necesidades de servicio del receptor, puesto que los paquetes excesivamente retrasados perderían su utilidad por caducidad. Consecuentemente, el problema se plantea cuando aparece la necesidad de conseguir que las celdas puedan alcanzar su destino cumpliendo el contrato de la conexión en todos los puntos que atraviese y, en particular, a la entrada de la red en la interfaz usuario-red (UNI) sin ser marcadas como descartables por la función de policía (UPC). Debe existir garantía de un retardo medio de entrega o CTD, para lo cual el tiempo de transferencia hacia la red (espera en colas ante la UNI) debe estar acotado y minimizado. Igualmente para la variación de retardo (CDV). Valores excesivamente grandes de CDV pueden comportar un funcionamiento todavía si cabe más nocivo que el propio CTD a causa de las restricciones temporales a las cuales el receptor se ve sometido. Las celdas deben alcanzar su destino con tiempo suficiente para que todavía resulten útiles a la secuencia temporal a la que pertenecen.

2. Esquema general de multiplexado

El multiplexado de diversos tipos de tráfico facilita la ubicación de recursos en red para su transmisión, gracias a la ganancia obtenida del propio multiplexado. Al mismo tiempo, permite gestionar de forma eficiente el ancho de banda empleado y controlar la calidad de servicio QoS propia de cada fuente tributaria. La transmisión de varios canales de vídeo junto a varios canales de voz asociados a cada uno de ellos puede ser gestionada de forma que ante limitaciones de ancho de banda por congestión en la red pueda aplicarse una disminución de calidad a través de los parámetros de codificación asociados a ellos, evitando el cierre de la transmisión.

La recomendación I.371 del ITU-T [5] ofrece una configuración de referencia para la transmisión de tráfico de N fuentes. Este modelo es demasiado simple para satisfacer las demandas requeridas por el tráfico de vídeo y no contempla el grado de congestión en el que la red se encuentre inmersa, el empleo de servicios específicos ofrecidos por la red o el ajuste de los codificadores de las fuentes. En [3] se propone un esquema realimentado (Fig.1) con capacidad de ajuste dinámico de acuerdo a los recursos disponibles instantáneamente en la red, sea cual fuere el tipo de conexión contratado con ella.

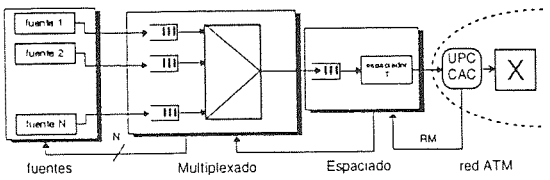


Figura 1

Para servicio ABR, la tasa de transmisión estará situada entre un pico PCR y un mínimo MCR. El valor dependerá del nivel de congestión a la cual se encuentre sometida la red. Este tipo de servicio considera el empleo de celdas de gestión RM que informan de la tasa admisible a través de la UNI. Su funcionamiento se basa en la generación y envío desde la fuente de este tipo de celdas junto al mismo torrente de la conexión ATM, de acuerdo a una proporción establecida, de forma que averigüen el estado de congestión de la red hasta el destino y retornen con información de la tasa disponible a través de todos los conmutadores atravesados en dicha conexión [1]. Cuando las celdas de tipo RM pasan por un conmutador, sus campos pueden ser modificados si las restricciones de dicho dispositivo son superiores a las indicadas. Cuando dicha celda sea recibida en destino, será retornada a la UNI donde se originó para informar del rastreo realizado. En consecuencia, existe un retardo temporal entre la tasa de entrada que el usuario emplea dictada por la realimentación de las celdas RM y la realmente admitida en dicho instante por la red.

El subsistema de espaciado [4] de la figura 1 intentará absorber pequeñas variaciones de la tasa admitida por la red, apoyado por el subsistema de multiplexación. En el caso de que la propia multiplexación - empleando el algoritmo que se pueda considerar adecuado - no pudiera superar por sus medios una disminución de tráfico, se debería solicitar a través de un método reactivo una variación en la tasa individual generada por las fuentes a costa de una inferior calidad, pero siempre evitando la pérdida de celdas por desbordamiento en las colas de espera y sin causar grandes retardos que causaran la validez de algunas de ellas.

Las conexiones ABR fueron diseñadas para aplicaciones sin inconvenientes en disponer de anchos de banda variables en el tiempo, por lo que el servicio ABR es apropiado para usuarios capaces de adaptar su tasa al ancho de banda disponible en cada instante de la transmisión. De hecho, es posible garantizar una tasa de pérdidas CLR nula para aquellas aplicaciones que se adapten de forma conveniente a las variaciones [2].

Las técnicas de multiplexación existentes ofrecen alternativas para el agregado de diversas fuentes de tráfico en las cuales suelen optimizar únicamente uno de los parámetros, en detrimento del resto. Además, tampoco son capaces de recordar el comportamiento desarrollado por una fuente, de forma que les permita administrar los recursos que ésta pudiera requerir en cada momento. Esta información puede ser de vital importancia para alcanzar rendimientos elevados en la multiplexación al tiempo que se acotarían valores bajos de CTD y CDV.

Los algoritmos de Round Robin ponderado o VirtualClock consiguen componer una mezcla procedente de las diversas fuentes de forma proporcional y equitativa a la tasa de los diversos flujos. Con ello, el CDV tiende a valores mínimos y se consigue acotar el CTD causado por el multiplexor. Sin embargo, en base a las restricciones temporales de las diversas fuentes y a la deseada pluralidad de las fuentes, la propuesta del Golden Ratio [9] generaliza el algoritmo de Round Robin para el caso de fuentes heterogéneas, consiguiendo un mejor espaciado de celdas consecutivas y, por consiguiente, menor CDV.

3. Algoritmo de control

El control de la tasa de las fuentes permite la reducción de la tasa de pérdidas en el sistema tanto como las fuentes sean capaces de adaptarse a las variaciones que se les indica. Cuando el multiplexor no pueda absorber las alteraciones en la tasa admitida por la red, el mecanismo de control deberá activarse. Un mecanismo efectivo evitará la acumulación de unidades en las colas de espera lo cual se traducirá en el minimizado de los parámetros de retardo (CTD y CDV) y en la optimización de la calidad relativa a los recursos disponibles en red.

3.1 Diseño básico

La monitorización del tráfico generado por cada fuente ofrece las directivas para la aplicación de un mecanismo reactivo que obligue a las fuentes a la variación de su tasa media medida en ventanas de tiempo de período T. La predicción de la tasa entregada por la fuente resulta esencial para la prevención de congestión en el sistema. La evaluación de celdas excesivas permite establecer una corrección en la tasa entregada por la fuente que evita la acumulación de unidades generadas esperando a ser multiplexadas. El comportamiento del predictor es regulado por el estado de la conexión contratada a la red. La repartición del ancho de banda disponible se efectúa en base a la tasa nominal a la cual operan cada una de las fuentes.

La figura 2 ilustra el mecanismo de realimentación inducido por el sistema de control

de la tasa de las fuentes. El sistema de cálculo P permite la distribución de los recursos proporcionalmente a las fuentes en función de las condiciones instantáneas.

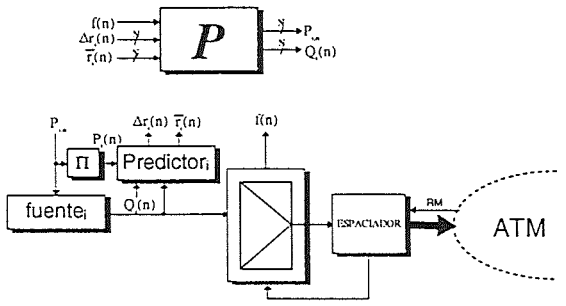


Figura 2

3.2 Parámetros fundamentales

Las fuentes operan a una tasa nominal de r_{oi} bits/seg. Cada periodo T, se establece un nuevo valor $P_{i,n}$ para la fuente i, en el instante n, que multiplica a la tasa entregada por esta fuente. Definiendo la tasa de penalización acumulada como

$$P_i(n) = P_{i,1} \cdot P_{i,2} \cdot \ldots \cdot P_{i,n} = \prod_{j=1}^{n} P_{i,j} \qquad (1)$$

entonces, la tasa transmitida por la fuente en el n-ésimo periodo será $r_i(n) = r_{oi} \cdot P_i(n)$. En el supuesto de una fuente que jamás viera alterada su tasa desde el sistema de control, $P_i(n)$ se mantendría a 1 y, en consecuencia, su tasa de transmisión seguiría a r_{oi}, su valor nominal.

$\bar{r}_i(n)$ corresponde a la tasa media generada por la fuente i-ésima en el instante n-ésimo, y se deriva de la monitorización que efectúa el predictor.

El parámetro f(n) refleja la relación entre la tasa disponible actual y la anterior tasa disponible. Es decir, es el factor de pérdida de ancho de banda.

$Q_i(n)$ sintetiza la calidad que va a controlar el predictor. En general, incluye la velocidad de transmisión admitida de acuerdo al funcionamiento del predictor, aunque el contenido exacto de $Q_i(n)$ depende del algoritmo definido en este subsistema. Por ejemplo, si el mecanismo de predicción fuese del tipo GCRA(I,L), $Q_i(n)$ se referiría al par (I,L).

El susbsistema predictor determina el número de celdas en exceso en el actual periodo. Definiendo $d_i(n)$ como la cantidad de unidades de transmisión (celdas ATM) no conformes o en exceso generadas por la fuente i-ésima en el n-ésimo periodo T, se deriva el indicador de exceso de tasa correpondiente, $\Delta r_i(n)$, como

$$\Delta r_i(n) = \frac{d_i(n) \cdot 53 \cdot 8}{T} \text{ bit/s.} \qquad (2)$$

3.3 Objetivos

Definidos los anteriores parámetros básicos del sistema, el sistema de control debe encargarse de repartir los recursos de acuerdo al comportamiento presente y pasado de las fuentes, penalizando a aquellas que en el transcurso de la historia hayan ejercido mayor perjuicio o presión sobre el sistema de multiplexación. En suma, las condiciones bajo las cuales debe operar pueden sintetizarse en las siguientes premisas básicas de repartición, ordenadas según importancia del criterio:

Cuando la tasa de transmisión deba reducirse,

i) eliminar el exceso debido a $\Delta r_i(n)$ de las fuentes.

ii) Penalizar a las fuentes que hasta el momento hayan sido menos sancionadas, de forma que todas ellas permanezcan igualmente cerca de su tasa nominal.

iii) El ancho de banda que todavía falte por ceder, deberá ser repartido proporcionalmente a la tasa nominal entre las diversas fuentes.

En el supuesto de que existan recursos en red para permitir un incremento en la calidad de la transmisión, la repartición se efectuará intentando favorecer a las fuentes previamente más perjudicadas. Los tres principios siguientes enumeran la prioridad con que debe aplicarse:

i) Permitir el paso de los posibles excesos $\Delta r_i(n)$ generados en las fuentes.

ii) Favorecer a las fuentes que en base a $P_i(n)$ hayan recibido mayor corrección.

iii) Repartir proporcionalmente el ancho de banda restante entre todas las fuentes.

Estos tres puntos son implementados a través del parámetro $Q_i(n)$ actuando sobre el mecanismo de predicción, de forma que, en la siguiente iteración, la fuente pueda transmitir a

mayor velocidad sin que por ello genere celdas en exceso.

3.4 Formulación

Se define el coeficiente de distorsión como la tasa en exceso frente a la tasa nominal de cada fuente,

$$c_i(n) = \frac{\Delta r_i(n)}{r_{oi}} \qquad (3)$$

y el coeficiente normalizado como

$$c_{n_i}(n) = \frac{c_i(n)}{\sum_{i=1}^{N} c_i(n)} \qquad (4)$$

En el caso de una fuente con tasa de transmisión inferior a la esperada por el predictor, $\Delta r_i(n)$ podría adoptar un valor negativo. En este caso, cabe esperar que en el siguiente periodo la fuente trate de mantener su tasa esperada, por lo cual este hecho no hará modificar el comportamiento del sistema de predicción, aunque si bien el ancho de banda no empleado podrá ser empleado por otras fuentes que sí hubieran generado más celdas de las esperadas, de acuerdo a la premisa i) de incremento de tasa de transmisión citada anteriormente.

En consecuencia,

si $c_i(n)<0 \rightarrow c_i(n) = 0$.

La tasa agregada total en exceso queda definida como

$$\Delta r_a(n) = \sum_{i=1}^{N} \Delta r_i(n) \qquad (5)$$

El ajuste de la tasa de las fuentes se efectuará de acuerdo al siguiente criterio:

$$\bar{r}_i(n) \rightarrow \bar{r}_i(n) - C_{ni}(n) \cdot \Delta r_a(n) \qquad (6)$$

y dado que $P_{i,n} = r_i(n) / r_i(n-1)$, resulta que:

Cuando $f(n) < 1$:

$$P_{i,n} = 1 - \frac{C_{n_i}(n)}{\bar{r}_i(n-1)} \cdot \Delta r_a(n) \qquad (7)$$

y, en caso contrario, definiendo Δr_{exceso} la diferencia entre la antigua tasa disponible en la red menos la nueva, incluyendo además el ancho de banda no empleado por las fuentes en dicho periodo (Notar que Δr_{exceso} será un valor negativo cuando aumenten los recursos en la conexión contratada):

$$P_{i,n} = 1 - \frac{\Delta r_{exceso} \cdot r_{oi} \Big/ \sum_{j=1}^{N} r_{oj}}{\bar{r}_i(n-1)} \qquad (8)$$

3.5 Cálculo de los coeficientes $Q_i(n)$

Los coeficientes $Q_i(n)$ son empleados como parámetro de calidad o tasa a admitir según los algoritmos de predicción. Por este motivo, las variaciones en la tasa disponible en la conexión contratada a la red se verán inmediatamente reflejadas en estos coeficientes de calidad.

La Fig.3 muestra el comportamiento esencial: A través de $Q_i(n)$ el predictor establece si ha habido exceso de tasa transmitida, lo cual induciría la inmediata reducción de tráfico, de acuerdo al comportamiento global experimentado por todas las fuentes pertenecientes a la misma agrupación de multiplexación.

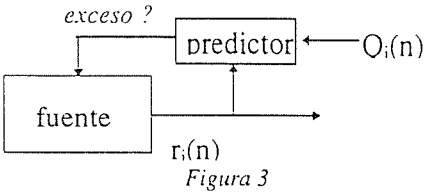


Figura 3

El ajuste de los coeficientes de calidad se realizará mediante la distribución de los recursos disponibles en la conexión de red proporcionalmente a las tasas nominales de los diversos generadores frente al total.

4. Respuesta del sistema de control

El algoritmo propuesto para el control de la tasa de las fuentes debe ser suficientemente rápido para que, de acuerdo a los requisitos de la sección 3.3, consiga adaptar la tasa de las fuentes sin causar la acumulación de unidades en las colas de espera. Por medio de simulación, se evalúan los principales parámetros de respuesta del mecanismo.

4.1 Objetivo

Con el fin de validar los requisitos requeridos, se estudia el tiempo de respuesta del algoritmo. Si la respuesta a variaciones resultaran excesivamente largas, el remanente de celdas generadas o el vacío de unidades por transmitir se traduciría en pérdida de eficiencia en el empleo de los recursos disponibles en la red.

Por otro lado, la simulación debe aportar resultados sobre el comportamiento de la respuesta en orden a determinar la estabilidad frente a variaciones continuas de tasa a la que las fuentes se pudieran ver sometidas, y a la justicia o *fairness* con que opera el sistema de control.

Como se ha visto en (5), puede existir una tasa agregada en exceso, que debe ser minimizada. Se va a demostrar que esta tasa en exceso sólo se manifestará durante los periodos en que se sufran cambios de velocidad de transmisión admitida en la red, de forma que la tasa de pérdidas y los retardos se podrán mantener tan bajos como se desee si las fuentes son capaces de adaptar la nueva velocidad tan pronto se les indica la variación. En suma, se trata de averiguar la capacidad de seguimiento del tránsito cursado frente al previsto, el cual sí predecirá de acuerdo a un ajuste perfecto al ancho de banda disponible en la conexión ABR contratada.

4.2 Esquema de simulación

La figura 4 muestra la relación entre los bloques constituyentes del mecanismo de control implementado en la simulación.

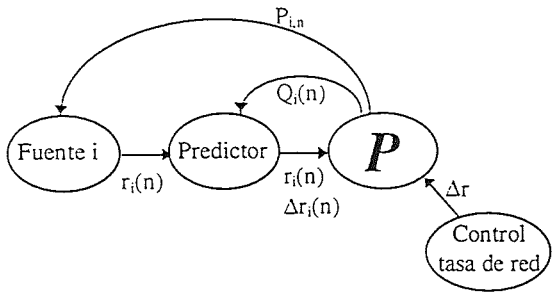


Figura 4

El subsistema de cómputo de parámetros *P* evalúa tanto los nuevos valores de calidad $Q_i(n)$ hacia cada predictor como la variación de tasa $P_{i,n}$ que se aplicará sobre cada fuente.

4.3 Condiciones de simulación

Se van a considerar dos tipos básicos de fuentes y dos tipos de comportamiento de la tasa disponible de red: De tasa constante, y de tasa variable. Las tasas variables serán modeladas por medio de perfiles senoidales, ya que ofrecen variación simétrica frente a un valor medio y permiten cubrir todas las posibles pendientes de respuesta.

Para todas las simulaciones, el caudal medio de la conexión será el de la tasa nominal agregada de todas las fuentes.

En todas las gráficas que serán presentadas, el eje de ordenadas corresponde a la tasa y el de abscisas al #periodo (eje de tiempos).

4.4 Estudio frente pérdida impulsiva de tasa en la red.

Se va a estudiar el comportamiento del sistema de control ante variaciones impulsivas de tasa en la red. Estas condiciones son las de mayor dificultad para conseguir la adaptación en el sistema de control, y realmente no se espera que las conexiones ABR en redes ATM ofrezcan este perfil de tasa. En realidad, cabe esperar variaciones suaves a las cuales, como se demostrará, el bloque de adaptación consigue una elevada eficiencia en el empleo de recursos.

Se considera la existencia de un escalón con una reducción de la tasa disponible en la red hasta el 1% del valor inicial, para recuperar al cabo de 10 periodos. Se obtienen los siguientes resultados:

La figura 5 muestra como, para diverso número de fuentes, y con tasa nominal disponible inicialmente en la red igual al caudal total generado por el número N de fuentes en cada curva, la relación entre la tasa disponible en la red y la tasa entregada por las fuentes están retardadas en un único periodo, pero consigue igualarse el valor deseado tras un transitorio de duración un periodo, de forma que, sea cual fuere el número de fuentes, el sistema consigue corregir sus tasas en un solo periodo hasta lograr que se emplee el 100% de los recursos reservados en la red.

La figura 6 refiere el comportamiento individual seguido por cualquiera de las fuentes, con cualquier número de ellas en el sistema. Por simulación se comprueba que el número de fuentes no influye en la velocidad de respuesta ni en el tipo de transitorio resultante. Todas estas fuentes sufren también por igual en los instantes de transición. Los instantes en los cuales la tasa aumenta son en los que existe mayor peligro de desbordamiento. La evolución de la nueva tasa en red vista por el sistema de control puede llevarle a confundir y conceder un incremento de tráfico superior al que finalmente queda establecido en la conexión con la red.

La tasa generada en exceso se manifiesta en la figura 7. Estos excesos deben ser vencidos mediante colas que permitan almacenar las celdas que acumuladas durante el periodo de transición. Las variaciones bruscas en la tasa de red no pueden ser absorbidas por el mecanismo de control de la tasa de fuente, y únicamente una memoria tampón puede evitar la pérdida de celdas, al precio de

acumular un cierto retardo. Por contra, los incrementos bruscos de tráfico admitidos en la red pueden ser causa de vacíos de transmisión.

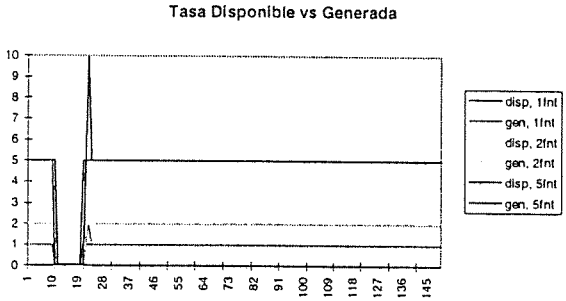


Figura 5

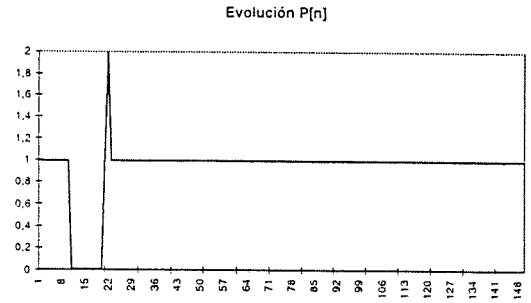


Figura 6

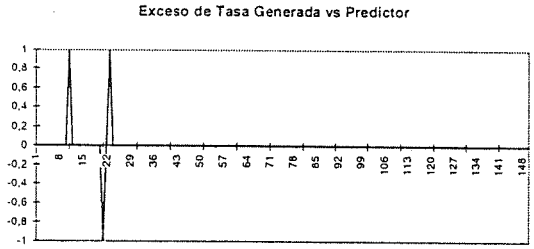


Figura 7

4.5 Estudio frente tasa variable en la red

La respuesta del sistema de control ante variaciones en la tasa de red constituye el principal desafío del sistema, con el fin de repartir equitativamente los recursos disponibles. Se considera en este caso fuentes de tasa constante sometidas a una conexión de red de tasa variable con característica senoidal. De forma parecida al apartado anterior, se obtienen las respuestas que aparecen en las figuras 8,9 y 10. Debido al retraso de un periodo, la tres figuras ponen de manifiesto que aunque el seguimiento de la tasa de efectúa para cualquier número de fuentes (fig.8) y todas ellas con igual comportamiento (mostrado en la fig.9), la fig.10 pone de relieve la existencia de una pequeña parte de tráfico que no sigue. Nótese que los generadores siempre entregan tráfico ligeramente por encima del admitido por la red. De esta forma se asegura el aprovechamiento de recursos. El caso peor se produce en el punto de inflexión de la senoide de la tasa de red, que coincide con el punto de mayor pendiente.

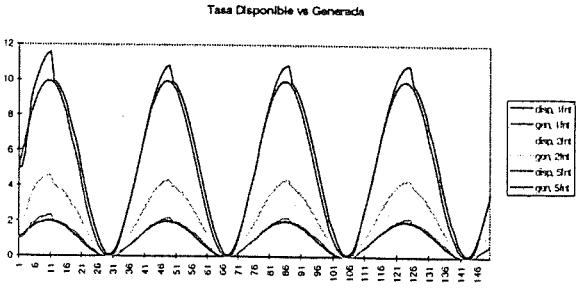


Figura 8

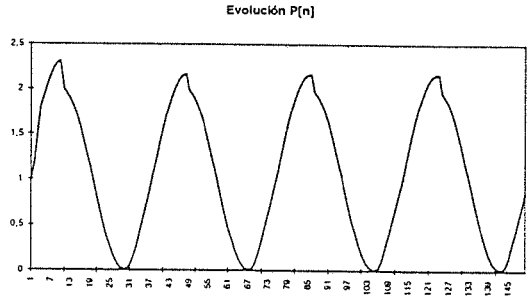


Figura 9

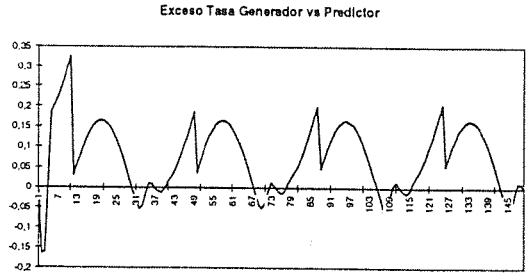


Figura 10

En cualquier caso, el peor seguimiento, se sitúa entorno a una tasa de exceso de 16,48 % de la tasa de la fuente, según valores de simulación, que es la que deberá ser salvada gracias al empleo de colas de espera ante el multiplexor. Consecuentemente, cabe esperar que los buffers deban contener hasta un exceso máximo del 16,48% de la tasa transmitida por la fuente. Este es, pues, un parámetro de diseño de las memorias de entrada al multiplexor.

4.6 Estudio frente tasa variable de las fuentes

Hasta el momento se ha establecido la neutralidad del mecanismo ante fuentes de idénticas características. Las fuentes de tasa variable se caracterizarán por la tasa media transmitida y el predictor empleado deberá estar de acuerdo a las características de la fuente. Por este motivo, el mejor predictor será una fuente sintética

de las mismas características que la fuente del usuario. La tasa de cada fuente será el ordinal que le corresponda. Así, de 5 fuentes, la primera transmitirá a tasa 1 y la quinta a tasa 5. Esto es aplicable a los próximos casos con tasa variable de fuente que serán estudiados en las siguientes secciones.

En cuanto a la red, en este apartado se empleará una conexión de red con tasa constante.

Considérese la situación con predictores no adaptados a las características de las fuentes. Usando los mismos predictores de ventana que en las anteriores simulaciones se obtiene los resultados que a continuación se presentan.

Se verifica que todos los coeficientes Pi[n] en cada instante n son prácticamente iguales para todas las fuentes (figura 12), para cualquier número de fuentes (las simulaciones fueron efectuadas para 1, 2 y 5 fuentes). En particular, para el caso de 5 fuentes, la media de las varianzas para cada realización n entre los Pi[n] de estas 5 fuentes presenta un valor de 0,0011, con una media de $P_i[n]$ de 1,1495. Estos es, el reparto del ancho de banda se realiza de manera prácticamente proporcional a la tasa de transmisión nominal de cada fuente.

Las figuras 11 y 13 muestran la consecuencia inmediata: Fuentes con mayor tasa nominal (por ejemplo, vídeo) producen mayor cantidad de celdas en exceso que las de menor tasa nominal (como el audio). Este es justamente el objetivo del control de tasa de las fuentes, ya que aquellos generadores con baja tasa no deben sufrir los problemas de ancho de banda de otros. En suma, se asegura el control de la velocidad de transmisión al tiempo que se garantiza la independencia entre fuentes y se evitan efectos de vecindad por el hecho de pertenecer a un mismo multiplexor.

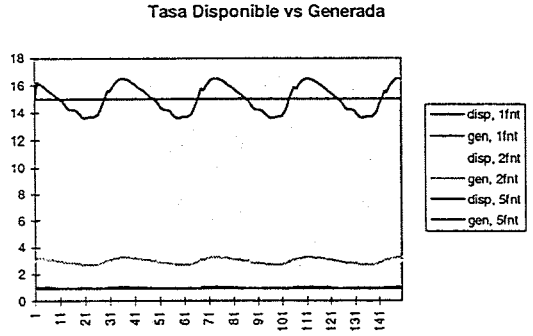


Figura 11

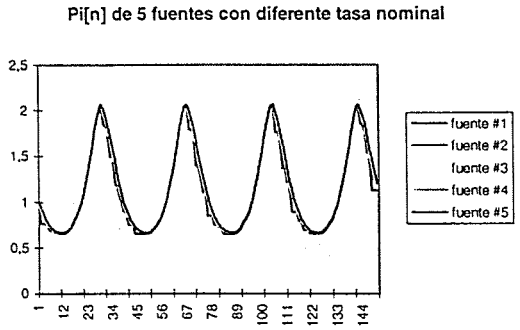


Figura 12

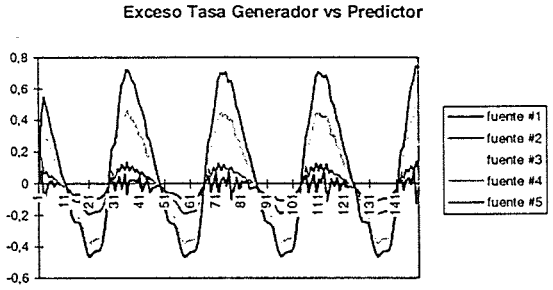


Figura 13

Consideramos a continuación las mismas fuentes pero empleando predictores con las mismas características que las fuentes, predictores adaptados a los generadores. En este caso, la simulación concede un interesante resultado: Con mismo comportamiento de la tasa de la red, y con las mismas fuentes que para el caso de las figuras 11 a 13, la figura 14 muestra como en este caso las fuentes siguen perfectamente el perfil senoidal sin verse sometidas a restricciones causadas por el predictor. Ello queda justificado y refrendado por los valores de $p_i[n]$ y de celdas en exceso encontrados en la simulación, que son estrictamente 1 y 0, respectivamente. Consecuentemente, se demuestra que el algoritmo de control opera tan idealmente como el predictor sea capaz de predecir la tasa del generador asociado.

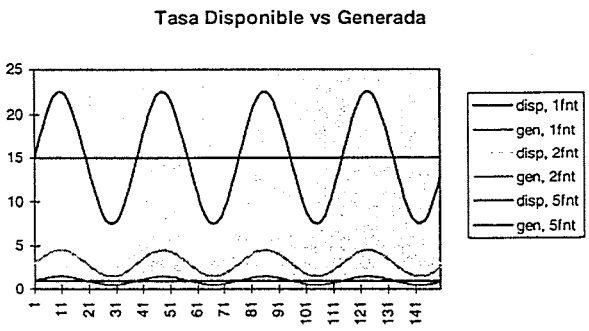


Figura 14

4.7 Estudio frente a tasa variable en las fuentes y en la red

Se añade a continuación una nueva variable al estudio: La variación de la tasa disponible en la conexión ABR contratada a la red, con perfil senoidal. De ello se obtienen los resultados de las figuras 15 y 16.

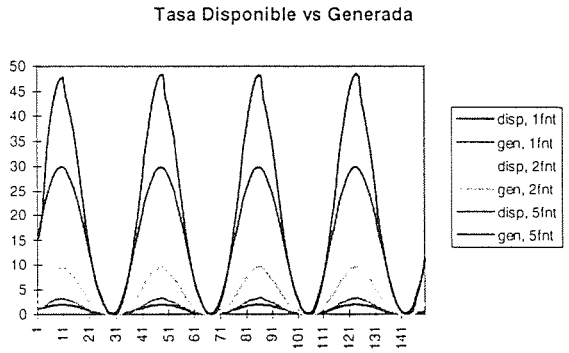


Figura 15

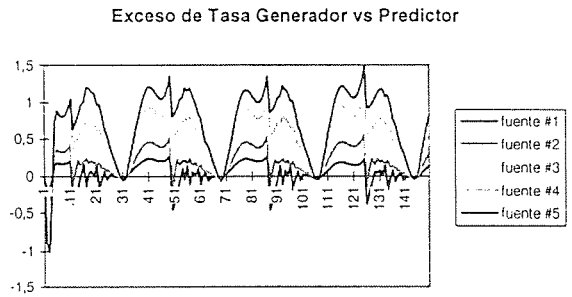


Figura 16

Al igual que en casos anteriores, se observa aquí que también el sistema es capaz de ajustar la tasa de las fuentes al ancho de banda disponible en la conexión ABR contratada a la red. La figura 15 es comparable a la 8, añadiendo que en el caso que aquí no ocupa se ha conseguido con tasa variable de red. Por otro lado, cabe observar que el exceso de tasa generada es esencialmente positiva, lo cual reafirma las consecuencias obtenidas de la figura 10, su homónima para el caso de tasa constante en la red.

4.8 Estudio frente tasa variable en fuentes no sincronizadas y red

Hasta ahora, todas las fuentes empleadas han presentado el mismo perfil de crecimiento en los mismos instantes. Con el fin de validar el funcionamiento en el caso más general, se finaliza el estudio con fuentes de diversa tasa media cuyo perfil varía senoidalmente con distintos instantes donde se alcanzan los máximos.

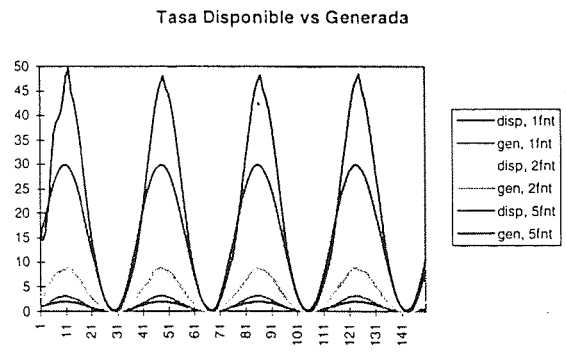


Figura 17

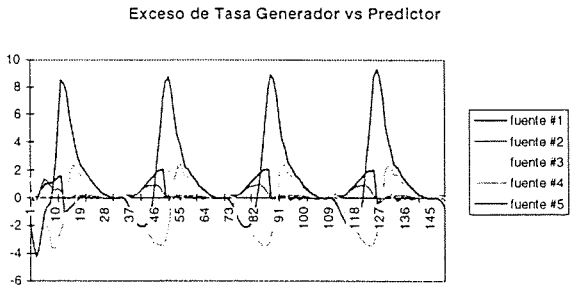


Figura 18

Aparentemente la figura 17 muestra todos los máximos en el mismo instante. Realmente, dichos máximos reflejan el máximo del ancho de banda admitido por la red, y por ello, las fuentes se benefician de ello con aumento de su calidad de transmisión. Por lo tanto, este esquema pone de relieve que las fuentes deberán variar su tasa de transmisión rápidamente si desean poder aprovechar el aumento de calidad que la red les autoriza. Sin embargo, por contrapartida, cuando la tasa de red decrece, la fuente puede exceder la tasa generada, como aparece en la figura 18, aunque como se observa, es directamente proporcional a la tasa nominal. Por ejemplo, la fuente #5 presenta mayor exceso porque transmite con velocidad 5 veces mayor que la fuente #1.

5. Conclusiones

En esta ponencia se ha presentado un mecanismo de alisado y multiplexación adecuado para tráfico de audio y vídeo controlado por un mecanismo de adaptación de la tasa de fuente. Se ha formulado y analizado su comportamiento, y se han extraído las siguientes conclusiones principales:

El algoritmo de adaptación asegura el control de la velocidad de transmisión al tiempo que se garantiza la independencia entre fuentes para evitar que puedan afectarse mútuamente.

Se demuestra que el algoritmo de control opera tan idealmente como el bloque de predicción sea capaz de pronosticar la tasa del generador al cual está asociado.

Se comprueba que el número de fuentes no influye en la velocidad de respuesta ni en el tipo de transitorio resultante, por lo cual este mecanismo resulta adecuado para entornos en los cuales se multiplexan pocas fuentes y donde es justamente más difícil conseguir alcanzar buenos niveles de eficiencia de los recursos reservados en la conexión contratada de la red.

Referencias

[1] F.Bonomi, K.W.Fendick. "The rate-based flow control framework for the Available bit rate ATM service". *IEEE network.* **9**, 2, 25-39 (1995).

[2] T. M. Chen, S. S. Liu, V. K. Samalam. "The Available Bit Rate Service for Data in ATM Networks". *IEEE Communications Magazine.* **34**, 5, 56-71 (1996).

[3] X.Hesselbach, S.Sallent, A.Barba. "Estudio de un alisador de Tráfico multiplexado de vídeo síncrono en ATM bajo servicio ABR". *Telecom I+D.* 307-312 (1996).

[4] F. Hübner, P. Tran-Gia. "Discrete-time analysis of cell spacing in ATM systems". *Telecommunication Systems3*, 379-395 (1995).

[5] ITU-T. "Traffic Control and Congestion Control in B-ISDN". *Recomendación I.371 del ITU-T* (1993).

[6] J.Mata. "Contribución a la gestión dinámica de recursos aplicada al control de fuente de vídeo de velocidad variable en la RDSI-BA". *Tesis doctoral.* DMAT UPC (1996)

[7] C.G. Omidyar, G.Pujolle. "Introduction to Flow and Congestion Control". *IEEE Communications Magazine.* **34**, 11, 30-32 (1996).

[8] K.K.Ramakrishan, P.Newman. "Integration of Rate and Credit Schemes for ATM Flow Control". *IEEE Network.* **9**, 2, 49-56 (1995).

[9] Z.Rosberg. "Cell Multiplexing in ATM Networks". *IEEE/ACM Transactions on Networking.* **4**, 1, 112-122 (1996).

Simulación y Modelos Analíticos en el Análisis del Tráfico de Voz en Redes BISDN

OLABE M.A., FERRO A., ESPINOSA K., OLABE X.*
Departamento de Electrónica y Telecomunicaciones - Grupo de Ingeniería Telemática
*Departamento de Ingeniería de Sistemas y Automática
E.T.S. de Ingenieros Industriales y de Ingenieros de Telecomunicación de Bilbao
Universidad del País Vasco-Euskal Herriko Unibertsitatea
C/ Alameda de Urquijo s/n, 48013 Bilbao
e-mail: {jtpolbam,jtpfevaa,jtpesacj,jtpolbax@bi.ehu.es}

Abstract:

This paper reviews the use of analytical and simulation methodologies for the performance evaluation of packetized voice transmission. An ATM switch statistically multiplexing voice channels is studied using five methodologies, including: two analytical models based on Markov modulated processes: a fluid flow analytical model; a standard simulation technique; and a discrete event simulation methodology with statistical inference.

1. Introducción

ATM (Asynchronous Transfer Mode) ha sido adoptada por la Unión Internacional de Telecomunicaciones como el modo de transferencia que permitirá la integración deseada de los diferentes tipos de tráfico soportados por B-ISDN (Broadband Integrated Services Digital Network). ATM utiliza los recientes avances de la tecnología VLSI y posee además la capacidad de optimizar la gestión de redes de conmutación de paquetes mediante la multiplexación estadística de una gran variedad de tráficos, como vídeo en demanda, vídeo-conferencia, transferencia de ficheros, y voz en paquetes, entre otros.

El proceso de multiplexación estadística está basado en la combinación de una serie de fuentes de tráfico cuya demanda es variable. Si la información que se desea transmitir varía en el tiempo, con valles y picos, y si además es posible combinar varias fuentes de tráfico de tal modo que los picos de unas coincidan con los valles de otras, entonces es posible transmitir simultáneamente más llamadas que si se utilizase como criterio de control de la red, la suma de las demandas máximas de las llamadas.

En el caso de tráfico de voz, antes de transmitir la señal por una red de paquetes se le puede someter a dos procesos de reducción de demanda: utilizando técnicas de compresión, y mediante la detección de actividad de la señal de voz (SAD, Speech Activity Detection). Estos procesos de reducción de demanda sin embargo convierten una señal caracterizada por una demanda alta y constante, en otra de demanda baja y variable ,en general desconocida y aleatoria.

La naturaleza aleatoria de la demanda de transmisión de voz por paquetes y las demandas de calidad de servicio establecidas para la redes ATM hacen que el problema de multiplexación de señales de voz no sea un problema trivial.

En este artículo se estudia el uso de métodos analíticos y de simulación para la gestión de redes que transmiten voz por paquetes. Se comparan dos métodos basados en cadenas de Markov (SMP y CTMC), un método analítico de fluidos [1], y un método de simulación e inferencia estadística descrito en [2]. Así mismo se incluyen los resultados de un método de simulación convencional.

El resto del artículo está organizado del siguiente modo: La sección 2 describe modelos probabilísticos de voz en paquetes con detección de actividad de señal [3], [4] y [5]; la sección 3 describe brevemente el método de simulación e inferencia estadística [2] y su aplicación al estudio de transmisión de voz por paquetes; en la sección 4 se presentan dos modelos probabilísticos de multiplexación basados en cadenas de Markov [6] y [7]; en la sección 5 se presenta el método analítico basado en flujo de fluidos contínuos con llegada y servicio uniforme [8]; y en la sección 6 se presentan resultados numéricos para la comparación cuantitativa de los diferentes métodos. El artículo concluye con un conjunto de conclusiones en la sección 7.

2. Modelo de Voz por Paquetes con Detección de Actividad

La gran proliferación de redes de comunicación basadas en conmutación de paquetes ha intensificado el interés y la actividad investigadora centrada en la transmisión de voz mediante paquetes. Un procedimiento clásico e inmediato para la reducción de la demanda de transmisión de voz es la detección de actividad de señal. La señal de voz ha sido en general modelada como una secuencia alternativa de fragmentos activos, o señal de voz propiamente dicha, y fragmentos inactivos o de silencio [3]. Diversos estudios han caracterizado las duraciones promedio de estos fragmentos con rangos que abarcan desde

0.4 a 1.2 segundos para la duración de las zonas activas, y de 0.6 a 1.8 para las zonas de silencio [4]. En los estudios de simulación presentados al final de este trabajo se han utilizado duraciones de 1.35 y 1.65 segundos como promedio de las zonas activas y de silencio respectivamente [5], para comparar el método de simulación e inferencia con los estudios presentados en [1].

Tanto la zona activa como la de silencio utilizadas para modelar la señal de voz pueden ser aproximadas mediante distribuciones exponenciales. La zona activa queda bien definida por este tipo de distribución. Las zonas de silencio no responden con tanta fidelidad al modelo exponencial, sin embargo este modelo es aceptado en un compromiso entre simplicidad de modelado y precisión.

Aunque se han realizado diversos análisis para evaluar el efecto del valor numérico de los promedios de las zonas activas y de silencio en el comportamiento de una red conmutada de paquetes, se reconoce la necesidad de evaluar el efecto de utilizar otras distribuciones no exponenciales.

Además de los estudios realizados para modelar voz con detección de actividad de señal, es preciso extender este trabajo a la aplicación conjunta de la detección de actividad así como al empleo de algoritmos de compresión, y su efecto en el modelado probabilístico de la voz en paquetes.

3. Simulación e Inferencia Estadística

Una descripción más detallada del método de simulación con inferencia estadística puede encontrarse en [2]. Este método está basado en el estudio de una variable, en este caso la pérdida de un paquete debido a la saturación del buffer, mediante el estudio de la causa que lo produce, el nivel de ocupación del buffer. Si se pudiese obtener un modelo válido del nivel de ocupación del buffer, sería posible entonces, utilizando métodos de extrapolación estadística, encontrar un modelo de la probabilidad de pérdida de paquetes debido a la saturación del buffer.

Utilizando un programa de simulación como COMNET III, es posible crear un modelo de una sencilla red ATM para la transmisión de tráfico con períodos activos de distribución exponencial intercalados con períodos de silencio con distribución también exponencial. Analíticamente esta red puede modelarse como un sistema M/M/1. Un sistema real con un buffer finito se caracteriza por una probabilidad de bloqueo, o de saturación del buffer, debido al carácter aleatorio del tráfico. Esta probabilidad será función del tamaño del buffer, de la intensidad de tráfico, así como de las características aleatorias del tráfico. Durante el proceso de simulación es posible registrar el nivel instantáneo de ocupación del buffer. La figura 1 muestra un segmento temporal del nivel de ocupación del buffer registrado durante una simulación.

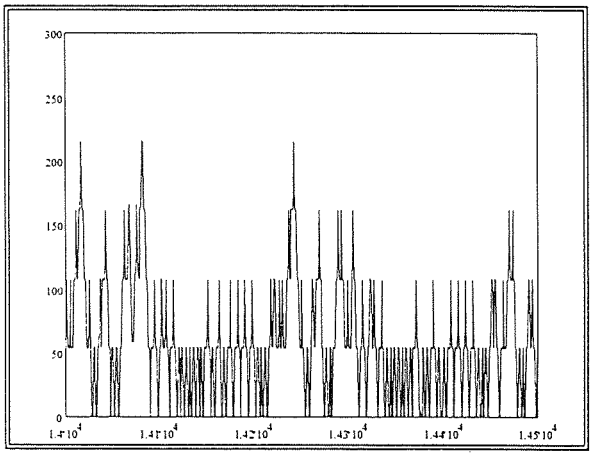


Figura 1. Nivel de Ocupación del Buffer durante un Período de Simulación

El proceso de modelar estadísticamente el nivel de ocupación del buffer está basado en la comparación de dos histogramas, el obtenido directamente durante el proceso de simulación, y el derivado de la función densidad de probabilidad del modelo candidato. La figura 2 muestra la probabilidad de ocupación del buffer utilizando veinte mil valores instantáneos de ocupación de éste durante la simulación. En esta figura se puede apreciar la evolución del histograma, y estimar los valores que la limitación temporal de la simulación no alcanza a reflejar.

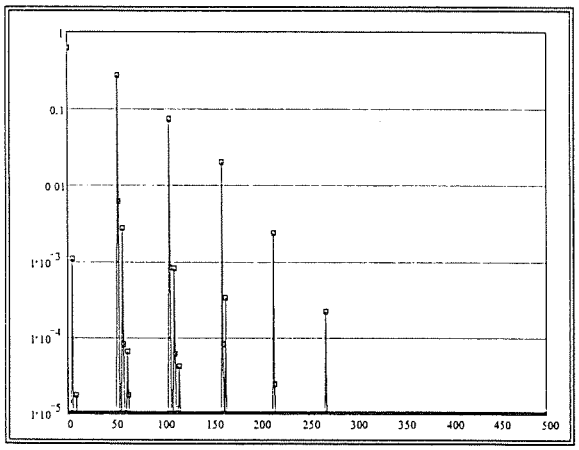


Figura 2. Probabilidad de Ocupación del Buffer Utilizando Resultados de Simulación

Mediante simple inspección del histograma de datos es posible seleccionar una o varias familias de distribuciones como modelos candidatas. Para cada distribución candidata es preciso calcular los parámetros de la función densidad de probabilidad. Este proceso se puede realizar mediante la resolución de un sistema de ecuaciones si este existe. En caso contrario es preciso realizar este proceso de parametrización mediante métodos numéricos.

La fase final del proceso consiste en la evaluación objetiva de la calidad del modelo obtenido. Este proceso de validación estadística consiste en medir la diferencia entre el histograma obtenido en la simulación y el histograma obtenido del modelo. Utilizando métodos estadísticos es posible estimar si la diferencia entre ambos histogramas corresponde a la diferencia esperada entre dos procesos estocásticos idénticos, en cuyo caso el modelo será validado, o si por el contrario la diferencia indica que los resultados de simulación y el modelo son procesos distintos, en cuyo caso es necesario rechazar el modelo, y repetir el proceso con una nueva distribución candidata.

4. Modelos de Markov Modulados

Para el estudio del comportamiento de redes ATM se han propuesto varios métodos analíticos con especial énfasis en la estimación de eventos con muy baja probabilidad. Estos eventos, denominados eventos raros, son característicos de sistemas como las redes ATM cuyos parámetros de calidad de servicio son del orden de uno en un millón o incluso inferiores. Estas metodologías incluyen la teoría de grandes desviaciones [9], la estimación de entropía termodinámica [10], modelos de Markov [4], y otros.

En esta sección se referencian dos modelos analíticos para la estimación de la distribución de la longitud de la cola en un multiplexador de voz por paquetes basado en cadenas de Markov. En ambos sistemas se tiene en cuenta la naturaleza independiente del proceso alternado de voz y silencio de cada uno de los canales de voz multiplexados.

Un tráfico multiplexado con N canales independientes de voz, cada uno representado por un modelo como el descrito en la sección precedente, produce un tráfico combinado que puede ser modelado por un sistema con N+1 estados. Cada estado representa el número de canales de voz activos en un momento determinado (desde un mínimo de cero canales, hasta un máximo de N canales).

El proceso de llegada de paquetes durante el período activo de cada canal y la naturaleza del proceso de servicio diferencian los dos modelos mencionados en esta sección.

En el primer modelo, SMP (Semi-Markov Model), descrito en detalle en [6], la variación de la longitud de la cola en el buffer es proporcional a la diferencia entre la capacidad del canal de transmisión y el número de canales activos. Es posible agrupar los N+1 estados mencionados anteriormente en dos conjuntos: el primero contiene los estados para los cuales la capacidad de transmisión es mayor que la demanda de tráfico; el segundo grupo contiene el resto de estados. En este modelo la generación de paquetes de un canal de voz responde a un proceso de tasa constante. El modelo resultante es un proceso Semi-Markov. Las probabilidades de transición entre estados en este modelo pueden ser obtenidas utilizando el método propuesto en [7], y la formulación dada en [6].

El segundo modelo analítico basado en cadenas de Markov es el CTMC (Continuous Time Markov Chain) [4] y [6]. Al igual que en el modelo anterior cada canal de voz se activa de acuerdo a un proceso de renovación alternado con duraciones exponenciales tanto para las zonas activas como para las de silencio. En este modelo CTMC la generación de paquetes de un canal de voz responde a un proceso Poisson y no de tasa constante como en el modelo anterior.

El método para definir las matrices de probabilidad de transición entre estados así como el procedimiento para su resolución numérica están descritos en detalle en [7] y [11].

5. Modelo de Llegada y Servicio Uniforme

Esta sección presenta el modelo analítico UAS (Uniform Arrival and Service) que fue específicamente desarrollado para la estimación de la probabilidad de bloqueo en redes de paquetes transmitiendo canales de voz con detección de actividad [8].

Este modelo de llegada y servicio uniforme, considera un multiplexador como si fuese un sistema físico en el que el buffer recibe mensajes de un número finito de canales de voz estadísticamente independientes e idénticos. Cada fuente de tráfico alterna asíncronamente entre dos períodos con distribución exponencial de actividad y de silencio. Durante los períodos activos cada fuente genera paquetes con una tasa constante. Al mismo tiempo el buffer se vacía a través del canal de transmisión a la velocidad máxima mientras haya paquetes esperando ser transmitidos.

La formulación de este modelo describe el equilibrio del estado del buffer mediante un sistema de ecuaciones diferenciales. La resolución de este sistema puede expresarse en forma cerrada. El comportamiento asintótico de la ocupación del buffer puede ser obtenido mediante el estudio de la componente exponencial de la fórmula final que posea el mayor exponente. Esta característica es de especial interés para la estimación de eventos con muy baja probabilidad.

Este modelo UAS asume que el número de paquetes generados durante un período de actividad es suficientemente grande como para poder aproximarse a un flujo continuo en el buffer de entrada del multiplexor. En este sistema la ocupación del buffer se representa mediante una variable continua "x". Un canal de voz genera paquetes a una tasa de V paquetes/segundo durante el período de actividad que dura en promedio "1/a" segundos. En este período la variable "x" se incrementará en "V/a" paquetes, también conocida como unidad de información. Un sistema con una capacidad de transmisión de "VC" paquetes/segundo tendrá una capacidad equivalente de "aC" unidades de información por segundo.

Si en un momento determinado hay "i" canales de voz activos se generarán "ai" unidades de información por segundo. El buffer, al mismo tiempo, se estará vaciando a una velocidad de "aC" unidades de información por segundo. El buffer se llenará si "i>C", y se vaciará si "i<C".

Con estas premisas se define F(i,x) o la función de distribución de probabilidad en el momento "t", con el estado "i", es decir, la probabilidad de que la ocupación del buffer sea igual o menor que "x" cuando haya "i" canales activos. La ecuación (1) muestra el proceso de llenado/vaciado del buffer en términos de la función "F".

$$F_i(t+\Delta T,x)=\left[N-(i-1)\right]\lambda\Delta t F_{i-1}(t,x)+(i+1)\alpha\Delta t F_{i+1}(t,x)+\left\{1-\left[(N-i)\lambda+i\alpha\right]\Delta t\right\}F_i\left[t,x-(i-C)\alpha\Delta t\right]+o(\Delta t) \quad (1)$$

Transformando la ecuación (1) en una ecuación diferencial, y asumiendo condiciones estacionarias después de un período de transición, la ecuación (1) se convierte en un sistema de N ecuaciones representadas en la ecuación (2).

$$\begin{aligned}
-C\alpha\frac{dF_0(x)}{dx} &= -N\lambda F_0(x)+\alpha F_1(x)\\
(1-C)\alpha\frac{dF_1(x)}{dx} &= N\lambda F_0(x)-\left[(N-1)\lambda+\alpha\right]F_1(x)+2\alpha F_2(x)\\
(2-C)\alpha\frac{dF_2(x)}{dx} &= (N-1)\lambda F_1(x)-\left[(N-2)\lambda+2\alpha\right]F_2(x)+3\alpha F_3(x)\\
&\vdots\\
(N-C)\alpha\frac{dF_N(x)}{dx} &= \lambda F_{N-1}(x)-N\alpha F_N(x)
\end{aligned} \quad (2)$$

Resolviendo este sistema de ecuaciones se obtiene la probabilidad "G(x)", descrita en la ecuación (3), de que la ocupación del buffer exceda el valor "x".

$$G(x)=1-F(x)=-ae^{zx}=\rho\cdot e^{-\frac{(1-\rho)(1+\gamma)}{(1-C)}x} \quad (3)$$

La ecuación (3) indica que el término asintótico de la probabilidad de saturación "G(x)" es independiente del número de canales multiplexados "N" si la capacidad de transmisión se normaliza. Modificando la probabilidad "G(x)", y transformándola en unidades de información, se obtiene la probabilidad de que el buffer exceda un valor de "i" paquetes, ecuación (4).

$$P[l>i]=G[\alpha i/V]=\rho\cdot e^{-\frac{(1-\rho)(1+\gamma)}{(1-C)}\alpha i/V} \quad (4)$$

6. Simulación y Resultados Numéricos

En esta sección se compara la estimación de la probabilidad de bloqueo o pérdida de paquetes debido a la saturación del buffer utilizando tres métodos analíticos (SMP, CTMC, UAS) y dos métodos basados en simulación, simulación convencional y simulación con inferencia estadística.

Los parámetros seleccionados para determinar las características de la red, el número de canales multiplexados, y la característica de actividad y silencio de la señal de voz, son aquellos utilizados en los estudios realizados en [1]. El promedio de los períodos de actividad es de 1.35 segundos, y el promedio de los períodos de silencio es de 1.65 segundos [5]. Estos valores proporcionan una fracción de actividad del 45%. La tasa de generación de paquetes es de 62.5 paquetes por segundo, con un tiempo de 16 ms entre paquetes.

Las figuras 3 a 6 muestran la distribución acumulada complementaria de bloqueo en función del tamaño del buffer medido en paquetes. Esta función representa la probabilidad de bloqueo de un paquete debido a la saturación del buffer. El número de canales de voz multiplexados es de 8, 15, 30 y 45 para las figuras 3, 4, 5 y 6 respectivamente. La capacidad de transmisión en cada estudio se ha elegido de manera que la intensidad de tráfico en todos los casos sea del 85%.

Los resultados de los cinco diferentes métodos han sido gráficamente representados utilizando el siguiente formato:

1. Simulación Convencional	o o o o o
2. Simulación e Inf. Estadística	x x x x x
3. SMP	-[]-[]-[]-[]-
4. CTMC	-+-+-+-+-
5. UAS	----------

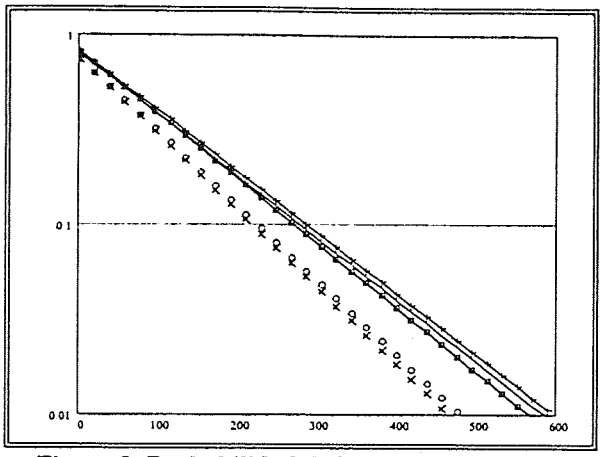


Figura 3. Probabilidad de Bloqueo en Función del Tamaño del Buffer. N = 8 Canales de Voz con Intensidad de Tráfico 85%.

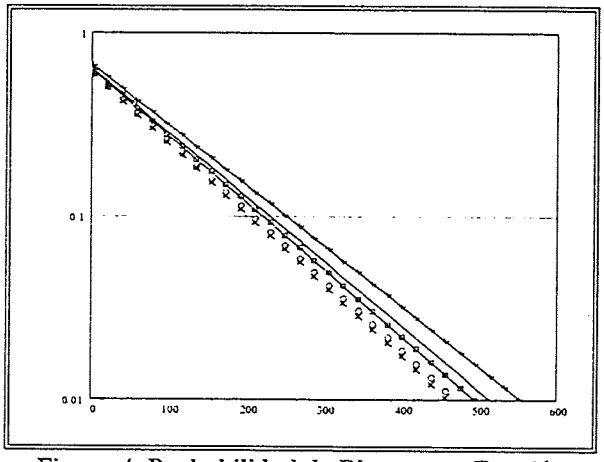


Figura 4. Probabilidad de Bloqueo en Función del Tamaño del Buffer. N = 15 Canales de Voz con Intensidad de Tráfico 85%.

En estas figuras es posible apreciar que los dos modelos basados en cadenas de Markov, así como el método de flujo de fluidos UAS, predicen sistemáticamente colas de mayor tamaño que las estimadas por los métodos de simulación.

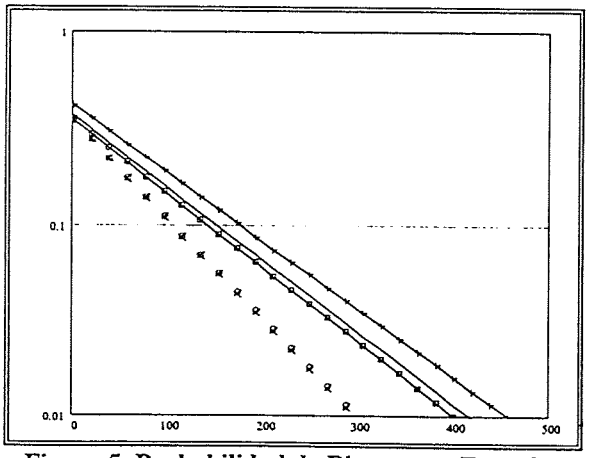


Figura 5. Probabilidad de Bloqueo en Función del Tamaño del Buffer. N = 30 Canales de Voz con Intensidad de Tráfico 85%.

Esta diferencia se incrementa a medida que se aumenta el número de canales multiplexados. En general cuanto mayor sea el número de canales multiplexados en un sistema mejor será el comportamiento de la red, ya que la característica estocástica de los procesos asociados a cada llamada hace que las variaciones individuales en una u otra dirección sean compensadas por las variaciones de otras llamadas independientes. Los resultados presentados en estas gráficas indican que los métodos analíticos tienden a subestimar esta tendencia.

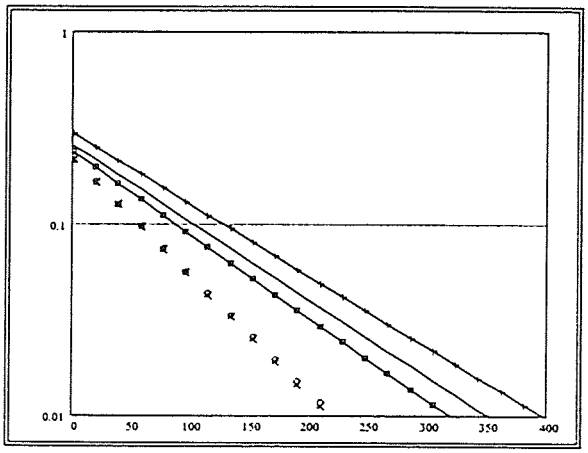


Figura 6. Probabilidad de Bloqueo en Función del Tamaño del Buffer. N = 45 Canales de Voz con Intensidad de Tráfico 85%.

7. Conclusiones

Un total de cinco métodos, tres analíticos y dos basados en simulación han sido utilizados para la estimación de la probabilidad de pérdida de paquetes en la transmisión multiplexada de voz con detección de actividad.

Los recientes avances en comunicaciones y la proliferación de redes basadas en conmutación de paquetes, han generado una nueva serie de estudios sobre la integración de voz por paquetes con otros tipos de tráficos tradicionalmente asociados a las redes de conmutación de paquetes.

En la transmisión de voz por paquetes es posible obtener mejoras en la gestión y comportamiento de la red si se dispone de un modelo preciso de las características de la señal que se desea transmitir, en particular la demanda mínima de transmisión que se ajuste a los parámetros de calidad de servicio establecidos en la red. Los parámetros de calidad de servicio utilizados en la transmisión de voz por paquetes en redes ATM incluyen: la probabilidad de pérdida de paquetes; el retardo promedio de transmisión de un paquete; la variación temporal del retardo; y la varianza del retardo.

Los estudios presentados en este artículo pueden ser extendidos para evaluar la sensibilidad de los métodos estudiados cuando se varía la fracción de actividad de la señal de voz, así como el tipo de distribución de los períodos de actividad y de silencio. Otros estudios pueden incluir la evaluación de estos métodos en la estimación de eventos con probabilidades de varios órdenes de magnitud inferiores a los presentados en este trabajo.

Referencias

[1] Daigle, J.N., and Langford, J.D.. "Models for Analysis of Packet Voice Communication Systems." *IEEE Journal on Selected Areas in Communications*. **SAC-4**, 6, 847-855 (1986).

[2] Olabe, M. A., F. Gondra, and J.L. Perez, *"Estimation of Cell Loss Probability in an ATM Multiplexer Using Simulation and Statistical Extrapolation"*, Proceedings of the 8th. European Simulation Symposium, Genoa, Italy, 556-559 (1996).

[3] Gopal, P.M., and Kadaba, B. "A Simulation Study of Network Delay for Packetized Voice." *Proc. of Globecom* (1986).

[4] Schwartz, M. *Broadband Integrated Networks.* Prentice Hall, Englewood Cliffs, N.J. (1996).

[5] Brady, P.T. *"A Model for Generating On-Off Speech Patterns in Two-way Conversations"*, Bell Syst. Tech. J. **48**, 2445~2472 (1969).

[6] Daigle, J.N., and Langford, J.D. *"Queueing Analysis of a Packet Voice Communication System."* Conf. Rec. IEEE INFOCOM'85, Washington, D.C. 18~26, (1985.).

[7] Neuts, M.F., *Solutions in Stochastic Models.* Baltimore, MD. The John Hopkins University Press. (1981).

[8] Anick, D., Mitra, D., and Sondhi, M.M. *"Stochastic Theory of a Data-Handling System with Multiple Sources."* Bell System Tech. Journal., **61**, 8, 1871~1894 (1982).

[9] Weiss, A., *"An Introduction to Large Deviations for Communication Networks"*, IEEE Journal in Selected Areas in Communications, **13**, 6. 938 ~ 952 (1995).

[10] Duffield, N.G., Lewis J.T., O'Conell N., Russel R., and Toomey F., *"Entropy of ATM Traffic Streams: A Tool for Estimating QoS Parameters"*, IEEE Journal in Selected Areas in Communications, **13**, 6, 981~990 (1995).

[11] Langford, J.D., "Queueing Analysis of a Packet Voice Communication System Via a Semi-Markov Process." Masters Thesis, Clemson University, Clemson, S.C. (1984).

Modelado de Tráfico Ethernet sobre ATM

A. Reyes Lecuona, J. J. Márquez, E. Casilari, A. Díaz Estrella y F. Sandoval
Dpto. Tecnología Electrónica, E.T. S. I. Telecomunicación
Universidad de Málaga, Campus de Teatinos, 29071, Málaga
Tlf. (95) 2132755, Fax (95)2131447, Correo electrónico: arcadio@dte.uma.es

Abstract:

In this paper a study of the influence of the statistical characteristics of a given data traffic in its queuing behaviour is presented. We propose several models based on an ON-OFF process that fit various statistics of the time spent in each state. Analysing the queuing behaviour of each model, it's concluded that first order statistics determine the queuing response for low buffer occupancy, while second order statistics are very important for high buffer occupancy.

1. Introducción

El desarrollo de las redes de banda ancha ha propiciado la aparición de nuevos tipos de tráfico que es necesario caracterizar y modelar con objeto de diseñar, dimensionar y probar tales redes. En este sentido, han aparecido recientemente muchos trabajos que inciden en el problema de la caracterización, modelado y generación de distintos tipos de tráfico, correspondientes a los diferentes tipos de servicio existentes en estas redes.

Por otra parte, la caracterización del tráfico nos proporcionará un conocimiento del mismo que resulta fundamental en la búsqueda de parámetros que describan de una forma realista el comportamiento del tráfico. En la actualidad, los parámetros descriptores de tráfico que propone el ATM forum y la ITU [1] se reducen a estadísticos básicos como son la velocidad media, la velocidad de pico, y algunos más complejos como el rafagueo o la duración de la ráfaga. Sin embargo, estos descriptores resultan insuficientes para describir el comportamiento del tráfico de una forma realista. En primer lugar, no contemplan determinados fenómenos de correlaciones del tráfico que son muy importantes a la hora de considerar el comportamiento del sistema por el que circula dicho tráfico. En segundo lugar no es fácil definir los parámetros relativos a las ráfagas, sobre todo cuando el tráfico presenta una superposición de muchos tipos de ráfagas, como de hecho sucede en la realidad.

En este sentido, en esta comunicación se propone un modelo de tráfico Ethernet sobre una red ATM y se presenta un estudio de la influencia de los estadísticos de primer y segundo orden sobre el comportamiento de una cola alimentada por este tipo de tráfico. Tomando como base de medidas de tráfico real realizadas en un conmutador ATM que multiplexaba tráficos de varias redes Ethernet IEEE 802.3 a 100 Mbps, se desarrollan varios modelos que imitan las características estadísticas del tráfico.

Esta comunicación está estructurada como sigue. En la sección 2 se realiza un estudio preliminar de las trazas de tráfico obtenidas. A partir de ese estudio se propone una estructura para modelar este tipo de tráfico. Asimismo se describen los estudios realizados sobre el modelado de tráfico Ethernet hasta la fecha por otros investigadores. En la sección 3 se describe el modelo propuesto y se determinan los distintos parámetros que lo caracterizan. En la sección 4 se presentan los resultados de las simulaciones realizadas con los distintos modelos y se comparan con los resultados obtenidos a partir del tráfico real en el mismo sistema de pruebas. Por último, en la sección 5 se discuten los resultados obtenidos y en la 6 se exponen las conclusiones de este estudio.

2. Estudio Preliminar del Tráfico Analizado

Las trazas analizadas en este estudio han sido cedidas por Telefónica I+D, que las obtuvo como resultado de monitorizar su red durante un periodo de 14 segundos. El sistema de monitorización toma una muestra por cada célula ATM, anotando el instante de tiempo en el que se genera con una precisión de 100 ns, el VPI (Virtual Path Identifier) y el VCI (Virtual Chanel Identifier).

El tráfico que se estudia corresponde al total generado por una sola red Ethernet. En el sistema monitorizado se multiplexaban varias redes locales Ethernet IEEE 802.3 a 100Mbps sobre un conmutador ATM de Fore con entradas a 155 Mbps. Por una de sus salidas se obtuvo una copia de todo el tráfico agregado, de forma que cada red LAN se asignó a un VCI diferente.

Posteriormente se separaron los diferentes canales virtuales, realizándose todo el estudio sobre el tráfico generado por cada una de las LAN por separado. En la figura 1 se muestra un fragmento de una de las trazas capturadas. En esta figura se puede comprobar que el comportamiento de las fuentes Ethernet puede ser modelado usando fuentes ON-OFF. Esto era de esperar, ya que cada vez que una máquina toma el bus Ethernet para transmitir un paquete, lo hace de forma exclusiva. La transmisión de este paquete se traduce en una ráfaga de

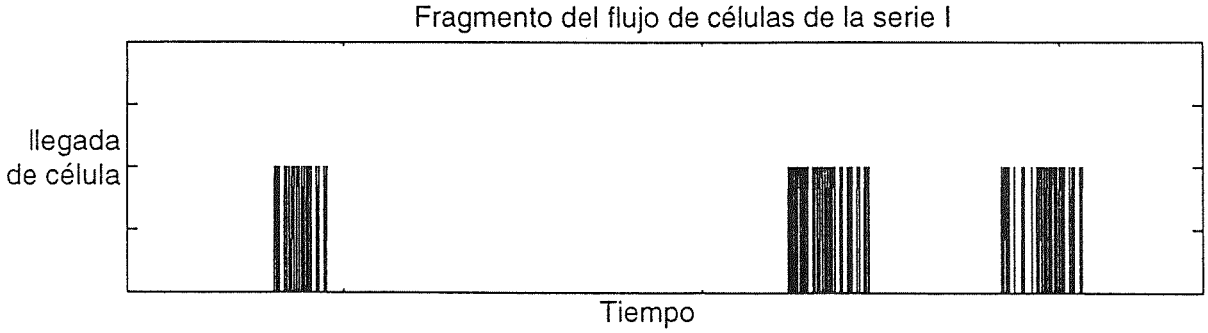


Figura 1. Fragmento del flujo de células. Se representa con una línea vertical cada célula del tráfico real de la serie I. Como puede comprobarse, un proceso ON-OFF puede modelarlo correctamente.

velocidad constante (100 Mbps). Entre dos de estas ráfagas se observa un periodo de silencio correspondiente al tiempo en que nadie está usando el bus Ethernet.

Parece por lo tanto evidente que una fuente ON-OFF modelará correctamente este tipo de tráfico. Sin embargo, es razonable pensar que la duración de los periodos ON y OFF no se distribuya exponencialmente. De hecho, sabemos que la duración de los periodos ON es proporcional al tamaño del paquete Ethernet que lo produce, y que dichos paquetes tienen una serie de longitudes características [2]. Por ejemplo, el establecimiento de conexiones de niveles superiores suelen generar paquetes muy pequeños, que se traducen en ráfagas de muy pocas células. Por otra parte el tamaño máximo de los paquetes Ethernet es de 1.500 bytes, por lo que la transmisión de ficheros grandes genera paquetes de ese tamaño.

R. O. Onvural [2] propone un modelo ON-OFF en el que, además, se consideran las distribuciones de probabilidad de los periodos OFF. Para una descripción de las características estadísticas del tráfico LAN, de donde se pueden extraer dichas probabilidades, véase [3].

En [4] podemos encontrar una amplia descripción de las distintas metodologías usadas para diseñar los parámetros de modelos MMPP y MMBP en el modelado de fuentes Ethernet.

Recientemente se han publicado varios estudios que centran su interés en las correlaciones que posee este tipo de tráfico. W. E. Leland realiza un estudio en el que demuestra que el tráfico generado en una red Ethernet posee dependencias a largo plazo, o lo que es lo mismo, un comportamiento autosemejante o fractal [5]. Este tipo de comportamiento se manifiesta en una correlación que no decae de forma exponencial, sino hiperbólica [6]. En los últimos años han aparecido multitud de trabajos que estudian las características autosemejantes del tráfico [7], por considerarlo muy importante a la hora de analizar su comportamiento en un sistema de colas, si bien también existen trabajos que restan importancia a este fenómeno [8].

En el presente estudio se analizará la influencia de la dependencia a corto plazo de las duraciones de los periodos ON en un sistema de colas. Para mostrar los resultados de dicho estudio, se utilizarán dos de las series capturadas, que denominaremos secuencia I y secuencia II, y que se muestran en las figuras 2 y 3 respectivamente. Se han elegido estas dos muestras por ser extremas en lo que se refiere a su estacionariedad. La serie I es muy estacionaria, mientras que la serie II no.

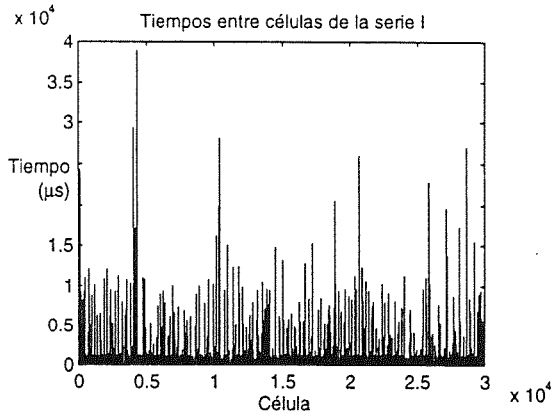


Figura 2. Tiempos entre células de la serie I. Para cada célula de la serie se representa la separación en microsegundos con la siguiente

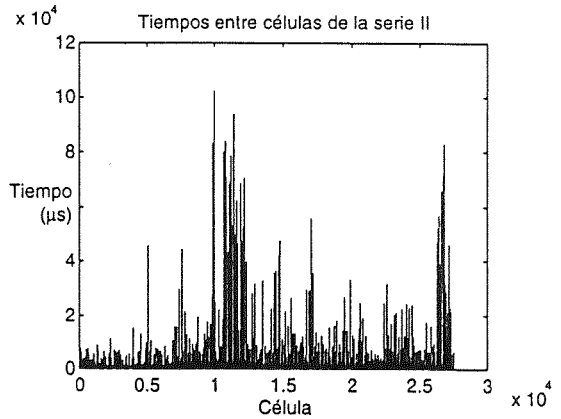


Figura 3. Tiempos entre células de la serie II. Para cada célula de la serie se representa la separación en microsegundos con la siguiente

3. Modelo Propuesto

Como se ha comentado en la sección anterior, en primer lugar debemos extraer la duración de los periodos ON y OFF. Para hacer esto se dispondrá un umbral γ para los periodos OFF, de tal forma que si el tiempo entre dos células es menor que γ se considerará que forman parte del mismo periodo ON, mientras que si dicha separación es mayor que γ, se considerará como un periodo OFF. En nuestro caso, la separación mínima entre dos células (velocidad de pico) será 2,7 μs, correspondiente a una velocidad de 155 Mbps. Se considerará un umbral γ=20 μs, ya que así se garantiza no perder periodos ON y sigue siendo un tiempo mucho menor que la mayoría de los periodos OFF.

En la figura 4 se representa la función de distribución de probabilidad de la duración de los periodos ON de la serie I, y en la figura 5 la misma función para la duración de los periodos OFF de la misma serie. En las figuras 6 y 7 se muestran las correspondientes a la serie II.

Analizando estas figuras, se puede comprobar que la aproximación exponencial para la duración de los periodos, aunque podría ser válida para los periodos OFF, no es realista en el caso de los periodos ON. La forma que presenta la distribución de probabilidad de los periodos ON obedece claramente a la existencia de paquetes de un tamaño determinado (1500 bytes), lo que provoca el escalón que se aprecia en las figuras.

Veamos ahora la correlación que existe entre las duraciones de los periodos. En la figura 8 se muestra la autocorrelación entre los periodos ON de la serie I. En la figura 9 representamos la correspondiente a las duraciones de los periodos OFF de la misma serie. Asimismo, en las figuras 10 y 11 se hace lo propio con la serie II. Las correlaciones cruzadas entre periodos ON y OFF se han considerado despreciables.

Analizando las figuras 8 y 10 podemos concluir que existe una dependencia estadística clara entre la duración de dos periodos ON consecutivos, lo que se pone de manifiesto al aparecer una autocorrelación apreciable. En la serie II esto último es mucho más importante. La causa de este fenómeno se puede comprender si volvemos a las figuras 2 y 3, donde se observa que esta segunda serie es menos estacionaria que la primera. Sin

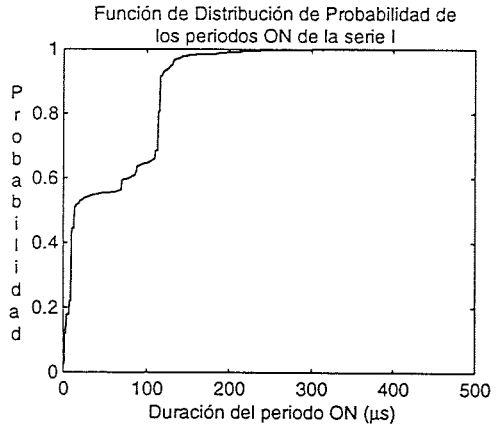


Figura 4. Función de Distribución de Probabilidad de la duración de los periodos ON de la serie I. Se observa claramente el escalon ocasionado por los paquetes de 1500 bytes.

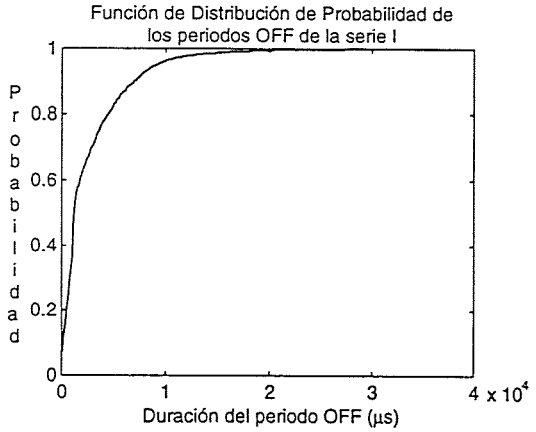


Figura 5. Función de Distribución de Probabilidad de la duración de los periodos OFF de la serie I. Se Puede afirmar que presenta un comportamiento aproximadamente exponencial.

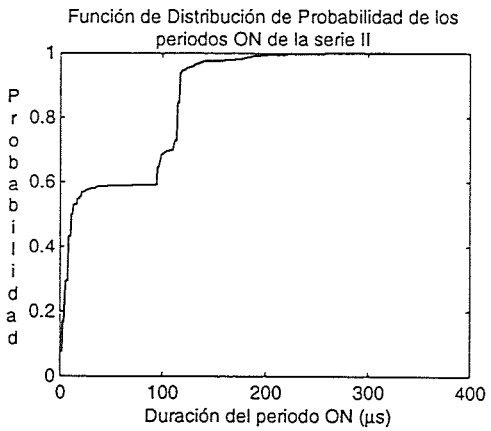


Figura 6. Función de Distribución de Probabilidad de la duración de los periodos ON de la serie II. Se observa claramente el escalón ocasionado por los paquetes de 1500 bytes.

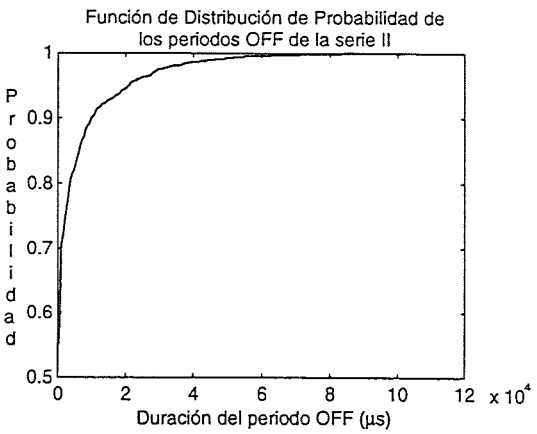


Figura 7. Función de Distribución de Probabilidad de la duración de los periodos OFF de la serie II. Se Puede afirmar que presenta un comportamiento aproximadamente exponencial.

embargo, analizando las figuras 9 y 11 se aprecia que la duración de los periodos OFF presenta mucha menos correlación, es decir, existe una independencia estadística mayor entre las duraciones de estos periodos OFF.

Una vez caracterizado el tráfico, se propondrá un modelo que incluya todas las características estadísticas observadas. Como ya se ha apuntado anteriormente, nuestro modelo será un proceso ON-OFF, tal y como se muestra en la figura 12, en la que se modelará la permanencia en el estado OFF mediante una variable aleatoria que posea la misma distribución de probabilidad que la medida en el tráfico real. La permanencia en el estado ON deberá incluir los efectos de la autocorrelacion observada. Esto se conseguirá con un modelo PAR (Proyected AutoRegresive) [9] modificado de primer orden [10], de forma que capturemos tanto la distribución de probabilidad como la autocorrelación en los primeros puntos.

El modelo PAR [9] es una modificación del modelo AR [11], que consiste en un filtro AR de orden k cuya expresión es la siguiente:

$$y(n) = b_0 \cdot x(n) + a_1 \cdot y(n-1) + \cdots + a_k \cdot y(n-k)$$

Si el orden del filtro es 1, y $x(n)$ es un ruido blanco, el valor del primer punto de la autocorrelación de $y(n)$ coincide con b_1. Si $x(n)$ es gaussiano de media cero, $y(n)$ también lo será. Se Puede proyectar la salida de ese filtro AR por una función de distorsión D(·) que sea la composición de la gaussiana con la inversa de la función de distribución que se desea obtener, obteniendo así una serie que llamaremos $z(n)$. De esta forma, se obtendrá una serie con la función de distribución deseada y con la autocorrelación aproximada. Si la función de distribución objetivo es parecida a una gaussiana, la función de distorsión será bastante lineal, con lo que la autocorrelación de $z(n)$ será parecida a la de $y(n)$.

Sin embargo, en nuestro caso las funciones de distribución de probabilidad de las duraciones de los periodos ON están muy lejos de ser gaussianas, por lo que la función D(·) será muy no lineal, y la autocorrelación de $y(n)$ se perderá bastante. La modificación que se propone al modelo PAR para resolver el problema, dado que nuestro filtro AR es de primer orden, es multiplicar el coeficiente a_1 por un coeficiente η, heurísticamente diseñado para corregir la autocorrelación. En la figura 13 se

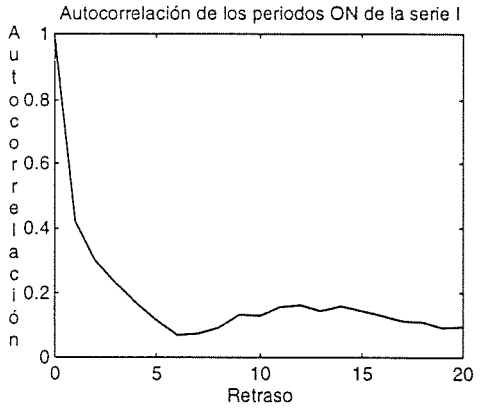


Figura 8. Función de Autocorrelación de la duración de los periodos ON de la serie I. Existe una clara dependencia estadística entre dos muestras consecutivas.

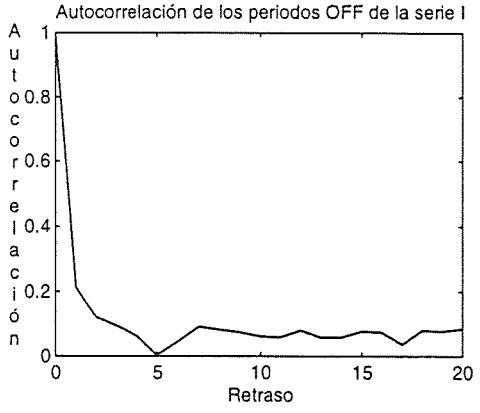


Figura 9. Función de Autocorrelación de la duración de los periodos OFF de la serie I. La dependencia estadística entre dos muestras consecutivas es menor que en el caso de los periodos ON.

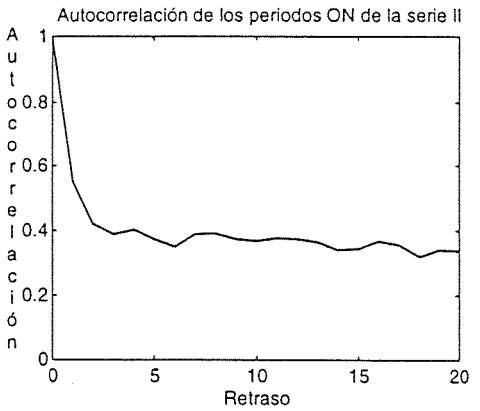


Figura 10. Función de Autocorrelación de la duración de los periodos ON de la serie II. Existe una clara dependencia estadística entre dos muestras consecutivas.

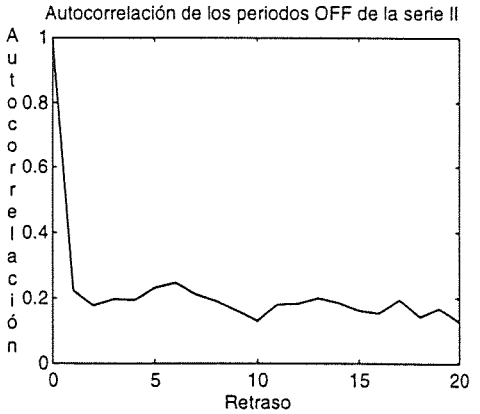


Figura 11. Función de Autocorrelación de la duración de los periodos OFF de la serie II. La dependencia estadística entre dos muestras consecutivas es menor que en el caso de los periodos ON.

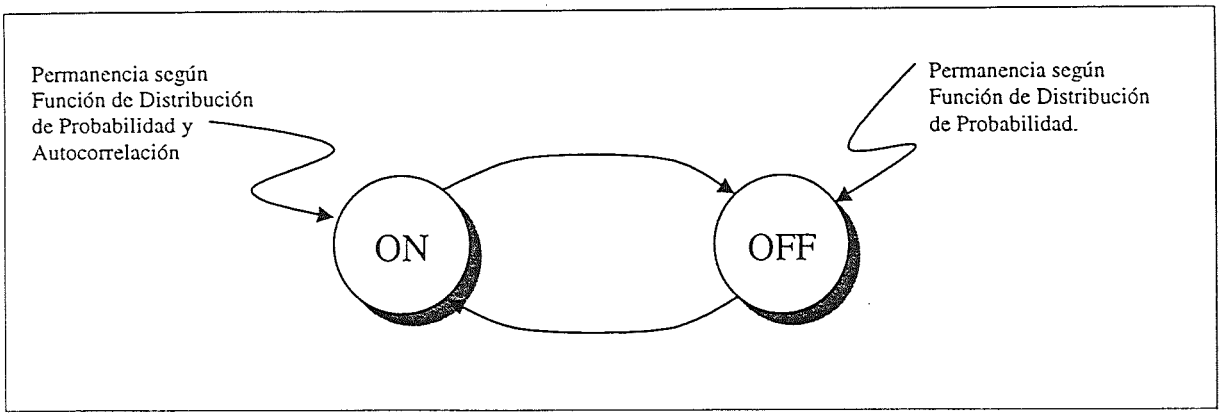


Figura 12. Diagrama del proceso ON-OFF que proponemos. El tiempo de permanencia en el estado OFF se modela según la distribución de probabilidad del tráfico real. Para el tiempo de permanencia en el estado ON modelaremos además la autocorrelación del tráfico real.

muestran las autocorrelaciones de los periodos ON de la serie I, y las de los modelos PAR y PAR modificado con un coeficiente η=1,2.

Sobre un simulador de eventos discretos programado en C++ se ha implementado este modelo para las dos series consideradas, modelando las funciones de distribución de probabilidad con una composición de tramos lineales a partir de histogramas de 300 niveles.

4. Resultados de las Simulaciones.

Para estudiar el efecto de los distintos estadísticos en el comportamiento del tráfico en un sistema de colas se han diseñado, además del modelo propuesto, una serie de modelos que se detallan a continuación:

- ON-OFF. Se llamará así a un modelo basado en un proceso ON-OFF con tiempos de permanencia en cada estado distribuidos exponencialmente. Ajustaremos por lo tanto el tiempo medio de permanencia en cada estado, así como la velocidad de transmisión durante el estado ON.
- NO-CORRELADO. Se usará también un modelo, como el propuesto en [2], en el que se adaptan las funciones de distribución de probabilidad de la permanencia en cada estado, pero no la correlación.
- CORRELADO. Se denominará así al modelo en el que se adaptan, tanto las funciones de distribución de probabilidad de ambos estados, como la de autocorrelación de la permanencia en el estado ON.

Las simulaciones llevadas a cabo consistieron en la generación de células con cada uno de estos tres modelos. Dichas células eran encoladas en un *buffer* de tamaño infinito, sobre el cual se midió la función de probabilidad de ocupación del mismo. Se ha considerado esta medida del comportamiento del tráfico porque en ella están incluidas todos los parámetros de calidad de servicio que se tienen en cuenta normalmente (probabilidad de pérdidas, retardo, *jitter*, etc.). La velocidad de servicio de la cola se ajustó para conseguir una tasa de utilización del servidor del 10%.

En la figura 14 se representa la ocupación

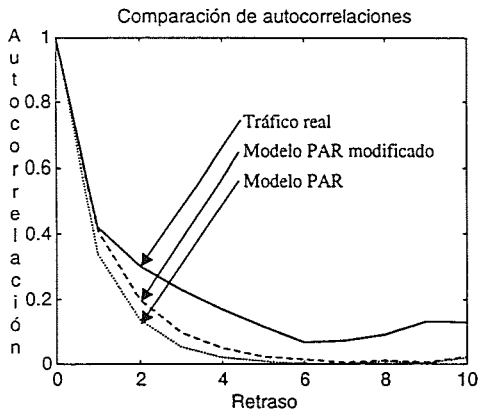


Figura 13. Función de Autocorrelación de la duración de los periodos ON de la serie I y del modelo PAR y PAR modificado. El filtro AR implementado es de primer orden, por lo que sólo se ajusta el primer punto de la autocorrelación. Con el modelo PAR modificado se corrige el error producido por la función de distorsión.

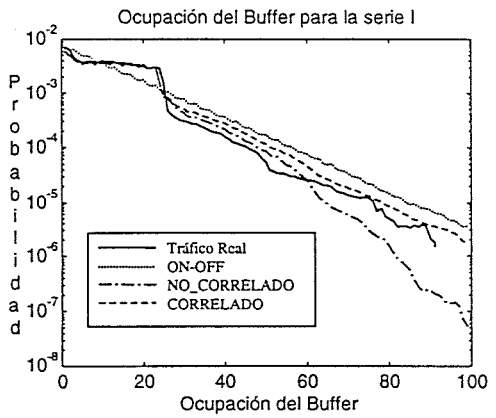


Figura 14. Función probabilidad de la ocupación del *buffer* para la serie I y un 10% de utilización del servidor. Se representan los resultados obtenidos para el tráfico real y para los tres modelos comentados en el texto. El modelo propuesto (CORRELADO) obtiene mejores resultados que los otros dos.

del *buffer* para los tres modelos y para la serie real. La curva de la serie real se ha obtenido introduciendo células en la cola de acuerdo con los tiempos entre células medidos. Las simulaciones se hicieron para un tiempo equivalente de 500 segundos.

Se pueden observar dos características muy importantes. En primer lugar, la ocupación del buffer, cuando es pequeña, es capturada por los modelos que incluyen la función de distribución de probabilidad de la duración de los periodos de permanencia en cada estado. El modelo ON-OFF produce una ocupación claramente exponencial, por lo que no puede reproducir el escalón que se aprecia en la figura 14 alrededor del tamaño 20.

En segundo lugar, parece necesario tener en cuenta la autocorrelación de los periodos ON si se quiere reproducir el comportamiento para ocupaciones altas. El modelo NO-CORRELADO pierde su eficacia al observar la zona de ocupaciones altas.

Veamos ahora los resultados obtenidos para la serie II. En la figura 15 se muestran los resultados obtenidos para dicha serie con los tres modelos y el tráfico real. Es importante destacar aquí que esta serie, aun presentando paquetes de un tamaño similar a los de la serie I, presenta una velocidad media menor. Como puede apreciarse en la figura 3, este fenómeno es debido a largos periodos de baja actividad. Por lo tanto, a pesar de realizarse los experimentos para una tasa media de utilización del servidor de un 10%, el comportamiento responderá al de un sistema más cargado, ya que durante los periodos largos de actividad la tasa de utilización será mayor.

Analizando la figura 15 se parecia que, efectivamente, es necesario modelar la distribución de probabilidad de la permanencia en los estados para reproducir el escalón, que ahora se produce alrededor de una ocupación de 30 células. Sin embargo, la mayor carga relativa de esta segunda serie revela un fenómeno muy interesante: la tendencia de la ocupación del *buffer* para ocupaciones altas únicamente depende de la autocorrelación de los tiempos de permanencia en el estado ON. Obsérvese como los dos modelos que tienen esos tiempos incorrelados (ON-OFF y NO-CORRELADO) presentan un comportamiento similar para altas ocupaciones. Comportamiento que, por otra parte, difiere significativamente del que presenta el tráfico real. Sin embargo, el modelo CORRELADO sigue a la fuente real tanto en baja ocupación como en alta.

Para corroborar estos últimos análisis se realizó un tercer experimento que consistió en simular de nuevo la serie I con todos los modelos, pero adaptando la velocidad de servicio de la cola de tal modo que tuviéramos una tasa de utilización del servidor de un 50%. En la figura 16 se muestran dichos resultados.

Estudiando estos resultados se debe concluir que los modelos presentados captan mejor el comportamiento del tráfico en situaciones de utilización baja del servidor de la cola. No obstante es muy importante destacar que la figura 16 corrobora los análisis realizados sobre la figura 15, ya que la autocorrelación de los periodos ON es fundamental para reproducir el comportamiento de la cola en la zona de ocupación alta. Como puede observarse, únicamente el modelo CORRELADO mantiene la probabilidad de ocupación alta en un nivel aceptable.

5. Discusion

Como resultado de los experimentos realizados podemos resaltar ciertos aspectos de interés.

La influencia de los estadísticos de primer orden ha sido evidente en la zona de ocupaciones

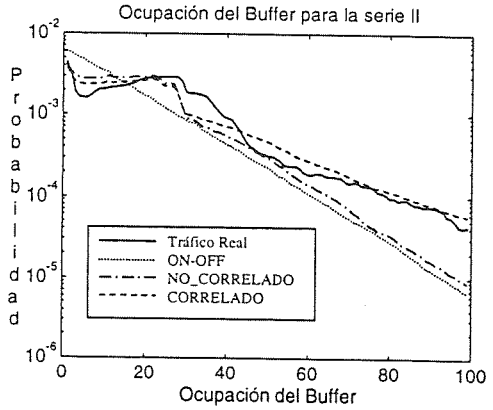


Figura 15. Función probabilidad de la ocupación del *buffer* para la serie II y un 10% de utilización del servidor. Se representan los resultados obtenidos para el tráfico real y para los tres modelos comentados en el texto. El modelo que proponemos (CORRELADO) obtiene mejores resultados que los otros dos.

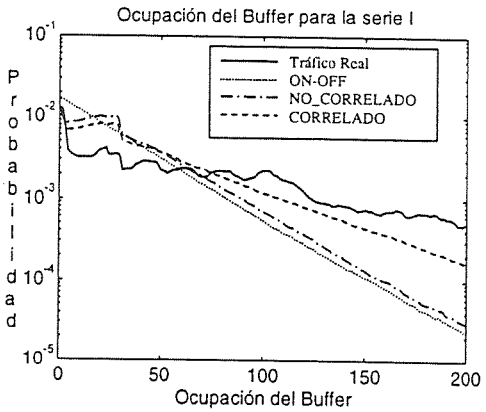


Figura 16. Función probabilidad de la ocupación del *buffer* para la serie I y un 50% de utilización del servidor. Se representan los resultados obtenidos para el tráfico real y para los tres modelos comentados en el texto. El modelo que proponemos (CORRELADO) obtiene mejores resultados que los otros dos.

bajas, mientras que la de los estadísticos de segundo orden se ha puesto de manifiesto en la zona de ocupaciones altas.

Por otra parte, si dos tráficos presentan un comportamiento similar en la zona de ocupación alta van a presentar una probabilidad de pérdidas semejante cuando el *buffer* sea finito. Sin embargo, parece claro que si se quiere conseguir que el tráfico sintético presente el mismo retardo y el mismo *jitter*, se debe conseguir que la ocupación del *buffer* presente una forma parecida en todas las regiones. De hecho, en este caso, la región más importante es la de ocupaciones bajas, ya que es la mas probable y la que, por lo tanto, más contribuye al retardo y al *jitter*.

Otra cuestión relevante es el comportamiento de los tráficos cuando la tasa de utilización del servidor es alta. Aquí se ponen de manifiesto fenómenos que no se puede capturar con los modelos propuestos. Quizás el tráfico real presente un comportamiento a largo plazo que no se puede apreciar en unas series tan cortas como las que se han analizado. Leland [5] presenta un estudio sobre tráfico Ethernet en el que muestra que existe un comportamiento fractal o dependiente a largo plazo, que podría influir en el comportamiento en una cola. En cualquier caso, sería necesario poder disponer de series más largas para poder estudiar este comportamiento, del que, por otra parte se conoce muy poco en lo que respecta a su influencia en los sistemas de colas.

Por otra parte, a la hora de diseñar un modelo se debe tener en cuenta qué comportamiento queremos reproducir, porque en función de ello, debemos imitar unas características estadísticas u otras. Pero, aún más importante, si pensamos en el diseño de algún tipo de control (admisión, vigilancia, etc.), debemos tener presente qué parámetros son los responsables de un determinado comportamiento. Por ejemplo, un control de vigilancia que controle sólo la media no será capaz de garantizar adecuadamente la calidad de servicio si el tráfico presenta funciones de distribución de probabilidad extrañas o autocorrelaciones, como es el caso del tráfico estudiado en esta comunicación.

6. Conclusiones y Futuros Trabajos

En este trabajo se ha presentado un estudio del comportamiento en una cola del tráfico de redes Ethernet sobre ATM. A través de varios modelos propuestos se ha observado la influencia de cada uno de los parámetros que describen el tráfico sobre los parámetros de calidad de servicio. Así, se ha comprobado que el parámetro más relevante para conseguir un ajuste del *jitter* es la distribución de probabilidad del tiempo de permanencia en cada estado. Sin embargo, la autocorrelación de dichos tiempos, en el caso del estado ON, es la característica más importante de cara a conseguir reproducir una probabilidad de pérdidas.

Los futuros trabajos en esta línea pasan por conseguir series más largas, en las que poder estudiar las dependencias a largo plazo y su influencia en el comportamiento en la cola. Si realmente se presentara dicha dependencia, su influencia sería importante sobre todo en la zona de ocupaciones altas, ya que es allí donde las autocorrelaciones producen efecto. Las dependencias a largo plazo [6][7] se traducen en autocorrelaciones que no decaen exponencialmente.

Agradecimientos

Este trabajo ha sido parcialmente financiado por la Comisión Interministerial de Ciencia y Tecnología CICYT, proyecto N° TIC96-0743. Asimismo, agradecemos a Telefónica I+D la cesión de las trazas de tráfico usadas en este estudio.

Referencias

[1] ITU-T, Recomendación I.371, *Traffic Control and Congestion Control in B-ISDN*, Génova, 1996.

[2] Onvural, R. O., *Asynchronous Transfer Mode Networks*, Norwood, Artech House, 1994.

[3] Gusella, R. "A Measurement Study of Diskless Workstation Traffic on an Ethernet", *IEEE Trans on Comm.*, pp 39-49, 1990.

[4] Arvidson, A. y Lind, C., "Using Markovian Models to Replicate Real ATM Traffics", *en ATM Networks: Performance Modelling and Evaluation, Vol 2*, Ed. Kouvatsos, D., Chapman & Hall, pp. 39-54, 1996.

[5] Leland, W. E., et al, "On the Self-Similar Nature of Ethernet Traffic", *IEEE JSAC*, 9, N° 3, pp. 284-293, Abril, 1993.

[6] Rose, O., "Estimation of the Husrt Parameter of Long Range Dependent Time Series", *Informe intern, Institute of Computer Science, Universidad de Würzburg*, Febrero, 1996.

[7] Willinger, W. Et al., "A Bibliographical Guide to Self-Similar Traffic and Performance Modeling for Modern High Speed Networks", *in Stochastic Networks: Theory and Applications*, Ed. Kelly, F. P. Et al, Clarendon Press, (Oxford University Press), pp. 339-366, 1996

[8] Ryu, B. K. Y Elwalid, A., "The Importance of Long-Range Dependence of VBR Video Traffic in ATM Traffic Engineering: Myths and Realities", *In Proc of ACM SIGCOMM*, San Francisco, 1996.

[9] Wu, J. L. et al., "Two models for Variable Bit Rate MPEG Sources", *IEICE Trans. On Comm.*, **E78-B**, N° 5, pp. 733-745, Mayo, 1995.

[10] Casilari, E. Et al, "Scene Oriented Model for VBR Video", *Enviado a IEEE Comm Lett.*

[11] Ohta, N., *Packet Video*, Norwood, Artech House, 1996.

Estimación de los parámetros de calidad para distintos tráficos en nodos MTA con enlaces múltiples

Mónica Aguilar Igartua, Francisco Barceló Arroyo, Joan García Haro
E. T. S. de Ingeniería de Telecomunicación de Barcelona
Universidad Politécnica de Cataluña
c/ Jordi Girona, 1-3, 08034 Barcelona
e-mail: {maguilar, barcelo, teljgh}@mat.upc.es

Abstract:

In this paper an ATM switching node is modeled as a finite storage M/D/*s* queueing system. Thus, characterizing the aggregation of traffic in the network, the realistic assumption of finite size buffers and the incorporation of several output transmission links. An approximate analytical method to easily compute the mean value and the quadratic variation coefficient of the number of cells in the buffer is proposed. The method is also useful to obtain the congestion probability that in this case, equals the Cell Loss Rate (CLR) since Poisson arrivals are assumed.

1. Introducción

Uno de los problemas relacionados con la evaluación de redes basadas en el Modo de Transferencia Asíncrono (MTA) es la ausencia de modelos sencillos que permitan obtener una primera aproximación de los parámetros de calidad a que están sometidos los distintos tráficos de la red [CHA]. Así abundan los modelos que consideran una caracterización precisa de dichos tráficos, diferente para CBR (*Constant Bit Rate*), rt-VBR (*real-time Variable Bit Rate*), nrt-VBR (*non real-time Variable Bit Rate*), ABR (*Available Bit Rate*) y UBR (*Unspecified Bit Rate*), pero cuyo cómputo práctico puede ser altamente complejo. Además estos modelos se basan generalmente en el modelado de los *buffers* y servidores en base a las siguientes situaciones extremas:

- Colas infinitas: suele aceptarse esta hipótesis para el caso de tráfico estadístico. En el caso de tráfico poco sensible al retardo y a la variación de retardo puede asumirse la situación de pérdidas muy bajas y *buffer* de capacidad ilimitada [DES, ELW].

- *Buffer* de capacidad nula (caso "*bufferless*"): para tráfico determinista muy sensible a los retardos y a sus variaciones. Si bien el *buffer* sigue siendo imprescindible en la práctica para acomodar variaciones de retardo debidas a múltiples factores, la aproximación "*bufferless*" representa un peor caso [KES].

- Un único servidor: de este modo se modela la multiplexación a la vez que se ignora la posibilidad de múltiples enlaces entre nodos que puede producirse en niveles altos de la red. En tal caso son frecuentes en la literatura los estudios de CAC (*Connection Admission Control*) basados en el concepto de ancho de banda equivalente y en la cota de Chernoff [KES, VEC].

El incorporar hipótesis más realistas supone tener en cuenta el tamaño limitado de los *buffers*. Este tamaño debe ser en general lo suficientemente pequeño para evitar CTD (*Cell Transfer Delay*) altos y a la vez lo suficientemente grande para mantener CLR (*Cell Loss Rate*) bajos. En cualquier caso es posible (y en la práctica necesario) tener un tamaño distinto de *buffer* en función del tipo de tráfico: más pequeño para tráficos deterministas en general más sensibles a los problemas de retardo, y más grande para tráficos estadísticos. Incluso sería posible tener diferentes tamaños para distintos tráficos que aun siendo de la misma clase (determinista o estadístico) tengan pactado parámetros de calidad diferentes. Este escenario no implica la necesidad de ubicar físicamente un *buffer* para cada tipo de tráfico, sino que el mismo *buffer* puede ser compartido por todos, siempre que se

descarten las celdas en función del tráfico al que pertenecen y del límite de *buffer* para dicho tráfico.

Consideramos también, la posibilidad de múltiples enlaces físicos entre dos nodos de la red lo cual debe modelarse obviamente como un sistema multi-servidor. A su vez, la existencia de varios enlaces puede servir para modelar el caso en el que un nodo tiene conexión con varios nodos remotos, ignorándose en la fase de diseño, en la cual se realiza la evaluación, qué parte de la carga está destinada a cada uno de ellos.

En este artículo se modela un nodo de conmutación MTA como una cola M/D/*s* [ALT] de capacidad finita. Es abundante la literatura existente sobre multiplexores MTA en los que se utilizan modelos G/D/1 para modelar el multiplexor en el nodo de acceso. En estos modelos, G suele representar algún proceso de alto coeficiente de variación (MMPP, Pareto, etc.) ya que el principal interés de los modelos es el caso más restrictivo al que pueden dar lugar principalmente los tráficos estadísticos y especialmente el VBR. También puede representar la G un proceso de llegadas NxD (combinación de deterministas) o llegadas acordes a una distribución geométrica. En ambos casos la cola M/M/1 es más restrictiva y suele utilizarse para modelar un peor caso. Sin embargo, en los conmutadores ubicados en niveles altos de la jerarquía de la red MTA cabe esperar un tráfico agregado que hoy por hoy todavía no ha sido caracterizado, en parte debido a la escasez de redes MTA lo "suficientemente cargadas" o con cargas que puedan ser consideradas "típicas" tanto en volumen como en tipos de tráfico que transportan. También en estos niveles altos es donde cabe esperar la necesidad de modelos con múltiples servidores que representen las diversas fibras que conectan dos nodos entre si o bien las salidas de un nodo hacia varios remotos.

A diferencia de la mayoría de trabajos que proponen evaluaciones de multiplexores MTA en base a soluciones exactas para modelos más o menos alejados de la realidad, como es el caso de los modelos G/D/1 sin *buffer* o con *buffer* infinito, los cálculos numéricos para evaluación que se proponen aquí son siempre aproximados. Desde una óptica de ingeniería consideramos que una aproximación es buena siempre que el error relativo en el que se incurre sea muy inferior a otros errores que forzosamente están presentes en el momento de realizar la evaluación o el dimensionamiento de la red. El más típico es el de las previsiones o medidas de tráfico que suponemos que va a transportar la red. Aun en el mejor caso de que pudiéramos medir la intensidad de tráfico real en la red que pretendemos evaluar (cosa que nunca sucede ya que la red que evaluamos siempre es diferente de la red existente) dicha medida sería imprecisa. Aunque el error relativo de la aproximación no sea pequeño, la aproximación también puede ser útil si no existen resultados mejores o exactos, o bien si éstos requieren de un elevado esfuerzo de programación o de cómputo para su obtención. Las aproximaciones presentadas aquí son de las dos clases mencionadas: por un lado aproximaciones muy precisas para el número medio de celdas en el *buffer* y su coeficiente de variación, por otro una aproximación burda pero sencilla de la probabilidad de congestión, que coincide con el CLR por haber asumido llegadas de Poisson.

El orden que se sigue es el siguiente: en la sección 2 se describe en profundidad el modelo utilizado para el nodo de conmutación, la sección 3 contiene la forma de cálculo de los dos primeros momentos del número de celdas que ocupan el *buffer*, en la sección 4 se presenta la forma de cálculo aproximado del CLR dado el CDVT (*Cell Delay Variation Tolerance*) y en la sección 5 se ofrecen resultados numéricos.

2. Modelo del nodo de conmutación

Si bien las líneas generales del modelo de nodo de conmutación asumido se han mencionado en la introducción, en esta sección se profundiza sobre las hipótesis del modelo. La forma en que caracterizamos un nodo de conmutación de nivel alto de la red MTA (no de acceso) es la que se representa en la Figura 1, es decir una cola M/D/*c* con capacidad limitada *B*. La justificación de la duración determinista del tiempo de servicio es clara dado que cada celda MTA de 53 octetos tarda exactamente $53x8/v_t$ segundos en ser transmitida, siendo v_t la velocidad de transmisión de cada enlace en bps.

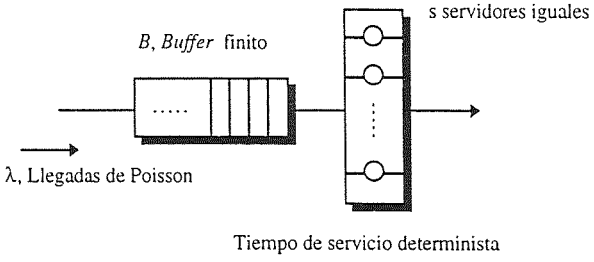


Fig. 1: Modelo del nodo de conmutación MTA.

Parece obvio que el proceso de llegadas de celdas al nodo constituye un proceso discreto dado que cada celda puede llegar únicamente sobre una base de tiempos ranurada. Sin embargo a nivel de tráfico agregado en un nodo de alta jerarquía que se supone que recibe una carga de tráfico alta, la aproximación mediante una cola continua no puede alejarse de la realidad. En cualquier caso debemos

ser conscientes de que la aproximación introducida por el modelo estriba en permitir que el servicio se inicie en cualquier instante de tiempo (continuo) más que en admitir la llegada en cualquier instante. De hecho debido a retardos de orden menor como CDV (*Cell Delay Variation*) en las líneas se van a producir llegadas de forma bastante continua, sin embargo, el servicio de la celda a la velocidad de transmisión del enlace sólo puede iniciarse en los inicios de *slot*. En cualquier caso la aproximación introducida deberá ser cotejada con simulaciones como se hace más adelante.

2.1 Llegadas de Poisson

Si bien la caracterización del tráfico a nivel de fuente (a la entrada del multiplexor) es actualmente objeto de múltiples estudios y discusiones, no puede decirse lo mismo del tráfico agregado que puede estar presente en niveles relativamente altos de la red. Es evidente la tendencia a asumir tráficos de fuente con distribución muy dispersa (MMPP o "*long-tail*") [CHO, FRO, LIP]. Sin embargo, a los conmutadores de nivel alto llega el tráfico ya muy mezclado y con toda seguridad procesado por espaciadores que habrán suavizado la gran dispersión que puede hallarse en algunas clases de tráficos de fuente [MER].

La caracterización del agregado de tráficos basada en herramientas analíticas puede resultar tremendamente compleja por mezclarse tráficos deterministas tipo CBR (NxD) con tráficos estadísticos tipo VBR que han probado poseer ciertas características fractales y de autosimilitud. La caracterización en base a medidas es detectada como altamente necesaria, pero resulta difícil en las circunstancias actuales encontrar una red con alta carga MTA, que a la vez pueda considerarse que lleva una combinación de tipos de tráfico típica, y que pueda ser objeto de medidas no confidenciales. Aunque existen algunas caracterizaciones como [MOL] que muestra ciertas propiedades de autosimilitud en tráfico MTA sobre WAN (*Wide Area Network*), estamos todavía lejos de poder generalizar tales resultados.

Ante esta situación el proceso de llegadas de Poisson resulta cuando menos una hipótesis razonable. Ante la ignorancia de la realidad el tráfico de Poisson no puede considerarse ni mucho ni poco disperso y por su característica de "proceso sin memoria" permite obtener conclusiones que de otro modo resultarían muy complejas o imposibles.

2.2 Múltiples servidores

El modelo con múltiples (s) servidores permite considerar el caso más general de múltiples enlaces entre los nodos [ALT, GAL]. Tal como se ha mencionado en la introducción, hasta la fecha se han publicado multitud de estudios de evaluación del multiplexor MTA asumiendo diferentes tipos de tráfico a la entrada. Dichos trabajos se han orientado en la mayoría de las ocasiones al Control de Admisión de Llamada (CAC) comprobando si existe la posibilidad de aceptar una nueva llamada sobre la única línea de salida del multiplexor manteniendo las calidades de servicio de todas las llamadas existentes y de la nueva.

En niveles altos la red MTA tiende a ser jerárquica y mallada en contraposición a lo que ocurre en el acceso donde la topología es en estrella. A diferencia de lo que ocurre en los multiplexores, a la salida del nodo de conmutación pueden existir diversos enlaces con el mismo nodo remoto. También es probable que existan diversas conexiones con diversos nodos remotos y que dadas las condiciones de evaluación desconozcamos qué tráfico va a cada remoto, lo más razonable es entonces asumir que cada enlace recibe una parte alícuota del tráfico. En cualquier caso es importante considerar que el modelo M/D/s solamente será útil cuando los enlaces mencionados compartan el mismo *buffer* para el acceso a cualquiera de ellos.

2.3 Cola de capacidad limitada

La capacidad del *buffer* que posee el nodo conmutador es un elemento a tener en cuenta e indispensable en el modelo para la evaluación de la calidad del servicio. En el caso de tráficos deterministas como puedan ser los CBR es necesario un *buffer* de capacidad muy limitada, y se prefiere descartar un número algo mayor de celdas que elevar el CDVT. Desde esta perspectiva hay que observar que el CDVT es proporcional a la capacidad del *buffer* en caso de que sólamente exista un tráfico determinista. Así el máximo retardo que puede sufrir una celda es igual a la capacidad del *buffer B* multiplicada por el tiempo que tarda en transmitirse una celda y dividido por el número de servidores (despreciamos otros tiempos de orden menor presentes en el conmutador). Este es el CDVT ya que el mínimo retardo que sufre la celda en el nodo es nulo, en caso de que el *buffer* esté vacío cuando llega la celda. Nótese que en cualquier caso nos estamos refiriendo únicamente al CDVT introducido por el nodo en cuestión, que no debe ser confundido con el CDVT extremo a extremo para el tráfico que se considera ya que éste puede atravesar múltiples secciones. Para calcular el CDVT global pueden usarse herramientas basadas en redes de colas [CHO].

A nivel de CAC es habitual la hipótesis de que el *buffer* tiene capacidad nula ("*bufferless*" cuando se trata con tráficos deterministas. Obviamente esta

hipótesis representa un peor caso que tiene en cuenta el hecho de que el *buffer* es muy pequeño. En cualquier caso esta hipótesis puede dar lugar a evaluaciones excesivamente conservadoras.

Para tráficos estadísticos, sobre todo ABR, suele asumirse un modelo con cola infinita. Con dicho modelo aparece CLR=0 al no existir pérdidas. Es habitual cuando se trata con estos tráficos la aproximación de la probabilidad de congestión por el fractil de la capacidad del *buffer*. Es decir se analiza la cola infinita y se identifica la probabilidad de congestión con la suma de probabilidades de todos los estados de ocupación del *buffer* por encima de su capacidad (estados que sólamente existen en el modelo pero no en la realidad). Si las pérdidas son realmente muy bajas la aproximación mencionada es extremadamente precisa [DES].

En esta ponencia se utiliza esta última aproximación para la evaluación de la media y varianza del número de unidades que ocupan el *buffer* (sección 4). Para la probabilidad de congestión se ofrece una aproximación empírica de modo que no es necesario realizar ninguna hipótesis extrema sobre la capacidad del *buffer*. De cualquier modo se comprueba numéricamente que la aproximación sólamente es válida cuando la capacidad del *buffer* al menos dobla el número medio de celdas que contiene.

3. Media y varianza del número de celdas en el *buffer*

La primera aproximación a la evaluación de la calidad del conmutador nos la proporcionará el valor medio del número de celdas que ocupan el *buffer*. La relación de dicho valor medio con la capacidad nos ofrece una primera idea intuitiva del orden de magnitud de las pérdidas. Dado que en nuestro modelo consideramos el CDVT y el tamaño del *buffer* B proporcionales, es claro que a menor CDVT mayor será el CLR por ser menor B. La varianza nos ofrece la perspectiva de la dispersión, de modo que a mayor varianza o coeficiente de variación cabe esperar un peor funcionamiento aun con el mismo valor medio. Es bien conocido que desde un punto de vista de evaluación los procesos con menor dispersión, tanto en llegadas como en duración de los servicios, proporcionan mejores funcionamientos.

3.1 Valor medio

Es bien conocido que en una cola infinita M/M/s los valores medios del número de unidades en la cola y el tiempo de espera en cola son respectivamente:

$$\overline{Q}_M = PD\frac{\rho}{1-\rho}$$
$$\overline{W}_M = PD\frac{d}{s-A} \qquad (1)$$

donde PD representa la probabilidad de demora que puede calcularse como Erlang-C (ρ, s) y $\rho=A/s$ representa la carga por servidor del sistema (A es el tráfico total ofrecido al conmutador). La duración media del servicio, es este caso fija, se representa por d y es el tiempo que tarda la celda en ser transmitida.

En nuestro caso los datos son el número de enlaces s y la velocidad de transmisión v_t de cada uno de ellos. Si conocemos la tasa a la que llegan las celdas λ (celdas por segundo), la velocidad binaria de llegada es obviamente $53x8x\lambda$. El tráfico ofrecido al conjunto de enlaces en Erlang es la tasa de llegadas multiplicada por la duración media del servicio, es decir $A= 53x8x\lambda/v_t$. Dicho tráfico puede ser visto también como la velocidad binaria de llegada respecto a la de un enlace.

Para la cola infinita M/D/s es conocida la siguiente aproximación para el número medio de unidades en cola [KIM].

$$\overline{Q}_D = \overline{Q}_M \times F$$
$$F = \frac{1}{2}\left\{1+\frac{(1-\rho)(s-1)\left(\sqrt{4+5s}-2\right)}{16\rho s}\right\} \qquad (2)$$

Del mismo modo y en base a la relación de Little, los tiempos medios de espera en la cola con servicio determinista y exponencial están también relacionados mediante F.

Esta excelente aproximación proporciona errores relativos del orden del 1% si la carga no es muy baja. Para cargas muy bajas la aproximación es inconsistente: el número medio de unidades en la cola con servicio determinista aparece mayor que en la cola con servicio exponencial, lo cual no es posible. En tal caso de carga muy baja debe utilizarse una aproximación más burda, pero consistente en todo el rango de cargas posibles [BAR]:

$$F = \frac{(1-\rho)s}{s+1}+\frac{\rho}{2} \qquad (3)$$

En cualquier caso no estamos considerando la capacidad limitada del *buffer*, tal como se menciona en la sección 2.3. Si deseamos precisar algo más el valor medio obtenido, una primera aproximación consiste en descontar del tráfico ofrecido las pérdidas que se calculan en la siguiente sección 4, y recalcular el valor medio del número de celdas en el nodo.

3.2 Varianza

Para tener una primera idea de la dispersión del número de celdas que ocupan el *buffer* con respecto a la media obtenida en el apartado anterior debemos acudir al segundo momento.

Para la cola infinita M/D/s el segundo momento ordinario puede calcularse de forma aproximada a partir del resultado para la cola M/G/s de [TIJ, HOO] que particularizado para el caso de distribución determinista del servicio resulta:

$$\sigma^2 = \frac{\rho^2}{1-\rho}\left\{\left(\frac{2(1-\rho)s^2}{s^2+3s+2}+\frac{\rho}{3}\right)PD+\overline{Q}_D\right\}-\left(\overline{Q}_D\right)^2 \qquad (4)$$

Al igual que ocurría con el valor medio, puede obtenerse una mejor aproximación descontando las pérdidas calculadas en la siguiente sección del valor de la carga utilizado en (4).

4. Probabilidad de congestión

Dado que el modelo utilizado representa una cola con llegadas de Poisson, la probabilidad de congestión debe coincidir con el CLR en base a la propiedad PASTA ("*Poisson Arrivals See Time Average*"). En este caso es forzoso tener en cuenta el tamaño limitado del *buffer* ya que de lo contrario no puede calcularse el CLR que aparece como nulo. Algunas aproximaciones utilizadas para cálculos en los CAC se basan, tal como hemos comentado, en el fractil de la distribución del número de unidades en cola asumiendo cola infinita, es decir se calcula la suma de todas aquellas probabilidades de estado por encima de la capacidad máxima del *buffer*.

La propuesta que presentamos para el cálculo de dicha probabilidad de congestión está basada en el método citado pero es puramente empírica, si bien está basada en algunos supuestos que resultan claramente intuitivos.

4.1 Cotas de las probabilidades de estado

Se sabe que el número de unidades en una cola infinita M/M/*s* sigue una distribución geométrica de parámetro ρ . Sin embargo, en el caso de la cola infinita M/D/*s* la relación entre la probabilidad de dos estados consecutivos *j* y *j*-1 es:

$$\lim_{k\to\infty}\frac{pj}{pj-1} = \tau < \rho \qquad (5)$$

donde el cálculo de τ supone resolver una ecuación trascendente [TIJ].

En cualquier caso resulta intuitivamente razonable que la cola (finita o infinita) M/D/*s* se comporte mejor que la M/M/*s*, y de (5) se desprende la misma conclusión de manera analítica aunque sólo para el caso de cola infinita.

Tenemos una cota superior para las probabilidades de estado en la cola infinita M/D/*s* que son las probabilidades para la M/M/*s*. También una cota inferior que sería la de la M/M/*s* con una carga τ, inferior ya que τ solamente vale para estados con un número muy alto de unidades en el sistema, mientras que para menos unidades el parámetro estará entre ρ y τ.

4.2 Carga equivalente

Nuestra tarea consiste ahora en encontrar una carga entre las dos cotas citadas que proporcione una buena aproximación para la probabilidad de congestión. Como valores candidatos a ser comprobados de forma numérica están aquellos valores de carga que en una cola infinita M/M/*s* proporcionarían los valores medios de número de unidades en cola y tiempo medio de espera obtenidos para la M/D/*s*. Si consideramos los tiempos medios condicionados a la situación de demora, lo cual supone dividir la ecuación (1) por *PD*, obtenemos las dos siguientes cargas equivalentes:

$$\rho_{eq}{}^{Q} = \frac{\rho F}{1-\rho-\rho F} \qquad (6)$$

$$\rho_{eq}{}^{W} = 1-\frac{1-\rho}{F} \qquad (7)$$

donde los superíndices *Q* y *W* representan las cargas equivalentes que igualan la media del número de unidades y el retardo condicionados respectivamente.

4.3 Estimación del CLR

Dado que se asume un proceso de llegadas de Poisson el CLR coincide con la probabilidad de congestión, es decir con la probabilidad de que el *buffer* esté lleno. La estimación que proponemos y que es validada numéricamente en la siguiente sección consiste en identificar la probabilidad de congestión de la cola finita con la probabilidad de que el sistema se encuentre en el estado *s*+*B*, es decir, todos los servidores ocupados y el *buffer* lleno. Dicha probabilidad se computa para una cola infinita M/M/*s* con carga equivalente.

De este modo podemos escribir la probabilidad de congestión identificada con el CLR como:

$$CLR = PD(1-\rho_{eq})\rho_{eq}{}^{B} \qquad (8)$$

La carga equivalente que debe utilizarse es la que iguala el número medio de celdas en el *buffer* de la ecuación (6), y que ha probado dar resultados más exactos.

5. Resultados numéricos

5.1 Simulaciones

Debido a que la aproximación (8) es empírica, el error en el que se incurre sólamente puede ser estimado mediante pruebas numéricas que cotejen los resultados de simulación con los cálculos propuestos. Dichas pruebas deben realizarse en un margen amplio de valores de carga, número de canales y capacidad del *buffer*.

Para mostrar la utilidad de la estimación presentada en el entorno MTA hay que trabajar con valores que puedan acercarse a la realidad de los sistemas MTA, lo cual implica trabajar con situaciones de CLR siempre inferiores a 10^{-6}. Para ello es necesaria la simulación de eventos raros [FRO] que no se ha realizado en los resultados que se presentan: hemos trabajado con CLR de hasta 10^{-6} (procesando 6 millones de celdas en cada simulación). De este modo hemos probado que el ajuste es bueno pero somos conscientes de que se precisan simulaciones adicionales. Aunque la generalización de los resultados obtenidos para CLR<10^{-6} suponen una conjetura razonable, ésta debe ser probada.

Por otra parte y tal como se ha mencionado la simulación realizada permite el inicio de un servicio (transmisión de una celda) en cualquier instante de tiempo, es decir se ha realizado una simulación sobre tiempo continuo en lugar de discreto. Aquí es válido el mismo argumento del párrafo anterior: es una conjetura razonable pensar que apenas cambiarán los resultados, pero debe ser probada mediante simulaciones adicionales.

5.2 Comparación con el método propuesto

En la Figura 2 se representa un caso con 5 enlaces y tamaño del *buffer* variable, utilizando la carga por canal como parámetro. Debe entenderse que el ancho de banda o velocidad total a la entrada del nodo es:

$$v_i = s \times \rho \times v_t \qquad (9)$$

Esta velocidad es obviamente una media ya las llegadas de celdas se producen en base a un proceso de Poisson. Debe mencionarse que en las pruebas realizadas el error relativo en que se incurre es de un máximo de un 25%, siempre que se consideren valores por encima del número medio de celdas que ocupan el *buffer*.

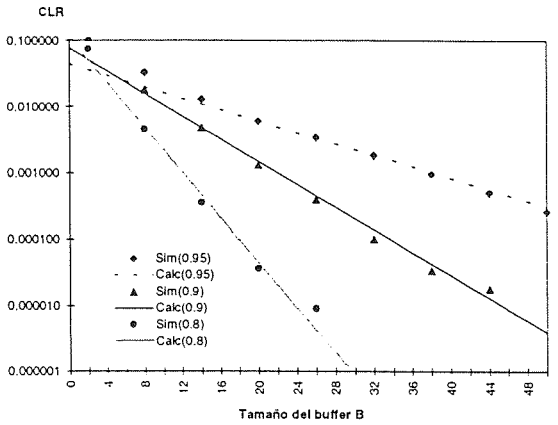


Fig.2 : CLR calculado según la estimación propuesta y resultado de simulación (5 enlaces).

La Figura 3 muestra también el CLR calculado y obtenido mediante simulación para un caso con 10 enlaces variando la carga y utilizando como parámetro la capacidad del *buffer*. Obsérvese que las velocidades de entrada al nodo y de transmisión de cada enlace influyen en el CLR sólo en la medida de la proporción en que se encuentran (a través de ρ y s). Lo mismo ocurre con el tiempo que tarda en ser servida una celda, que no influye en el CLR ya que la capacidad del *buffer* está fijada en número de celdas. Las gráficas, por tanto están normalizadas con respecto al tamaño de una celda.

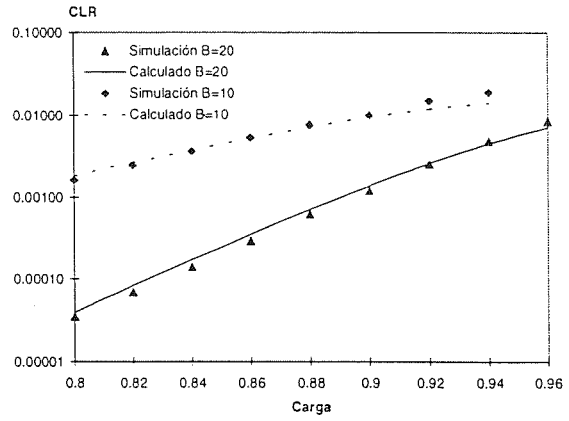


Fig. 3: CLR en función de la carga para un nodo con 10 enlaces de salida.

6. Diferenciación mediante prioridades

En esta sección se propone diferenciar los tráficos determinista y estadístico mediante un sistema de prioridades HOL (*Head of Line*). Este método de prioridad sin interrupción ubica una celda prioritaria en el nodo por delante de todas las celdas con nivel de prioridad inferior y detrás de todas las de prioridad superior: mantiene el carácter FIFO de la cola. La utilización de prioridades en el nodo MTA para obtener calidades de servicio diferenciadas para diferentes tráficos ha sido propuesta por diversos autores [CHOU, OUY].

Parece razonable considerar las celdas pertenecientes a tráfico determinista como de mayor prioridad y las de tráfico estadístico de prioridad inferior. Debido a las restricciones de CDVT el tráfico determinista dispone de una cantidad de *buffer* mucho menor que el estadístico y parece lógico priorizar dicho tráfico para evitar CLR excesivos. De forma más general cabe esperar que el nivel de prioridad sea independiente de la característica determinista o no del tráfico y en cambio sea asignado en función de la calidad (CLR y CDV deseados).

Una particularización de lo anterior puede ser el caso con dos tipos de tráfico: CBR y ABR. El tráfico CBR requiere en general un bajo CDVT y por tanto B pequeño, mientras que el ABR es muy poco sensible a retardos y debe poseer espacio en *buffer* B grande para minimizar su CLR. El *buffer* compartido tiene por tanto dos umbrales B_{CBR} y B_{ABR}, uno para cada tipo de tráfico y mayor para ABR (pudiendo considerar éste último como el tamaño total de la cola). Dichos umbrales actúan de manera que si una celda CBR llega y el *buffer* está ocupado por encima del umbral para CBR, ésta es descartada (obsérvese que si no fuera descartada llegaría al enlace con un retardo mayor que CDVT para CBR). La asignación de un mayor nivel de prioridad al tráfico CBR permite en este caso mantener su CLR a un nivel bajo que no sería posible alcanzar sin prioridad.

6.1 Método de estimación

La forma que proponemos de aplicar las estimaciones propuestas en las secciones anteriores al caso de que existan dos niveles de prioridad es extremadamente sencilla y basada en el algoritmo propuesto en [BAR]:

- El número medio de celdas prioritarias en cola se calcula aplicando las ecuaciones (1) a (3) utilizando la probabilidad de demora para el total de carga. Sin embargo la variable ρ se sustituye únicamente por la carga prioritaria. Si la carga prioritaria es baja debe utilizarse (3) en lugar de (2).

- Se obtiene el número medio total de celdas en cola. El número medio de celdas no prioritarias es la diferencia entre el total y el número medio con prioridad.

- El CLR para el tráfico prioritario se calcula aplicando (8) con la probabilidad de demora global y la carga equivalente considerando sólamente tráfico prioritario. El tamaño del *buffer* es el disponible para tráfico prioritario.

- El CLR para el tráfico no prioritario se aplica (8) con el tamaño del *buffer* total.

Debe entenderse que seguimos en el terreno de la aproximación con márgenes de error en torno al 25% que no permiten calificar la estimación de precisa. La propuesta presentada es por tanto una primera estimación de las calidades y una primera visión de diseño, que para nada descarta la aplicación posterior de métodos más precisos y sofisticados de evaluación.

6.2 Resultados numéricos

En la Figura 4 se representan los CLR para tráfico prioritario y normal (p y n en la gráfica) en un nodo con 10 enlaces y carga del 90%. De todas las pruebas realizadas, se presentan aquí los resultados para un tamaño total del *buffer* que es cuatro veces el umbral de tamaño para descartar celdas prioritarias y que el tráfico de cada nivel de prioridad representa el 50% de la carga total.

Se ha utilizado el mismo tipo de simulación que en la sección anterior, esto es simulación en tiempo continuo sin utilizar técnicas de eventos raros. Se han realizado numerosas simulaciones variando los parámetros de carga y proporción en la que se encuentran los tráficos de ambas clases, obteniéndose en todas resultados similares de precisión para la aproximación propuesta.

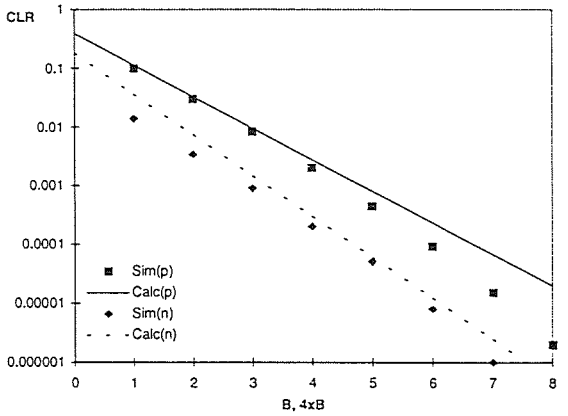


Fig. 4: Nodo con dos tipos de tráfico y prioridad HOL (10 enlaces).

7. Conclusiones

En este trabajo se ha presentado un método para realizar un primera estimación de algunos parámetros de calidad de un nodo de conmutación MTA situado en un nivel jerárquico de multiplexación alto de la red. Las hipótesis de partida (llegadas de Poisson, múltiples enlaces) son adecuadas para modelar nodos de nivel alto recibiendo tráfico agregado. Los parámetros estimados son la media y varianza del número de unidades en cola y el CLR. Dichos parámetros pueden estimarse también en el caso de que haya dos tipos de tráfico con diferentes umbrales para descartar celdas en el *buffer*. Los autores conjeturan que el procedimiento es generalizable a más de dos niveles de prioridad. Campo en el que estamos trabajando actualmente.

Las estimaciones de la media y varianza del número medio de celdas en el *buffer* son precisas mientras que las del CLR son burdas, aunque en todos los casos son de cómputo extremadamente simple (fórmulas cerradas y compactas). Esta facilidad de cómputo las hace aptas para un primer acercamiento al problema de diseño, permitiendo evaluar rápidamente el orden de magnitud de la calidad obtenida. Es obvio que en fases más avanzadas del diseño de la red es necesaria una evaluación más fina, mediante procedimientos más precisos que a la vez requieren un esfuerzo mucho mayor de cómputo.

Agradecimientos

Este trabajo ha sido financiado por el proyecto de investigación SIGLA (CICYT TEL96-1452).

Referencias

[ALT] K. Altinkemer, I. Bose, "Asynchronous Transfer Mode Networks with Parallel Links and Multiple Service Classes", *5th Int. Conf. On Telecommunications Systems*, pp. 594-601, 1997.

[BAR] F. Barceló, V. Casares, J. Paradells, "The M/D/C Queue with Priority: Application to trunked Mobile Radio Systems" *IEE Electronics Letters*, Vol **32**, No. 18, pp. 1644-1645, August 1996.

[CHA] W. C. Chan, E. Geraniotis, "Near-Optimal Bandwith Allocation for Multi-Media Vitual Circuit Switched Networks", *IEEE INFOCOM'96*, 1996.

[CHO] C.-H. Chou, E. Geraniotis, "Efficient Computation of End-to-End Performance Measures for Multi-Link ATM Networks with Multi-Media Traffic", *IEEE INFOCOM'95*, April 1995.

[CHOU] A. K. Choudhury, E. L. Hahne, "Dynamic Queue Length Thresholds for Multipriority Traffic", *ITC'15 International Teletraffic Congress*, pp. 561-570, June 1997.

[DES] E. Desmet, B. Steyaert, H. Bruneel, "Tail Distributions of Queue Length and Delay in Discrete-Time Multiserver Queueing Models, Applicable in ATM Networks", *ITC'13 International Teletraffic Congress*, pp. 1-6, 1991.

[ELW] A. Elwalid, D. Heyman, T. V. Lakshman, "Fundamental Bounds and Approximations for ATM Multiplexers with Applications to Video Teleconferencing", *IEEE Journal on Selected Areas in Communications*, Vol. **13**, No. 6, August 1995.

[FRO] V. S. Frost, B. Melamed, "Traffic Modeling For Telecommunications Networks", *IEEE Communications Magazine*, March 1994.

[GAL] G. Gallassi, G. Rigolio, L. Verri, "Resource Management and Dimensioning in ATM Networks", *IEEE Network Magazine*, May 1990.

[HOO] M. H. van Hoorn, "Algorithms and Approximations for Queueing Systems", *CWI Tract Centre for Mathematics and Computer Science*, 1984.

[LIP] L. Lipsky, J. E. Hatem, "Buffer Problems in Telecommunications Networks", *5th Int. Conf. On Telecommunications Systems*, pp. 556-566, 1997.

[MER] G. Mercankosk, T. Moors, A. Cantoni, "Multiplexing Spacer Outputs on Cell Emissions", *IEEE INFOCOM'95*, April 1995.

[MOL] S. Molnár, A. Vidács, "On Modeling and Shaping Self-Similar ATM Traffic", *ITC'15 International Teletraffic Congress*, pp. 1409-1420, June 1997.

[OUY] T. Ouyang, A. A. Nilsson, "Performance Modeling of an ATM-SMX in the NCIH with CBR, VBR, and UBR Traffic", *ITC'15 International Teletraffic Congress*, pp. 731-740, June 1997.

[SAT] Y. Sato, N. Yamanaka, K. Sato, "ATM Network Resource Management Techniques for CBR Virtual Paths/Channels", *IEICE TRANS. COMMUN.*, Vol. **E79-B**, No 5, pp. 684-692, May 1996.

[TIJ] H. C. Tijms, "Stochastic Modeling and Analysis: A Computational Approach", John Wiley & Sons, 1986.

[VEC] G. de Veciana, G. Kesidis, J. Walrand, "Resource Management in Wide-Area ATM Networks Using Effective Bandwidths", *IEEE Journal On Selected Areas in Communications*. Vol. **13**, No. 6, pp. 1081-1089, August 1995.

Grupo II: Impacto Social de las Tecnologías de Información en la Sociedad

Seguridad

Seguridad en Internet

ANA ARROYO MUÑOZ , IÑAKI ARRIBA GONZÁLEZ DE DURANA
ROBOTIKER, AREA DE TELECOMUNICACIONES
PARQUE TECNOLÓGICO DE ZAMUDIO, EDIF. 202 48170 ZAMUDIO
Correo electrónico: ana@robotiker.es, ignacio@robotiker.es

Abstract:

This paper gives a general approach to build up a Internet security policy, bearing in mind that must be part of the organization's global security. The process has three main phases: designing the policy, technical implementation of defences and finally running or operating mode. The most common defence is the firewall in its variety of architectures. Some of these devices also helps to construct virtual private networks, that extents organization's Intranet further its local limitations.

1. Qué es la política de seguridad y para qué sirve.

El objetivo de diseñar una política de seguridad es proteger los activos de la empresa de los posibles robos, daños, etc., que puedan causar personas ajenas a la organización o de dentro de la organización. La organización debe entender y saber el coste de su información: cómo afecta a la toma de decisiones datos "posiblemente" incorrectos, qué pasaría si la información confidencial de la empresa se hiciera pública, cual es el coste (en tiempo y en credibilidad) de la interrupción del servicio. Para poder valorar las amenazas hay que saber si la información que posee la empresa es valiosa. Merece la pena diseñar y realizar una política de seguridad si los recursos que se quieren proteger son cruciales para la organización.

El riesgo de la conexión a la red Internet queda reflejado en su tamaño: 50 millones de usuarios, 30.000 redes, más de 10 millones de ordenadores en 137 países. Es una comunidad muy numerosa y en aumento. La red Internet no debe ser vista sólo como fuente de información, o como infraestructura de conexión entre sedes de una misma empresa, también hace que la organización sea más visible alcanzable ya que dicha conexión abre otras posibles vías de entrada a la organización (además de las puertas físicas del edificio), lo que supone más oportunidades para ser atacados.

No toda la gente que utiliza Internet tiene las mismas intenciones: hay empresas que hacen marketing de sus productos, personal investigador que comparte información, personas que entran en sistemas accidentalmente y otros que entran con ánimo de bloquear el servicio, obtener información confidencial, etc., que a la postre son los que causan los perjuicios y son de los que hay que defenderse.

La seguridad en Internet debe ser contemplada como parte del plan de seguridad de la empresa, es un aspecto más a tener en cuenta en la securización del sistema de información de la organización. La securización de la infraestructura informática es una parte del sistema de información. Otros aspectos de la seguridad incluyen la seguridad física (control de acceso al edificio, departamentos,..), la gestión de documentos y sus soportes (copias de seguridad), los documentos desechados (para reciclar o en las papeleras), costumbres del personal (copiar información, ..), etc.

Este plan estará plasmado en la política de seguridad, que debe quedar recogida en un documento.

Definir y diseñar una política de seguridad significa desarrollar planes y procedimientos que salvaguarden los recursos de las red contra pérdidas y daños. Debe dejar claro:

- los tipos de recursos y servicios que se van a proteger;
- su nivel de importancia;
- los personas de las cuales se les quiere proteger.

Antes de instalar un cortafuegos o cualquier medida de seguridad hay que identificar los recursos que se quieren proteger, su valor y su vulnerabilidad.

En caso de aplicar medidas de seguridad al personal, hay que considerar que hay que proporcionar acceso transparente sin que dicha política suponga una carga para los usuarios internos, ya que lo contrario puede traer consecuencias no deseadas, como que los propios usuarios se salten dichas medidas.

2. Implantación y puesta en marcha del plan de seguridad.

Las etapas básicas que se deben contemplar para poner en marcha y hacer el seguimiento de la seguridad de la empresa están reflejadas en la Fig. 1. A continuación se realiza una breve descripción de las mismas.

1. Definición de la política de seguridad.
   Tienen que participar áreas y estamentos de toda la organización para que aporten sus necesidades y se sientan involucrados en el proceso global. El

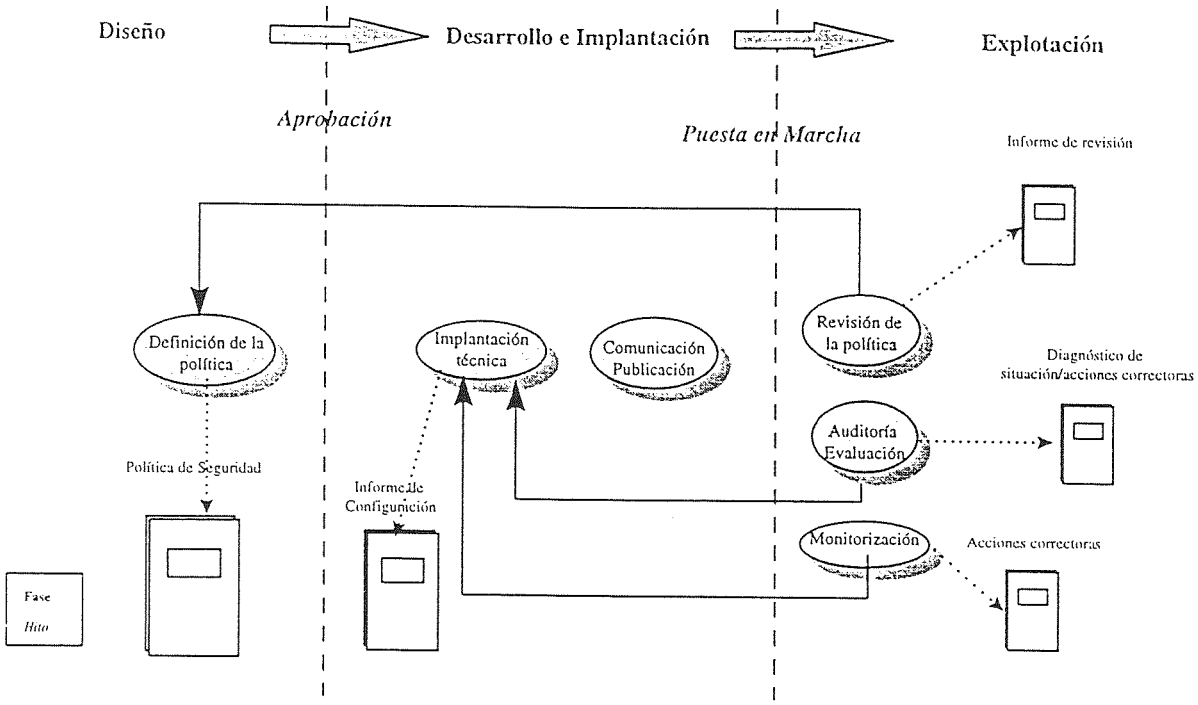


Figura 1. Fases para la puesta en marcha de una política de seguridad

resultado de esta actividad se plasma en el documento "Política de Seguridad". En el apartado siguiente se detallan los pasos a seguir para diseñar esta política.

2. Obtener la aprobación de la política.
   Una vez definida la política, debe estar respaldada por los órganos de decisión de la empresas ya que afectará a todos los empleados.

3. Implantación tecnológica de la política.
   La política se define en términos de alto nivel. La implementación técnica es la instalación del software y hardware y procedimientos que se han definido para soportar la política aprobada, es decir, se instalarán y configurarán las medidas que protejan nuestros activos. Es conveniente recoger la información de la implementación de la política e ir actualizándola según se vayan produciendo los cambios en la configuración.

4. Diseminación de la política.
   El objetivo de esta fase es publicar la política y hacerla llegar a todos los empleados a los que les afecte. Una forma directa es organizar charlas internas para explicar el porqué de los nuevos procedimientos, las responsabilidades de cada usuario, las sanciones en caso de no seguir las normas, etc.

   La comunicación de la puesta en marcha de la política debe ir acompañada de un calendario en el cual se indique la duración del periodo de transición y la fecha de arranque *real*. Es necesario un periodo de adaptación y concienciación a las nuevas formas y procedimientos de trabajo. Además, se podrán corregir las desviaciones, detectar aspectos que no se han considerado, etc.

5. Revisiones periódicas.
   Una vez puesta en marcha la política de seguridad, habrá que realizar revisiones periódicas del plan para ver si sigue en consonancia con los objetivos de la empresa o si ha habido cambios que deben quedar reflejados en el plan. La periodicidad puede ser anual, y el resultado de este trabajo será un informe diagnóstico de la nueva situación, a partir del cual se deberá modificar la política de seguridad según corresponda.

6. Auditorias/Evaluación continúa a los mecanismos de seguridad implantados.
   Además de llevar a la práctica la política de seguridad definida en el plan, hay que estar probando continuamente que las medidas de seguridad implantadas en la organización son fiables. Se deben contemplar métodos y herramientas para comprobar dicha seguridad. La seguridad de los sistemas se probarán con herramientas para redes y de hackers. Las pruebas deben representar ataques que puedan ocurrir desde dentro y fuera de nuestras defensas. Para los ataques externos se puede usar herramientas SATAN(System Administrator's Tool for Analyzing Networks), COPS (Computer Oracle and Password System) y CRACK. Los

ataques internos pueden ser a través de líneas de teléfono, etc.
El resultado de esta auditoría debe ser un informe que corrobore la validez de las medidas de seguridad, o que recoja los problemas de seguridad detectados y las acciones pertinentes para su corrección.

7. Monitorización del sistema.
El objetivo de la monitorización de la red es detectar cualquier actividad inusual dentro del sistema. Los accesos no autorizados o simplemente los intentos, hay que detectarlos cuanto antes, por ello, hay que monitorizar la actividad del sistema de manera periódica. La frecuencia debe ser diaria (durante todo el día), ya que si se hace semanal o mensualmente se pueden producir las intrusiones durante el intervalo de monitorización.

3. Diseño y definición de la política de seguridad.

Un comienzo para abordar el diseño de la política de seguridad es examinar los siguientes aspectos:

- Qué tipo de recursos se tratan de proteger.
- De qué tipo de gente hay que proteger los recursos y cómo de vulnerables son.
- Cómo de probables son los ataques o amenazas.
- Cómo de importante es el recurso.
- Periódicamente revisar si hay cambios organizativos que deban quedar reflejados en la política de seguridad.

Las actividades que se deben realizar para obtener un plan de seguridad son :

- Identificar los recursos
- Identificar las amenazas
- Analizar los riesgos
- Determinar el uso de la red y las responsabilidades de los usuarios
- Identificar y prevenir problemas de seguridad
- Política de control y monitorización del sistema.

Es importante implicar a la gente adecuada en el diseño de la política. El grupo de trabajo no debe estar formado sólo por técnicos, que son los que conocen los detalles para implementar la política, deben participar personas con autoridad que avalen los resultados del grupo. Además, deben participar otras áreas que actualmente tengan a su cargo tareas relacionadas con seguridad (gestión de la información confidencial de productos, etc.).

La política debe tener, como se ha mencionado, el respaldo y aprobación de la Dirección.

También hay que dejar claro cuales son las responsabilidades de cada uno en el mantenimiento de la seguridad del sistema. El CERT estima que el 80 % de los problemas de seguridad de la red son por que se eligen palabras de paso inseguras. Por ejemplo, cada usuario será responsable de elegir una palabra de paso segura (se deberá indicar qué normas se deben seguir para tener una palabra de paso segura), ya que un mal uso de la misma puede comprometer la seguridad de todo el sistema. También deben quedar claras las responsabilidades del administrador del sistema.

3.1 Identificación de los recursos.

Es importante identificar todos los activos que se pueden ver afectados por problemas que impidan su disponibilidad, confidencialidad e integridad.

Según la RFC 1244 [1] los recursos del sistema son los siguientes, entre otros:

- hardware: procesadores, tarjetas, teclados, terminales, estaciones de trabajo, ordenadores personales, servidores, impresoras, discos, líneas de comunicación, servidores de terminales, dispositivos de interconexión como gateways, routes, bridges, repetidores;
- software: aplicaciones con sus programas fuente, objeto, utilidades, programas de diagnóstico, sistemas operativos, programas de comunicaciones;
- datos: durante la ejecución, almacenados y archivados, copias de seguridad, registros de las auditorías, bases de datos, en tránsito en líneas de comunicación;
- personas: empleados locales, remotos, proveedores, usuarios casuales, etc.
- documentación: de programas, hardware, de sistemas, procedimientos administrativos locales.

Toda esta información se puede recoger, como propone Karanjit [2], en una hoja de trabajo que tenga los siguientes campos:

- Recursos de red: numerados y con una descripción y la importancia que se le da. No es lo mismo proteger catálogos comerciales que se suponen están a libre disposición, que información sobre el plan estratégico de la empresa. Este punto en cuanto a información pero no hay que olvidar el acceso al router, o recursos de disco, etc.
- Tipo de usuarios de los cuales hay que proteger el recurso: usuarios externos, internos, por áreas de la empresa, etc.
- Posibilidad de ataque al recurso: alto, bajo, medio, etc. Una palabra que estime la importancia que tiene la información o los recursos.

- Medidas que se pueden implementar para proteger los recursos: se pueden aplicar medidas como permisos que proporciona el sistema operativo para los ficheros y directorios, auditorias y alertas para los servicios de red, routers y cortafuegos para proteger a servidores y dispositivos de red.

3.2. Identificación de las posibles amenazas.

Una vez identificados los recursos que es necesario proteger, se deben identificar las amenazas a dichos recursos. Existe una parte importante de los ataques que se producen desde dentro, principalmente por empleados resentidos o despedidos. Otro ataque interno, no voluntario, es "social engineering", está muy extendido y es efectivo. Se investiga una organización y se usa la información para engañar a los empleados o administradores para conseguir su permiso de acceso.

El protocolo IP no fue diseñado pensando en su seguridad, los paquetes se envían sin ninguna consideración de autenticación ni de confidencialidad, esto supone que un paquete capturado puede se legible y son esa información se puede conocer el sistema. La amenazas mas comunes son:

- acceso no autorizado (benignos, accidentales, malignos): todo acceso a cualquier recurso sin permiso previo, el más común es utilizar la cuenta de un usuario válido
- revelación de información: puede ser voluntaria o involuntaria de información confidencial.
- negación del servicio: cuando se bloquea la máquina por causas que son difíciles de prever a priori. Por ejemplo la red se puede bloquear por tráfico masivo, un virus que se propaga y consume tiempo del procesador.

En cualquier caso se deben determinar qué servicios son absolutamente esenciales y qué efecto puede tener su pérdida. Se deben tener planes de contingencia para recuperarse de dichas pérdidas.

3.3. Análisis de riesgos.

Con este análisis se trata de evaluar qué recursos de la red merecen la pena proteger y cuáles son más importantes que otros. Los riesgos se ordenan por su nivel de importancia. El análisis de riesgos es la base del plan de seguridad.

Para el análisis se realizarán las siguientes estimaciones:

- estimación del riesgo de perder el recurso (en base a lo vulnerable que sea): se puede cuantificar con un valor del 0 - no riesgo- al 10 - riesgo alto.
- estimación de la importancia de este recurso: lo mismo.
- disponibilidad: se trata de medir la importancia de la disponibilidad del recurso en el tiempo.
- integridad del recurso: si los datos son consistentes - sobre todo las bases de datos.
- confidencialidad: para los ficheros de datos a los cuales se quiere restringir el acceso.

Karanjit [1] propone una fórmula para valorar el peso y la importancia que se le da al recurso:

$$PRi = Ir * Pi$$

donde

Ri= Riesgo del recurso i
Pi= Importancia del recurso
PRi= El peso del recurso dentro del sistema

La valoración de los activos puede ser numérica o también se puede hacer en lenguaje natural: alto bajo, medio, etc.

Con esta información ordenamos los recursos por orden de importancia y valoraremos el coste de su defensa.

3.4. Identificación del uso de la red y responsabilidades.

En esta apartado hay de identificar aspectos que son importantes para el plan de seguridad como son los derechos y obligaciones de los usuarios.

- quién está autorizado para utilizar recursos y servicios: se trata de hacer una lista con los usuarios o grupos de usuarios del sistema. Serán usuarios internos, externos, con determinados privilegios, etc. Por ejemplo, "Todos los usuarios solo podrán utilizar los servicios de Internet siguientes: accesos a servidores Web y correo electrónico", en este caso quedarían prohibidos la utilización de telnet, X11, etc.
- determinar las responsabilidades de los usuarios y del administrador del sistema, es decir, definir e identificar que se entiende por uso correcto de los recursos: se trata de definir el AUP (Accepable Use Policy) para la red. En este apartado se debe definir claramente que los usuarios son responsables de sus acciones y de la información que manejan. (No tiene sentido poner un cortafuegos potente si luego un usuario coge un disquete y se lo da a un usuario no autorizado). Debe dejar claro que no se puede entrar en cuentas ajenas y que saltarse la seguridad está prohibido. La idea es poner por escrito que: NO está permitido entrar en cuentas ajenas sin permiso, averiguar palabras de paso, interrumpir el servicio, leer ficheros que no tengan permiso de lectura para todo el mundo, modificar ficheros de los cuales no son propietarios. También se puede referir a las

licencias o copyright del software no permitiendo duplicidad de uso de licencias.

- determinar qué tipo de privilegios: se identificará quién puede otorgar privilegios (para saber que privilegios se dan y a quién). En un sistema complejo el administrador puede utilizar una hoja con la siguiente información (además sirve para recoger las peculiaridades del sistema): número de recurso con el que se ha identificado en la descripción de recursos, descripción del recurso, descripción del tipo de acceso al recurso (ej.: lectura a directorios), permisos del sistema operativo de lectura, escritura, ejecución (tipo Unix)
- por último, deben indicarse los procedimientos a seguir en caso de que, como seguro que sucederá, alguien haga caso omiso de la política.

3.5. Identificar y prevenir problemas de seguridad.

La política de seguridad define qué es lo que se debe proteger pero no cómo deben protegerse los recursos. Debe existir una sección dentro del documento que hable de los procedimientos para prevenir problemas de seguridad.

Además de realizar el análisis de riesgos sobre los recursos, hay que identificar qué otras áreas son vulnerables dentro de las que se mencionan a continuación:

- Puntos de acceso: puntos de entrada utilizados por usuarios no autorizados. Estos pueden ser servidores de terminales, routers, estaciones de trabajo que tengan modems conectados, etc. En este último caso si la conexión se realiza para sesiones telnet (que no utilicen TCP/IP) puede no existir problema, pero si utilizan SLIP o PPP se está creando un punto de acceso que representa un riesgo para toda la red.
- Errores (Bugs) de software: los intrusos se pueden aprovechar de errores que haya en el sistema operativo, sobre todo en Unix que es un sistema muy conocido en su codificación. Los administradores del sistema deberán encargarse de obtener los parches necesarios para subsanar estos errores, y además es recomendable que estén en listas de correo donde se publican los fallos que se van detectando en diversos programas (no solo los del sistema operativo, si no de todas las herramientas que se tengan instaladas).
- Amenazas internas: los de la organización tienen fácil acceso al software de red, por lo que pueden poner patas arriba la seguridad. Se puede capturar lo que va por la red con un analizador y ver la palabras de paso que se envían en los servicios como telnet, ftp, rlogin, etc.
- Seguridad física: tiene que ver con la localización física del servidor/recurso. Si el ordenador en sí no está seguro (una habitación cerrada o de acceso restringido), el mecanismo software de seguridad puede ser fácilmente saltado.

3.6. Política de control y monitorización.

Los controles que se seleccionan son la primera línea de defensa en la protección de la red. En el análisis de riesgos se habrán definido cuales son los objetivos a proteger. Si la mayor amenaza son los usuarios externos, se puede pensar en soluciones tipo cortafuegos o router. Si la mayor amenaza es el acceso no autorizado de los usuarios internos, se puede pensar en establecer procedimientos de accesos automáticos y seguros.

Si del análisis de riesgos se desprende que un recurso es crítico para el funcionamiento de la red se pueden utilizar múltiples estrategias para protegerlo (defensa en profundidad) y así en caso de que una de las barreras sea atravesada o falle, la siguiente está todavía activa.

Los accesos no autorizados o simplemente los intentos, hay que detectarlos cuanto antes, por ello hay que monitorizar la actividad del sistema de manera periódica. La frecuencia debe ser diaria (incluso durante todo el día se pueden ir lanzando ciertos comandos), ya que si se hace semanal o mensualmente se pueden dejar las brechas abiertas durante el intervalo.

Para monitorizar la seguridad de la red se utilizan utilidades del sistema operativo o herramientas comerciales o de dominio público. También se pueden desarrollar utilidades propias. Todos los sistemas operativos almacenan información acerca de los accesos en ficheros especiales. Por lo general estas utilidades se fijan en las siguientes cosas:

- una cuenta que tiene actividad fuera de las horas normales puede significar que esté siendo utilizada por un intruso,
- buscar gran número de intentos de entrada fallidos en un periodo corto de tiempo, esto indica que han intentado adivinar las claves de acceso (en Unix se guardan en el fichero syslog). Sobre todo si son del administrador del sistema.
- utilizar programas del sistema para detectar programas que se estén ejecutando,
- los cortafuegos producen ficheros que recogen todos los accesos a la red.

Existen herramientas de libre distribución de control y seguimiento de accesos y otras que chequean la integridad del sistema.

Las primeras, permiten tener una información (mediante ficheros de trazas) de todos

los intentos de conexión que se han producido sobre un sistema (el que se señale) y ataques de forma sistemática a puertos TCP y UDP. Permiten tener un control sobre todos los paquetes que entran por el interfaz de red de la máquina -IP (TCP,UDP) e ICPM- , y analizar los paquetes a nivel de aplicación -Telnet, ftp, smtp, login, shell, etc.). Los programas más comunes son: tcp-wrapper, netlog, argus, tcpdump, satan, iss, courtney, gabriel, nocol, tcplist.

Las herramientas que controlan la integridad de los datos y programas del sistema se basan en chequeos a los ficheros y el envío de alarmas de posibles modificaciones de ficheros y de programas "sospechoso" que puedan estar ejecutándose en la máquina de forma camuflada. Los programas más comunes son: COPS, Tiger, Crack TripWire, chkwtmp y chklaslog, CMP, etc.

3.7. Contenido de la política de seguridad.

Además de contener los puntos que se mencionan a continuación, el documento debe tener las siguientes características: ser claro y preciso, fácil de entender y leer, organizado.

1. Descripción general de los recursos que se han considerado críticos para la empresa, campus, etc.
2. Alcance de la política, quién está sujeto a la política de seguridad y en qué términos, las responsabilidades de cada usuario de los recursos.
3. Procedimientos para actuar en caso de ataque/incidentes desde dentro o fuera de la organización.
4. Procedimientos para actuar en caso de que los usuarios se salten la política de seguridad, es decir, sus responsabilidades.
5. Procedimientos para prevenir problemas de seguridad. Dispositivos y política de control y monitorización.

4. Cortafuegos.

Los cortafuegos son elementos que aumentan la seguridad del sistema y reducen los riesgos de la conexión a Internet. Un cortafuegos es un sistema o grupo de sistemas que refuerza la política de control de accesos entre dos redes. Esta tarea se puede llevar a cabo basándose en dos mecanismos distintos: bloquear o permitir el tráfico en la red. Además, pueden llevar un registro tanto de las entradas como de las salidas de la actividad en Internet.

Los cortafuegos proporcionan una protección segura contra un ataque externo, se colocan como una barrera entre la red interna de la organización y la red exterior. Actúan como un cuello de botella por el que debe cruzar todo el tráfico procedente del exterior y sólo el tráfico autorizado pasará a través de la barrera.

Los beneficios principales de los cortafuegos son:

- permite al administrador controlar y monitorizar el acceso que se hace a los recursos de la red interna desde el exterior.
- simplifica la gestión de la seguridad de la red gracias a concentrar y combinar la mayoría de las funciones de seguridad en un único dispositivo con una interface única. Se deberán tomar las medidas seguridad necesarias para este dispositivo.

Los cortafuegos no pueden proteger la red de ataques que no pasen a través de ellos. Muchas organizaciones con estrictos sistemas de seguridad en las conexiones a Internet, han descuidado por completo medidas de seguridad elementales como: control de cintas y discos magnéticos, accesos dial-in y dial-out vía modem, etc.

Es conocido que las amenazas internas suponen un mayor peligro ya que estos usuarios disponen de los medios, el móvil y la oportunidad. Contra estas acciones el cortafuegos es ineficaz.

Tampoco pueden detectar nuevas amenazas para las cuales no han sido diseñados, como son nuevos protocolos, virus, etc.

Un cortafuegos no es la panacea de la seguridad, debe ser sólo una parte más de la arquitectura de seguridad de toda la organización.

4.1. Tipos de Cortafuegos.

Existen cortafuegos que trabajan a nivel de red (routers) y los que trabajan a nivel de aplicación. Esta división cada vez es más difusa ya que hay dispositivos que se basan en el filtrado dinámico de paquetes y que mantienen información sobre el estado de las conexiones que le atraviesan y del contenido de los paquetes para aprobar o rechazar el paso de los paquetes. Conceptualmente este modelo es un híbrido de los dos anteriores. Para el usuario final parece que se limita a funcionar en el nivel de red, pero en realidad examina todo el tráfico que lo atraviesa.

El modelo más simple de defensa es el **Screening Router** que opera en el nivel de red y basa sus actuaciones en el contenido de las cabeceras de los paquetes TCP/IP. El router analiza la cabecera de cada paquete de datos que recibe y decide, en función de unas reglas de filtrado - que habremos definido de acuerdo a nuestra política de seguridad - si deja pasar el paquete o lo rechaza.

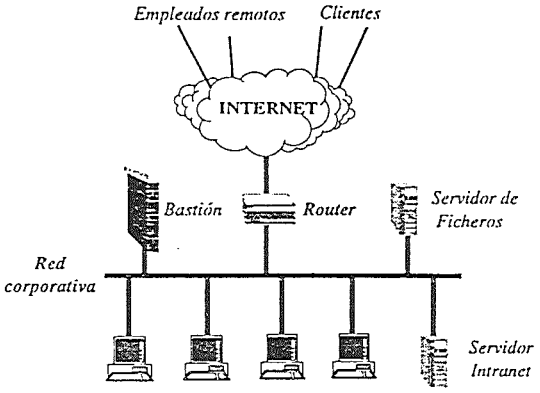


Figura 2. Screened Host screenning router

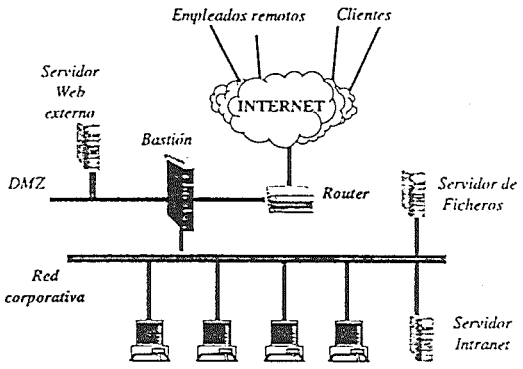


Figura 3. Bastión con 3 interfaces de red y

Esta es la forma más rápida, flexible y barata de cortafuegos, aunque también la más vulnerable a ataques. Son incapaces de filtrar algunos protocolos, por lo general no contemplan mecanismos de alerta del sistema y de auditoría y requieren un conocimiento profundo de los protocolos de comunicaciones.

Debido a estas limitaciones, los routers suelen estar complementados con otros dispositivos que trabajen a nivel de aplicación y que poseen información completa del tráfico de red, como por ejemplo un "bastión" o un "dual-homed host". **Bastión** es el punto de entrada y salida de la red y es crítico para la seguridad del sistema. Debe estar bien protegido y someterlo a auditorias regularmente. Esta configuración da lugar a lo se denomina **Screened Host** (ver fig. 2).

En este caso el router estará configurado para recibir y enviar paquetes desde el ordenador bastión, no permitiendo otro tipo de enrutamiento. Si el router es atacado con éxito, y se consigue que se envíen paquetes a otro destino que no sea el bastión, la red corporativa queda comprometida.

El siguiente modelo de defensa, es el **Dual-Homed Host**, que proporciona un mayor grado de aislamiento entre la red interna y el exterior. Es un ordenador con dos tarjetas de red (en caso de tener más interfaces de red se le denominan **Multi-Homed**) y puede bloquear completamente cualquier tráfico entre la red interna y el exterior. en caso de que los datos de una aplicación necesiten cruzar el cortafuegos, se necesita de un software específico llamado Proxy que se encarga de realizar la transvase de los datos entre ambas aplicaciones, es decir el "almacenamiento y reenvío" de los datos. Como en el caso anterior, para defender este ordenador, se puede colocar entre él y la red Internet un router, (Ver fig. 3) creándose un área llamada DMZ (De militarized Zone). En esta área se puede colocar el servidor Web externo, el servidor FTP, y otros que se estimen oportunos destinados a los usuarios externos.

Otra fórmula de cortafuegos son las **Screened subnets** (ver fig. 4). Consiste en crear un segmento de red aislado, considerado DMZ, que sea la única forma de comunicación entre la red externa y la red interna. A este segmento se conectan, por un lado, la red interna de los usuarios de la empresa a través de un router y por otro, también a través de un router, la red exterior. En este mismo segmento se encuentra el bastión. Los routers deben estar configurados para encaminar y recibir el tráfico hacia y desde el bastión.

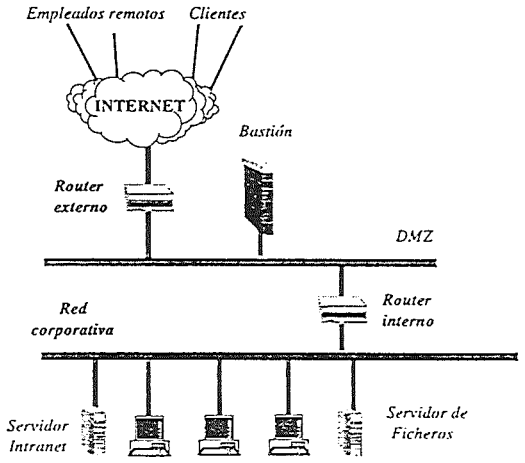


Figura 4. Screened subnets.

Al igual que el resto de sistemas de seguridad, ninguno de los modelos citados es totalmente seguro. Normalmente se deberá buscar un compromiso entre seguridad y facilidad de uso.

Cuanto más rigurosa sea la comprobación de la identidad del usuario realizada por el cortafuegos, más se sentirá éste interrumpido e incomodado. Este factor deberá tenerse en cuenta a la hora de elegir un modelo u otro ya que los usuarios descontentos tienden a buscar formas de saltarse las reglas que les molestan.

5. Redes Privadas Virtuales.

Una Red Privada Virtual (VPN) consiste en la interconexión de un grupo de oficinas de una compañía de manera segura a pesar de que la conexión entre las oficinas se realiza sobre una red insegura como puede ser Internet.

La utilización de Internet para la interconexión de oficinas remotas de una compañía presenta dos ventajas fundamentales:

- Disminución derivados de los costes de comunicación debido al ahorro que supone el menor coste en Internet frente a la conexión mediante líneas dedicadas, sobre todo para el caso de conexiones internacionales.
- Es la mejor forma de completar la intranet de las empresas, aumentando la información accesible de una manera sencilla, integrada y flexible.

Sin embargo, la utilización de tramos de redes públicas expone a la empresa a dos amenazas:

- Acceso no autorizado desde Internet a la red interna.
- Captación de los mensajes de la empresa a su paso por Internet.

La solución es el empleo de los sistemas VPN (Redes Privadas Virtuales) que combinan técnicas de encriptación con técnicas de autentificación para asegurar la red.

La estructura de una VPN se basa en los elementos encriptadores que se encuentran situados en los extremos de las redes privadas de cada oficina. La mayor parte de las VPNs trabajan sobre una red IP estándar y dado que la encriptación de los datos se realiza en la capa IP los paquetes encriptados pueden ser enrutados normalmente.

El elemento encriptador puede ser un router o un firewall (ver fig. 5) que implementan encriptación o una unidad de encriptación dedicada. Este último caso puede ser una implementación hardware o software. La opción dedicada se hace necesaria cuando las redes a interconectar disponen de routers y firewalls de distintos vendedores que no podrán trabajar juntos.

Algunas VPNs encriptan todo el tráfico que se produce entre las distintas oficinas, mientras que otras VPNs son capaces de distinguir entre las distintas aplicaciones y encriptar sólo aquel tráfico sensible que no debe ser capturado. La primera aproximación es más sencilla e inherentemente más segura, sin embargo mediante la segunda no se gasta capacidad de proceso en tareas innecesarias por lo que se obtiene un mayor rendimiento.

Por otra parte muchas VPNs encriptan no solo los datos, sino también las cabeceras de los paquetes IP, con lo que consiguen ocultar las direcciones IP del emisor y receptor de la comunicación. Para poder encriptar la cabecera del paquete IP se introducirán la cabecera y los datos en un nuevo paquete IP mediante una técnica denominada tunneling.

5.1. Intercambio de claves.

La encriptación se realiza mediante un conjunto de claves secretas compartidas por los dos extremos de la comunicación, de manera que mantener secreta la clave es un punto crítico del sistema. El problema viene a la hora de compartir las claves entre ambos extremos de la comunicación.

Algunos sistemas de intercambio de claves se basan en métodos manuales, el administrador de red usa el teléfono o correo certificado para transportar la clave a otro lugar. Éste es un método tedioso y potencialmente peligroso, ya que un hacker podría tener acceso al correo o teléfono de la compañía. La dificultad del intercambio de la clave hace que su actualización sea poco regular lo que facilitaría su descifrado.

Otros sistemas de intercambio de claves se basan en la utilización de claves de sesión distintas para cada sesión y que se intercambian a través de la red durante la inicialización. Estas claves se intercambian mediante criptografía de clave pública. Cada encriptador en la VPN dispone de un par de claves adicional, una pública y una privada. La clave pública se distribuye libremente a todos los lugares de la VPN, mientras que la clave privada se mantiene de forma segura en cada encriptador. Al comienzo de una sesión, la clave de sesión se combina con la clave pública del receptor mediante un algoritmo matemático que sólo permitirá al poseedor de la clave privada correspondiente a esa clave pública obtener la clave de sesión.

5.2 Autenticación.

Los métodos de autenticación también se basan en la criptografía de clave pública. La autenticación se realiza a dos niveles:

- A nivel de paquetes, los elementos encriptadores de cada oficina añaden un sello digital, o checksum a los datos encriptados. Este sello es una secuencia corta de bits que ha sido generada

combinando los contenidos del paquete con la clave privada del emisor. El receptor comprueba que los datos no han sido alterados y mediante la clave pública del emisor comprueba la autenticidad del emisor.

- A nivel del encriptador, los encriptadores nuevos son autentificados manualmente o mediante el uso de certificados. En el modo manual las claves privadas y públicas son cableadas durante el proceso de fabricación. Los productos que emplean certificados digitales generan de manera aleatoria su propio par de claves. A continuación, encripta la clave pública que ha generado con la clave pública de una Autoridad de Certificación (generalmente un servidor seguro dentro de la red de la empresa, o una agencia de certificación ajena a la empresa) para que ésta lo apruebe. Una vez que ha recibido la aprobación de la Autoridad de Certificación (cuando se ha comprobado la autenticidad del emisor), la clave pública se distribuirá a todos los encriptadores de la red. Mediante este método sólo debe transportarse mediante una conexión segura (teléfono, correo,...) la clave pública de la Autoridad Certificadora.

5.3 Usuarios remotos.

La utilización de VPN permite generalmente la inclusión de usuarios remotos en la red de la empresa. Para ello, el usuario remoto dispone de software que puede cargar en su PC o laptop y que le permite dialogar con los encriptadores de la red de la empresa de manera segura empleando los mismos algoritmos que los dispositivos VPN emplean para dialogar entre ellos. De esta manera la comunicación se puede realizar mediante un modem o RDSI para conectarse a Internet siendo la comunicación segura desde usuario remoto.

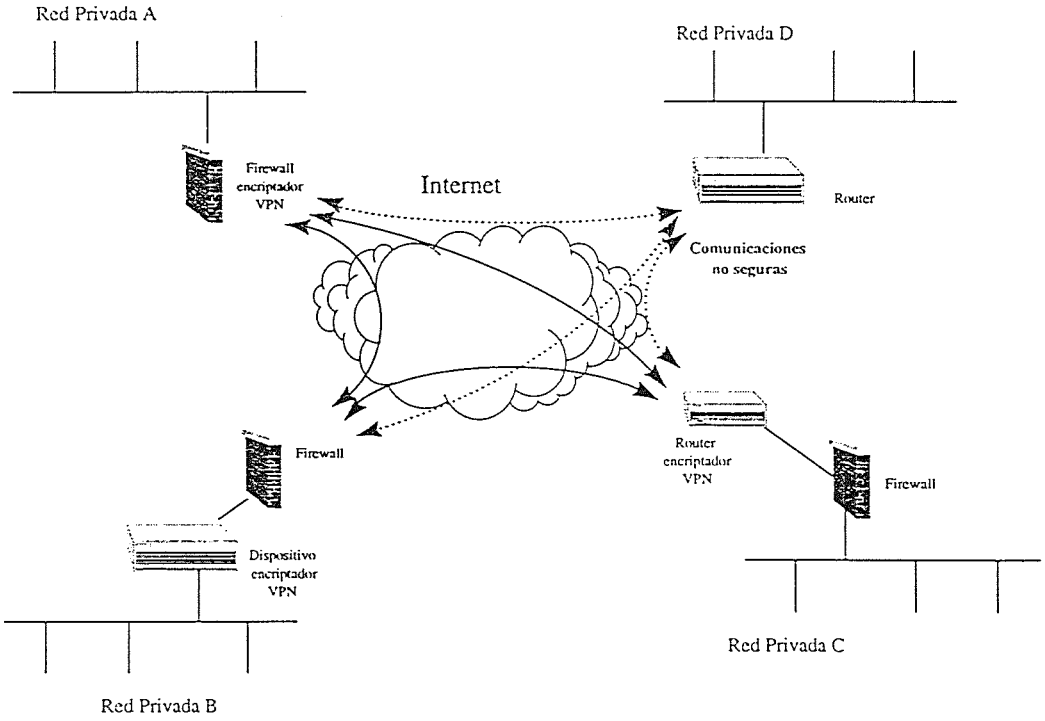


Figura 5. Red Privada Virtual con distintos elementos encriptadores

La única cosa que los usuarios remotos deben hacer es identificarse mediante una clave de acceso o algún sistema de autenticación basado en tokens (tarjetas chip,...).

La presencia de agentes de encriptación en los PCs de los usuarios permiten construir otro tipo de VPN en la que los puntos de control ya no se sitúan en los extremos de las redes privadas, sino en los propios PCs (ver fig. 6). En este caso, se dispone de una autoridad de certificación centralizada que repartirá los certificados a cada uno de los usuarios de todas las oficinas, y las comunicaciones entre todas las máquinas de la red será securizada. Este tipo de red VPN permite un control más granular de los permisos ya que la Autoridad Certificadora puede conceder permisos distintos a los usuarios según grupos que se adaptan dinámicamente.

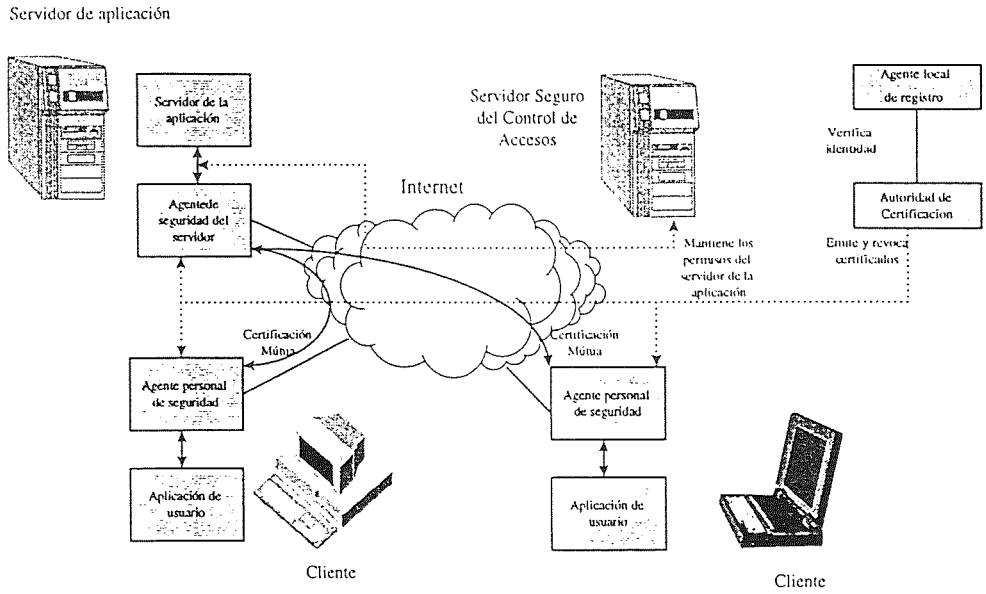


Figura 6. Red Privada Virtual con los puntos de control en los ordenadores personales

Sin embargo, la descentralización de la seguridad añade riesgos ya que un mal uso por parte de un usuario puede poner en peligro todo el sistema.

5.4. El problema de la estandarización.

En la actualidad la tecnología VPN dista mucho de disponer de un estándar. En la actualidad los productos de VPNs soportan un conjunto de estándars establecidos por el grupo de Seguridad IP (IPsec) del IETF que define como los encriptadores VPN deben autentificarse mutuamente y negociar un algoritmo de encriptación y claves. Por ello, la mayoría de los productos que emplean un sistema manual para la gestión de claves deberían trabajar juntos, pero en realidad el IPsec todavía se encuentra en proceso de estandarización RFCs.

Para la gestión de claves el IPsec todavía se encuentra seleccionando entre tres protocolos.

Por ello de momento la única posibilidad es emplear todos los productos del mismo proveedor para asegurarse de que no van a existir problemas de integración.

Conclusiones

Internet es una herramienta imprescindible para mejorar la competitividad de las empresas, sin embargo al conectarse a Internet la empresa se enfrenta a los siguientes objetivos contrapuestos:

- Abrir las redes internas a los clientes, vendedores y empleados remotos o móviles.
- Proteger la información y los recursos internos, de los competidores, espías y hackers.

Para mantener la conexión a Internet bajo control se debe definir una política de seguridad. La política de seguridad debe contemplar el global de la empresa, no sólo la conexión a Internet, y debe involucrar a todos los empleados de la red de la empresa. Una herramienta fundamental a la hora de controlar la conexión a Internet son los cortafuegos (firewall) que actúan como una membrana entre la red interna y la externa, permitiendo el paso a los servicios aceptados en la política de seguridad, pero negando el paso al resto.

Finalmente, para aprovechar toda la potencialidad de la Intranet de la empresa, la Intranet debe interconectar todas las sucursales, creando así una Intranet global, o Red Privada Virtual (VPN). Este tipo de redes obligará a enviar la información encriptada para protegerla de la mirada de los nodos intermedios de la red.

Referencias

[1] Network Working Group. RFC 1244 *Site Security Handbook* (1991).

[2] Karanjit, Siyan y Hare, Chris. *Internet Firewalls and Network Security*. Indianapolis: New Riders Pblishig (1995).

[3] Arnold, S.L. "Security policies for the Internet" *Sesión SE020 de las conferencias de DECUS U.S*, en San Francisco (Diciembre 1995).

[4] Cruz Aguado, Francisco. "Seguridad en Redes y Sistemas", *Boletín de la red nacional de I+D, RedIRIS*, 19-32, 35 (Abril 1996).

[5] Mañas, J.A. "Seguridad en Internet" *Seminario Comercio Electrónico en Internet. Madrid de IIR* (Noviembre 1996).

[6] Cray Andrew "Secure VPNs: Lock the Data, Unlock the savings", *Data commuications*, (Mayo 1997).

[7] CheckPoint Software Technologies Ltd. *Privacy in Public Networks Using Checkpoint FireWall-1*. (Junio 1996).

Seguridad en Internet: Utilización de Marcas de Agua en Imágenes Digitales

Iñaki Goirizelaia, Juanjo Uncilla, Eduardo Jacob, Xabier Andiano
Area de Ingeniería Telemática
Dpto. de Electrónica y Telecomunicaciones
ETSII y de IT Bilbao
Alda. Urquijo s/n - 48013 Bilbao
{jtpgoori, jtpungaj, jtpjatae, jtaanazj} @ bi.ehu.es
Tfno.: (94) 4 27 80 55 - Fax: (94) 4 41 40 41

Abstract:

This paper presents a new method for signing digital images using high frecuency components as guide to insert the digital watermark. It also presents an implementation that allows us verify the autenticity of the watermark by using Java applets.

1. Introducción

La distribución, reproducción y manipulación de imágenes digitales sobre redes de ordenadores se ha convertido recientemente en una operación tan simple como hacer clic en nuestro ratón. Sin embargo, esta utilización plantea serias dudas sobre la legalidad de las operaciones que realizamos con las imágenes que se encuentran a nuestra disposición, obviándose en numerosas ocasiones el posible perjuicio que, debido a una distribución y utilización inadecuada de dichas imágenes, se puede ocasionar a terceras partes.

La aparición de las marcas de agua digitales pretende ofrecer una solución a la problemática planteada. Su utilización consiste en codificar el copyright en la imagen mediante la realización de modificaciones sobre la misma, con el objetivo de proporcionar una clara prueba sobre quién es el propietario de la imagen. La realización de este proceso sobre la imagen deberá ser imperceptible para el usuario, no afectando por tanto a su calidad.

En el presente artículo se presenta un nuevo método que permite realizar la codificación de marcas de agua en imágenes digitales basado en la utilización de los componentes de alta frecuencia, los cuales son utilizados como guía para difuminar la marca en la imagen. Esta técnica presenta a priori mayor resistencia a operaciones de compresión tipo JPEG, aspecto muy importante de cara a la optimización de la transmisión.

2. Antecedentes y estado actual del tema

El uso de marcas de agua como método para proteger la propiedad intelectual es relativamente reciente, pero para ello se utiliza una tecnología muy probada: criptografía, comunicaciones de espectro amplio y teoría de ruido. Los métodos más utilizados para introducir marcas de agua emplean el dominio del espacio, del tiempo o de la frecuencia. Utilizar el dominio de la frecuencia tiene la ventaja de distribuir la marca por todo la imagen (información), con lo que es más resistente a cortes parciales o reducciones; sin embargo un filtro estándar en frecuencia o un algoritmo de compresión, que habitualmente filtran las frecuencias menos significativas, puede dañar la marca de agua. Otra posibilidad es insertar la marca en las bandas de color y luminancia (suelen utilizar ésta, pues tiene mas información en imágenes en color) o en el contorno o la textura de la imagen [1].

Para la realización de este proceso en el dominio del espacio existen varias propuestas, basadas en la modificación del bit menos significativo de los pixels de la zona elegida para insertar la marca [2].

Para la extracción de la marca, en primer lugar hay que seleccionar las localizaciones donde se encuentra (tanto si se marca en el dominio de la frecuencia como en el del espacio). Este proceso suele requerir el original o la marca añadida para realizar la comparación. También es posible extraer la marca sin el original, para lo cual el algoritmo debe detectar propiedades específicas y patrones del documento marcado [3].

Si hay una jerarquía en la inserción de marcas (inserción de varios códigos de identificación y extracción por separado), debe cuidarse el orden de inserción pues la inserción de una marca sobre una imagen ya marcada puede dañar tanto la imagen como la marca (creará ruido adicional que puede degradar la información hasta hacerla irrecuperable). Por otra parte la aparición de varias marcas sobre la misma imagen puede crear problemas a la hora de determinar la autoría de la misma [4].

En el mercado existen algunas herramientas que permiten marcar información, divididas en dos categorías, la primera basada en huellas binarias (FBI, Fingerprinted Binary Information) incluidas en la información (desarrollo realizado por la empresa HighWater FBI), y la segunda que emplea números que oculta en los documentos (desarrollo realizado por NEC y la Universidad Católica de Lovaina). El primer método sobreimpresiona una marca modulando un patrón de ruido del mismo tamaño que la imagen. El segundo método es más rápido, ya que modifica sólo un subconjunto de los datos, aunque es menos robusto y más vulnerable al ataque. Otros desarrollos como el de Digimarc, SysCoP desarrollado por el Instituto Fraunhofer y por Argent de DICE permiten incluir información adicional de identificación del autor (ISBN de un libro, o el nombre del autor). Nuestro trabajo encaja dentro de este último enfoque.

Pese a la existencia de tecnología disponible en el mercado, ésta aun está en su fase inicial, y su mayor reto es ofrecer sistemas en tiempo real que proporcionen protección a ataques sencillos. Está claro que la seguridad total es imposible, pero aun hay que contrastar los métodos actuales con varias operaciones de procesado de señal [1].

Subvencionado por la Unión Europea se ha desarrollado el proyecto TALISMAN [5], cuyo objetivo es definir un modelo funcional común para la inserción de marcas y etiquetas en imágenes, considerando todos los actores (los propietarios del copyright, los distribuidores de la información, los proveedores de servicios de red y los consumidores). En este proyecto se analiza la tecnología existente y mediante la coordinación con los organismos de normalización (ISO, DIVAC, etc.) pretende poner a disposición de los autores mecanismos que les permitan proteger sus derechos de autor.

Es necesario destacar la influencia decisiva de Internet en la importancia creciente de este tipo de desarrollos, por la capacidad de comunicación que supone y su fácil acceso. Por ello cualquier desarrollo debe estar soportado por las nuevas tecnologías disponibles en la red.

3. Definición de la marca de agua

Se puede definir el concepto de marca de agua como una imagen definida por el usuario que, difuminándola en la imagen original a marcar, es utilizada para poder comprobar la propiedad de la misma. Para ello la marca de agua debe cumplir los siguientes requisitos [6]:

- Ser invisible a la percepción humana.
- Ser invisible estadísticamente.
- Rápidamente extraible por el propietario.
- Robusta a distorsiones accidentales o intencionados en la imagen (filtrados, compresión, re-muestreos, recortes, etc.).

El concepto de marca de agua, tal y como se utiliza en nuestro trabajo, se puede definir como una cadena de segmentos lineales de longitud y dirección variable definida por el usuario (ver figura 1).

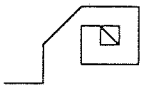


Figura 1: Ejemplo de marca de agua

Esta marca se codifica de acuerdo con el código de cadena (cc), según se muestra en la figura 2.

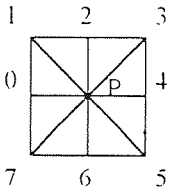


Figura 2: Codificación del código de cadena

La utilización de éste código de cadena juntamente con la longitud de cada segmento nos permite definir la marca de agua de la siguiente forma :

$$cc_1, l_1; cc_2, l_2; cc_3, l_3; .....; cc_n\ l_n$$

donde cc_n indica el código de cadena del segmento y l_n la longitud en pixels.

4. Introducción de la marca de agua en la imagen

La idea básica de nuestro trabajo se basa en la utilización de aquella información de la imagen que puede ser empleada como guía para esconder la marca de agua. El método propuesto para la inserción de la marca de agua en la imagen, se fundamenta en la detección de componentes de alta frecuencia de la imagen, es decir, aquellas zonas donde se presentan transiciones de intensidad importantes. La razón para elegir estas zonas de la imagen es que serán precisamente estas zonas las menos afectadas por los algoritmos de compresión al ser las zonas de máxima información.

La detección de transiciones se realiza mediante la utilización de diversos operadores para realce de flancos. En nuestro trabajo se utiliza el operador gradiente como operador que posibilita la detección de aquellas zonas de la imagen que serán utilizadas como guía para la introducción de la marca de agua [7].

Dada la función I = f (x, y), el operador gradiente se define como un vector cuyo módulo se calcula empleando la siguiente fórmula:

$$(1)\ \Delta_X = \frac{\delta f(x, y)}{\delta x} = \frac{f(x + dx, y) - f(x, y)}{dx}$$

$$(2)\ \Delta_Y = \frac{\delta f(x, y)}{\delta y} = \frac{f(x, y + dy) - f(x, y)}{dy}$$

donde el módulo M y la dirección θ se calculan mediante las expresiones:

$$(3)\quad M = (\Delta_X^2 + \Delta_Y^2)^{1/2}$$

$$(4)\quad \theta = \arctan (\Delta_Y / \Delta_X)$$

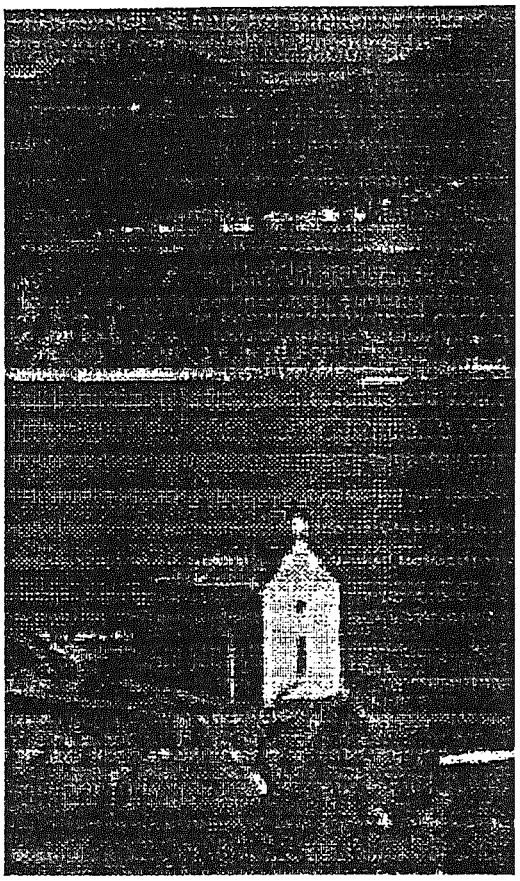
Al ser las imágenes funciones discretas podemos considerar los diferenciales dx y dy, en términos de números de pixels entre dos puntos, pudiendo reescribir las ecuaciones anteriores de la siguiente forma:

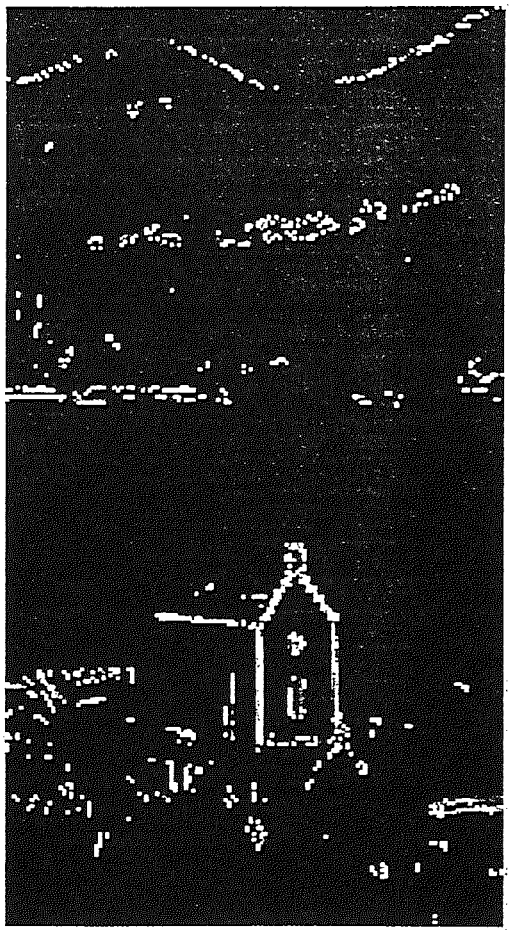
$$(5)\ \Delta_X = \frac{\delta f(x, y)}{\delta x} = \frac{f(x + n, y) - f(x, y)}{n}$$

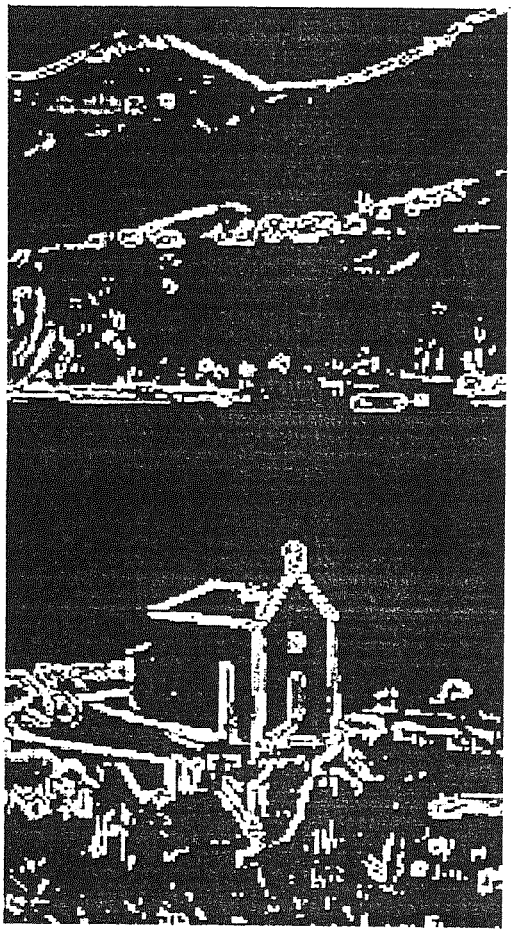
$$(6)\ \Delta_Y = \frac{\delta f(x, y)}{\delta y} = \frac{f(x, y + n) - f(x, y)}{n}$$

En la figuras 3a, 3b y 3c se presentan los resultados de aplicar el operador gradiente a una imagen (fig. 3a), remarcando aquellos puntos cuyo módulo del gradiente supera unos determinados valores.

Una vez definidos los puntos de la imagen donde se producen las transiciones en los valores de intensidad de la imagen, se procede a codificar estos puntos generando una lista segmentos conectados de acuerdo con el código de cadena previamente definido. La información almacenada por cada segmento es el punto inicial, el final y el código de cadena (ver figura 4).







Figuras 3a, 3b, 3c: Imagen original y resultado de aplicar el operador gradiente con distintos umbrales.

Como puede apreciarse los segmentos encadenados se codifican siguiendo pautas similares a las empleadas en la codificación de la marca de agua. Esta codificación consiste en almacenar en una lista encadenada el punto de inicio del segmento con su correspondiente código de cadena. Esta información se almacena para todas las cadenas de segmentos que se encuentran en la imagen.

Como consecuencia de la realización de este preprocesado de la imagen, se obtiene una estructura de datos que contiene todos los conjuntos de segmentos encadenados de la imagen, correspondientes a aquellas zonas donde se producen las transiciones de intensidad de la imagen.

Los criterios de inserción de la marca de agua se basan en las dos estructuras de datos fundamentales de que disponemos: la marca de agua y el conjunto de cadenas de segmentos rectilíneos. El proceso puede resumirse en los siguientes pasos:

1. Se toma el primer segmento de la marca de agua y la primera cadena de segmentos rectilíneos.

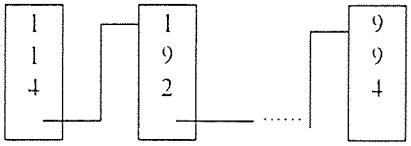


Figura 4: Lista de segmentos lineales

2. Encontramos en dicha cadena de segmentos rectilíneos, el primer segmento cuyo código de cadena coincide con el código de cadena del primer segmento de la marca.

3. Desde el centro del segmento seleccionado, en dirección perpendicular al mismo se selecciona una matriz de 9 puntos tal y como se muestra en la figura 5.

4. Marcamos la matriz de 9 pixels modificando el bit menos significativo.

5. El proceso se repite hasta que se terminen todos los segmentos que hemos definido para la marca de agua.

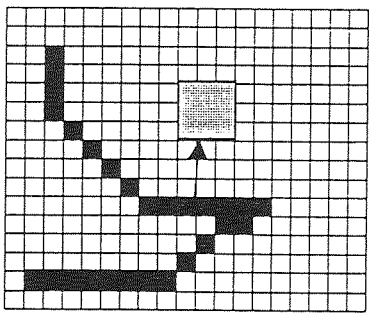


Figura 5: Selección de los pixels a marcar

5. Arquitectura del sistema distribuido de validación

Dada la difusión y facilidad de acceso a la información que supone Internet, el sistema de distribución de las imágenes marcadas debe hacer frente a los retos que supone emplearla. Esta facilidad de acceso a la información es en este caso un inconveniente desde el punto de vista de garantizar los derechos de autor. Por otro lado esa facilidad de acceso hace que los clientes potenciales sean numerosos lo que obliga a pensar en desarrollos que permitan ofrecer, por un lado, un servicio barato, rápido y fácil de utilizar y por otro sistemas de fácil instalación y mantenimiento.

La tecnología web reúne todos estos requisitos. Se pretende ofrecer imágenes marcadas (firmadas) de muy alta calidad (la marca no debe alterar la calidad de la imagen) de forma que se pueda demostrar la autenticidad de las mismas. No se trata de sólo de construir un sistema de distribución de imágenes basado en un servidor web, sino ofrecer a los autores y a los clientes un sistema conjunto de distribución y de validación fiable.

Para integrar el sistema de marcado y validación en un servidor web, debemos considerar cuatro elementos:

- Los **programas de marcado y validación** que son la base de la aplicación.
- La **base de datos** de soporte de la información (imágenes, marcas y datos de autores y clientes).
- **El servidor Web** a través del que se debe ofrecer el servicio.
- **El cliente Web**, que es el único software requerido en el cliente para acceder a los servicios ofrecidos.

El camino clásico para dar solución al problema y relacionar los cuatro elementos ha sido el establecimiento de una arquitectura basada en el CGI (Common Gateway Interface) [8] como se puede ver en la figura 6.

En esta arquitectura a los programas de marcado y validación se les conoce como programas de más allá del servidor, ya que para el cliente estos programas están *ocultos* detrás del servidor. En el servidor Web se crean unas páginas especiales llamadas *forms*. Estas páginas pueden ser accedidas desde el cliente como cualquier otra utilizando su URL ((1) y (2)). Una vez en el cliente, se escribe en ellas la información que necesita el programa CGI (por ejemplo un código de cliente, el tipo de operación a realizar, etc.), y se envían de vuelta hasta el servidor (3).

Cuando esta información llega al servidor Web, utilizando el interfaz CGI éste ejecuta el programa adecuado pasándole la información recogida desde el cliente (4). Este programa realiza sus cálculos y operaciones con la información adquirida, ejecuta las consultas necesarias sobre la base de datos ((5) y (6)), y los resultados obtenidos se vuelven a pasar al servidor Web en forma de página HTML utilizando de nuevo el interfaz CGI (7), y desde aquí, siguiendo el camino habitual, la página creada dinámicamente es enviada al cliente (8).

Emplear esta arquitectura plantea los siguientes inconvenientes:

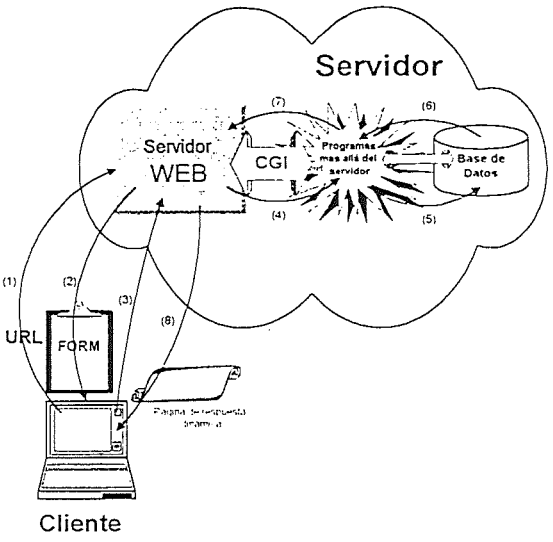


Figura 6: Esquema de funcionamiento usando CGIs

- El tipo de información requerible del usuario a través de los FORM está limitada.

- Cada iteración tipo pregunta-respuesta entre el cliente y el programa de mas allá del servidor, hay que realizarla vía protocolo HTTP con el importante gasto de recursos que conlleva el establecimiento y consiguiente liberación de la comunicación TCP en cada uno de estos pasos.

- Este bajo rendimiento degrada la sensación de interactividad.

Otra alternativa para implementar el sistema es la utilización de applets Java. El esquema de funcionamiento seguirá los siguientes pasos:

- Se pedirá una página especial al servidor que hará de puerta hacia la aplicación.

- El servidor servirá lo demandado.

- En esta página especial vendrá pegado el applet que una vez llegado al navegador del cliente se ejecutará en la misma página visualizada.

Con esto se consigue disminuir la carga del servidor ya que el applet se ejecuta en la máquina cliente. Para crear el camino de comunicación entre la aplicación en el cliente y la base de datos en el servidor es preciso seguir los siguientes pasos:

- Cuando el applet necesite trabajar contra la base de datos del servidor, construirá una vía de comunicación TCP hasta el puerto donde escucha el servidor de base de datos. Para ello se utilizarán las facilidades que ofrece el

lenguaje de programación Java para la gestión de comunicaciones a nivel de transporte utilizando sockets.

- Hecho esto será necesario incluir en los applets un módulo especial encargado de la comunicación con la base de datos hablando su propio protocolo a través del socket TCP. Así las consultas a la base de datos se podrán realizar de una manera directa.

Por lo tanto a la hora de diseñar los applets habrá que concebirlos compuestos por dos módulos como se refleja en la figura 7.

Esta arquitectura presenta ciertas características que la diferencia de las aplicaciones clásicas ofrecidas desde un web:

- Aumento del grado de interactividad debido a la conexión directa a la base de datos y abandono de la conexión vía HTTP.

- Se abandona por completo el modelo basado en CGI que obligaba a enviar datos en claro por la red. Se pueden intercambiar datos de forma segura implementando sobre la conexión directa hasta el servidor (socket) mecanismos de encriptación (clave pública - clave privada). Para ello el applet del cliente no se conectará directamente contra la base de datos sino que lo hará contra una parte de la aplicación que permanecerá en el servidor (3), y es esta aplicación la que trabajará contra la base de datos ((4) y (5)), para devolver los datos requeridos al cliente (6). El esquema de funcionamiento queda recogido en la figura 8.

Las principales ventajas que aporta esta arquitectura frente al esquema clásico vía CGI pueden resumirse en:

- Se libera al servidor de la ejecución de una parte importante de los programas.

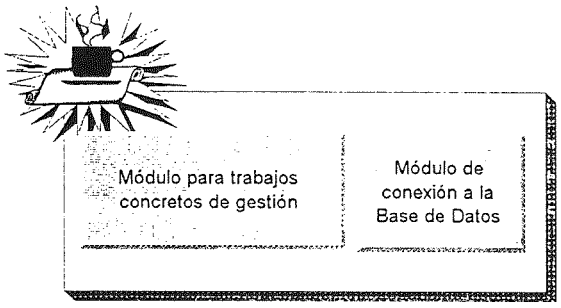


Figura 7: Diseño general de los applets

- Se consigue una interactividad real entre usuario y programa ya que ambos confluyen en la misma localización.

- Para trabajar contra la base de datos, se creará una conexión directa hasta la misma abriendo una conexión TCP hasta el puerto donde escuche el servidor de base de datos. Así:

  - El consumo de recursos es mucho menor. Antes había que crear y destruir varias conexiones TCP. Ahora en cambio, la conexión TCP se realiza una sola vez (directa hasta el servidor de base de datos, sin pasar por el servidor Web) y se mantiene activa hasta que se deje de necesitar relación con la base de datos.

- Permite la creación de un canal de comunicación encriptado entre cliente y servidor, mediante el uso de la tecnología de clave pública y clave privada, incluyendo la posibilidad de firma (para evitar el *ataque del hombre en el medio*). Esta alternativa ofrece más garantías de independencia de la plataforma empleada que el uso de sistemas tipo SSL o SHTTP.

Estas razones nos hacen proponer una arquitectura basada en applets Java para ofrecer el servicio distribuido de validación. Esto nos permite disponer de un sistema doble de seguridad: la propia del canal encriptado más la marca de agua en la imagen.

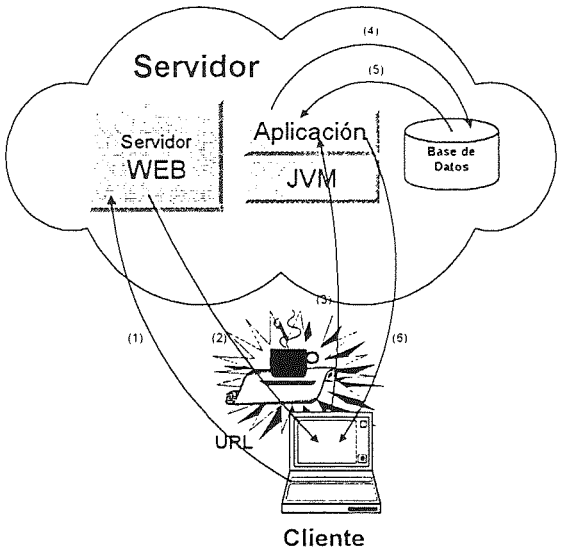


Figura 8: Arquitectura general del sistema

El sistema permite en primer lugar la creación de la marca de agua mediante la unión de dos informaciones, por una parte datos del autor (su dni), y por otra un sello de tiempo certificado [10]. Esta operación es sencilla de realizar puesto que ambos valores son números que mediante una simple operación pueden convertirse en una secuencia interpretable como un lista de códigos de cadena y longitudes de las mismas, como se ha explicado en el apartado 3. Por uniformidad en el desarrollo este software también se ha desarrollado en Java.

Este sistema permite disponer de marcas propias para cada autor y a la vez ofrece un mecanismo de validación en caso de marcas sucesivas o intentos de apropiación indebida de la propiedad intelectual.

Al insertar el autor la imagen en el servidor (para lo cual se ofrece un sencillo interfaz usando también un applet), automáticamente se genera la marca y se marca la imagen. Se ofrece la posibilidad de que el autor modifique el dato a introducir (por defecto es su dni).

Para descargar la imagen marcada, tras autenticar al cliente, se establece una conexión encriptada (mediante la técnica de clave pública y clave privada, K_{pu}, K_{pr}); éste elige la imagen que desea y, se le envía por el canal establecido.

Para verificar que una imagen está marcada y que es original, se descarga un applet que permite enviar la imagen al servidor para proceder a su comprobación. Para ello se emplea un operador de comparación binaria y tras determinar si la imagen se encuentra en la BD se extrae la marca (el método de desmarcado sigue el mismo método que el de marcado), para verificar si tiene la marca que le corresponde.

6. Conclusiones

Ante el aumento constante del intercambio de material multimedia en Internet, son varios los desarrollos que tratan de ofrecer a autores y a clientes potenciales sistemas para el intercambio de información de forma segura. No se pretende sólo garantizar que el acceso a la información es el acordado, sino que posteriormente no se va a hacer un uso ilegítimo de la misma, es decir, comerciar con segundas copias. Una de las propuestas que progresivamente va ganando adeptos es la utilización de marcas de agua digitales para marcar tanto imágenes como audio e incluso vídeo.

El trabajo presentado describe un sistema sencillo para marcar imágenes digitales y su posterior distribución a través de Internet. Dada una marca de agua definida mediante su código de cadena, se establece un método robusto que permite difuminar dicha marca en la imagen sin afectar a la calidad de la misma, siendo posible en todo momento comprobar si dicha imagen tiene la marca o ha sido alterada, así como la existencia, mediante un sello de tiempo certificado, de posibles segundas marcas.

Para su distribución en Internet se ha diseñado una arquitectura que posibilita la comprobación en tiempo real de la existencia de las marcas de agua empleando applets Java, lo que permite su uso en cualquier plataforma. El sistema gestiona una base de datos donde se almacena la información sobre las marcas utilizadas por diferentes autores y así como las fechas de marcado.

Referencias

[1] Zhao, J. "Look, It's not there". *Byte International*, 7-12. (1997).

[2] Steve Walton, S.. Image Authentication for a slippery new age. *Dr. Dobb's Journal*, , 18-26. (1995).

[3] Ó Ruanaidh, J.J.K., Boland, F.M., Sinnen, O. "Watermarking Digital Images for Copyright Protection". *Electronic Imaging and the Visual Arts. Florencia. http://cuiwww.unige.ch/~oruanaid/eva_pap.html.* (1996).

[4] Craver, S., Memon, N., Yeo, B., Yeung, M. "Can Invisible Watermarks Resolve Rightful Ownerships?". *IBM Research Report. 96-07.* (1996).

[5] Delaigle, J.F. "ACTS AC019 Proyect: Tracing Authors' Rights by Labelling Image Services and Monitoring Access (TALISMAN). Parts I, II & III". (1996).

[6] Swanson, M.D., Zhu, B., Tewfik, A. H. "Transparent robust image watermarking". *Proceeding of IEEE International Conference on Image Processing.* Vol III, 211-214 (1996).

[7] Goirizelaia, I., Uncilla, JJ. Igarza, JJ., Pérez, F., Romo, J. y Espinosa, K. "Modelización de contornos mediante la búsqueda de segmentos lineales y circulares en imágenes de niveles de gris". *XII Simposium Nacional de la Unión Científica Internacional de Radio.* (1997).

[8] Rowe, J. "Building Internet database servers with CGI". *New Riders* (1996).

[9] Gosling, J., Joy, B., Steele, G. "The Java Language specification". *Addison-Wesley Publishing Company* (1996)

[10] Ford, W, Baum, M.S. *Secure Electronic Commerce*. New Jersey: Prentice Hall (1997).

Plataforma de negociación de servicios de seguridad en Internet

DAVID REBOLLO MONEDERO, JORDI FORNÉ MUÑOZ y JAVIER JARNE PARDO
DEPARTAMENTO DE MATEMÁTICA APLICADA Y TELEMÁTICA
ESCUELA TÉCNICA SUPERIOR DE INGENIEROS DE TELECOMUNICACION DE BARCELONA
UNIVERSIDAD POLITÉCNICA DE CATALUÑA
C/ JORDI GIRONA 1y 3, MÓDULO C-3, CAMPUS NORD
08034 BARCELONA
Correo electrónico: jforne@mat.upc.es
Tel: (93) 4016011

Abstract:

This paper presents a security system designed to provide secure communications over Internet. Although specific security mechanisms can be developed for each application, we propose an integrated solution able to provide security services for all types of applications running over TCP/IP.

In this system sensitive applications can specify a considerable number of parameters in order to define the key management as well as the following secure communication. This procedure is carried out by using a set of primitive functions.

The security system computes combinations of ¡algorithms and mechanism and decides the best way to fulfil the specifications, minimizing the computational cost due to the cryptographic operations.

1. Introducción

1.1 Motivación

A medida que avanzamos hacia lo que algunos denominan *la sociedad de la información*, el papel de las redes teleinformáticas en el mundo de las comunicaciones adquiere una importancia cada vez mayor. Fruto del vertiginoso desarrollo tecnológico y de las nuevas necesidades sociales nació y crece a pasos agigantados lo que puede considerarse no sólo como la red de redes sino como todo un fenómeno : Internet.

A lo largo de la historia de Internet, tratada ampliamente en [1], Internet ha experimentado un crecimiento imparable, desde su origen en 1969 como ARPANET, creada por la Agencia de Proyectos de Investigación Avanzados del Pentágono con el objetivo de permitir que informáticos e ingenieros trabajando en proyectos militares por toda Norteamérica pudiesen compartir valiosos y potentes ordenadores, hasta nuestros días, en los que se ha convertido en la mayor red telemática del planeta –véase la Tabla 1.

Esta red permite que decenas de millones de usuarios de todo el mundo se conozcan, trabajen en grupo, compartan ideas, obtengan información multimedia interrelacionada sobre todos los temas imaginables, negocien, intercambien correo electrónico, transfieran ficheros y un larguísimo etcétera.

Desafortunadamente, este desarrollo ha venido acompañado de una creciente preocupación causada por los riesgos inherentes a toda gran red en la que un gran número de personas comparten multitud de recursos. Dichos riesgos se manifiestan en todos los servicios y las formas de comunicación que proporciona Internet: el correo electrónico, la transferencia de ficheros FTP (File Transfer Protocol), los servicios de conexión telnet o

Tabla 1. Incremento de hosts, de dominios y de redes en Internet desde enero de 1995 a enero de 1996.

Fecha	Hosts	Dominios	Redes		
			Clase A	Clase B	Clase C
Enero 1995	4 862 000	71 000	91	4 979	34 340
Julio 1995	6 642 000	120 000	91	5 390	56 057
Enero 1996	9 472 000	240 000	92	5 655	87 924

X-Windows, los servicios de información World Wide Web, los grupos de noticias, el servicio de directorio finger, las conversaciones electrónicas conocidas como chat, etc. Por citar algunos de estos problemas, considérese que el correo electrónico puede ser interceptado y leído por una persona distinta al destinatario, y lo mismo para un número personal requerido en una transacción monetaria, que la información a la que se accede puede ser falsa o conducir a engaños, o que alguien se puede hacer pasar por otra persona en una sesión de chat.

Es fácil percatarse de que la seguridad no es un tema sencillo. Las medidas para contrarrestar todas estas amenazas son tan numerosas y diversas como lo son los propios ataques contra la seguridad. Una visión general de los peligros en las autopistas de la información se proporciona en [2], donde se hace patente que en algunos casos las precauciones estriban en el comportamiento de los propios usuarios por encima de las tomadas en la implementación del sistema. Otros peligros y soluciones se presentan en [3] y [4], centrándose en sistemas UNIX.

En respuesta a estos problemas, cabe resaltar el gran esfuerzo que se está realizando a nivel mundial para incorporar en las redes de ordenadores mecanismos de seguridad que controlen el acceso a recursos, garanticen integridad y confidencialidad de las comunicaciones, etc. Pueden mencionarse a título ilustrativo, entre otras, las instituciones siguientes: el ANSI (American National Standards Institute), la ISO (International Organization for Standardization) la ITU (International Telecommunication Union) y el NIST (National Institute of Standards and Technology). Y entre los proyectos, los sistemas PGP (Pretty Good Privacy) y PEM (Privacy Enhanced Mail) para la seguridad en correo electrónico, el protocolo SSL (Secure Socket Layer) utilizado por el Netscape Navigator, software de navegación en la Web, y el paquete de funciones C conocido como SecuDE (Security Development Environment) para la implementación de sistemas de comunicación seguros. Todo ello pone de manifiesto el enorme interés que suscita el tema.

Vista la complejidad del escenario presentado, es comprensible contemplar sólo una porción de los problemas que se plantean. El presente estudio se circunscribe pues a un escenario parcial o restringido, en la línea del análisis de las amenazas contra la seguridad del estándar de arquitectura de seguridad ISO/IEC 7498-2, que conduce a la definición de un conjunto reducido de servicios de seguridad para contrarrestar tales amenazas, tratados en el siguiente apartado. Dentro de este marco, se considera en particular la comunicación basada en la familia de protocolos TCP/IP, que es la que se utiliza en Internet. Con mayor concreción, el estudio se centra en el protocolo de transporte orientado a conexión TCP – el más frecuente– empleado en un entorno UNIX.

1.2 Fundamentos de seguridad

1.2.1 Amenazas contra la seguridad

Se define como **amenaza o ataque contra la seguridad** cualquier acción que compromete la seguridad de la información perteneciente a una persona u organización.

Los tipos de amenazas contra la seguridad pueden caracterizarse entendiendo la comunicación entre dos ordenadores como un flujo de datos de una fuente a un destino, como se muestra en la Fig. 1. En esa misma figura aparecen también representados las cuatro clases en las que pueden clasificarse las amenazas contra la seguridad :

- **Interrupción.** Una parte del sistema resulta destruida o no disponible en un momento dado. La destrucción de una parte del hardware o el corte de una línea de comunicación son ejemplos de interrupción.

- **Intercepción.** Una entidad no autorizada accede a una parte o a la totalidad de la información. Ejemplos de este ataque son la copia ilícita de ficheros o programas transmitidos a través de redes de datos utilizando analizadores de redes o la lectura de correo electrónico por una tercera persona en cualquiera de los ordenadores por los que pasa el mensaje antes de alcanzar el destino

- **Modificación.** Una entidad no autorizada no sólo accede a una parte de la información, sino que además es capaz de modificar su contenido.

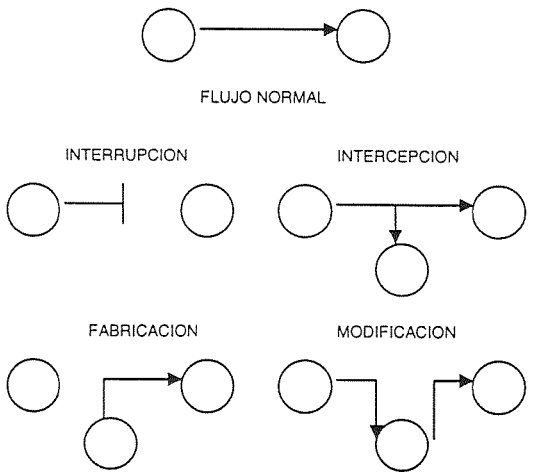


Figura 1. Amenazas contra la seguridad

Ejemplos de tales modificaciones son la alteración de ficheros de datos, alteración de programas y modificación de mensajes mientras son transmitidos por la red.

- **Fabricación.** Una entidad no autorizada envía mensajes haciéndose pasar por un usuario legítimo. Entre los ejemplos se incluye el envío de correo electrónico suplantando la identidad de otro usuario.

Para una descripción más detallada, consultar [5].

1.2.2 Mecanismos de seguridad

Un **mecanismo o técnica de seguridad** es un procedimiento diseñado para detectar, prevenir o recuperarse de un ataque contra la seguridad.

Existe una gran variedad de mecanismos de seguridad diseñados para contrarrestar las amenazas enumeradas, si bien ninguno de ellos por sí solo es capaz de hacer frente a todas las clases de ataque. El presente estudio hace uso de un cierto tipo de técnicas de seguridad : los algoritmos y mecanismos criptográficos.

A continuación se expone de forma concisa y esquemática una clasificación de los algoritmos y mecanismos criptográficos en base a su comportamiento. Aunque incompleta, es suficiente como primera aproximación a la comprensión del funcionamiento de los algoritmos criptográficos. Una completa y rigurosa explicación y valoración sobre un gran número de algoritmos criptográficos aparece en [6]. La clasificación en cuestión es la siguiente :

- **Algoritmos de cifrado.** Transforman la información de modo que ésta no sea inteligible para una entidad no autorizada. Hay dos grandes grupos de algoritmos de cifrado :

  - **Cifrado simétrico o de clave secreta.** En la comunicación de un mensaje, la entidad emisora y la receptora poseen ambas una misma clave secreta desconocida por el resto de entidades. Dicha clave determina la transformación del mensaje y hace posible la correcta transformación inversa para recuperar el mensaje original.

  - **Cifrado asimétrico o de clave pública.** A diferencia del cifrado simétrico, no se tiene una sola clave compartida sino una pareja de ellas. Una de las claves, llamada pública, es conocida por todas las entidades, mientras que la otra, llamada privada, es conocida por una única entidad. En la comunicación de un mensaje, la entidad emisora aplica una transformación al mensaje haciendo uso de la clave pública de la receptora, que es conocida. Para poder leer el mensaje deshaciendo la transformación, la entidad receptora emplea su clave privada, que no es conocida por ninguna otra entidad.

- **Funciones hash.** A partir de un mensaje, generan un bloque de longitud fija y reducida que es función de dicho mensaje, de modo que una variación mínima del mensaje original produciría un resultado en la función hash distinto. Entre otras propiedades, una función hash debe cumplir que no sea computacionalmente factible obtener dos mensajes con un mismo resultado de la función hash.

- **Algoritmos de checksum criptográfico.** Similares a las funciones hash, excepto por el hecho de que funcionan con una clave que determina la transformación.

- **Algoritmos de firma digital.** A semejanza de los algoritmos de checksum criptográfico, generan un bloque de longitud fija a partir de un mensaje y de una clave, y cualquier variación en el mensaje produce una firma digital distinta. Como en el caso del cifrado asimétrico, interviene una segunda clave. La utilizada para realizar la firma es la clave privada. El proceso de verificación de la validez de la firma requiere solamente el uso de la clave pública. Conocida la entidad a la que pertenece la clave pública, cualquier otra puede verificar que el mensaje fue firmado por esa entidad, pues es la única que tiene acceso a la clave privada.

- **Generación de números aleatorios y pseudoaletorios.** A diferencia de los generadores de números aleatorios y pseudoaleatorios utilizados en otras áreas, los utilizados en criptografía deben cumplir una propiedad adicional. Una entidad que no sea la que genera los números pero que conoce una serie finita de ellos no debe ser capaz de predecir el siguiente número que va a ser generado.

- **Mecanismos o protocolos de gestión de claves simétricas.** Intercambio de mensajes que permite que dos entidades convengan en el uso de una clave para utilizar en un cifrado simétrico o en un checksum crptográfico de forma que ninguna otra entidad conozca cuál es esa clave, proporcionando además a cada una de las entidades participantes la seguridad de que la otra entidad es quien dice ser y que no suplanta la identidad de otra.

Con frecuencia ocurre que los algoritmos de un tipo se sirven o contienen algoritmos de otro tipo. Por ejemplo, un gran número de mecanismos de gestión de claves se construyen a partir de algoritmos de firma digital y para generar las claves simétricas emplean generadores de números aleatorios.

1.2.3 Servicios de seguridad

Se entiende por **servicio de seguridad** aquél que mejora la seguridad de los sistemas de procesado y de comunicación de la información de un grupo determinado de personas o de una organización. Un servicio de seguridad hace frente a ataques contra la seguridad empleando uno o varios mecanismos.

En el estándar de arquitectura de seguridad de la ISO se definen cinco servicios :

- **Autentificación.** Proporciona la certeza de la identidad de una entidad. Contrarresta al ataque de fabricación. Debe satisfacerse en un algoritmo de gestión de claves, y emplea entre otros los algoritmos de firma digital y cifrado asimétrico.

- **Confidencialidad.** Protege contra el acceso no autorizado a parte o a la totalidad de una información. Contrarresta el ataque de intercepción. Se basa en alguno de los tipos de cifrado.

- **Control de acceso.** Protege contra el uso o la manipulación no autorizada de recursos. Un ejemplo conocido de servicio de control de acceso y autentificación, llamado Kerberos, hace uso de algoritmos de cifrado simétrico y de protocolos de gestión de unas unidades de información para la autentificación denominadas tickets.

- **Integridad de los datos.** Protege contra la modificación, el borrado o la substitución de información. Contrarresta el ataque de modificación. Emplea algoritmos de checksum criptográfico, funciones hash y firma digital, y debe satisfacerse en un mecanismo de gestión de claves.

- **No repudio.** Protege contra la negación de una entidad que participa en una comunicación de haber enviado un mensaje –repudio de origen– o de haberlo recibido –repudio de entrega. Emplea básicamente algoritmos de firma digital y funciones hash.

La Tabla 2 establece una serie de analogías entre los servicios de seguridad y mecanismos de seguridad relacionados con la vida cotidiana.

Tabla 2. Analogías entre los servicios de seguridad y la vida cotidiana

Servicio de seguridad	Ejemplo de la vida cotidiana
Autentificación	Carné con identificación fotográfica Huellas dactilares
Control de acceso	Llaves y cerrojos
Confidencialidad	Tinta invisible Carta lacrada
Integridad	Tinta indeleble
No repudio	Firma notarizada Correo certificado

En [5], [6] y [7] aparece un excelente y exhaustivo análisis de las técnicas y protocolos empleados en los diferentes servicios de seguridad.

2. Escenario a proteger

El escenario o entorno del sistema de seguridad que se propone consiste en una serie de máquinas UNIX interconectadas a través de redes de comunicación, empleando la arquitectura de protocolos TCP/IP. En la Fig. 2 se muestra un ejemplo esquemático de este escenario, en el que aparecen una LAN e Internet como redes que interconectan los ordenadores.

En ese escenario, dos procesos, cada uno ejecutándose en un ordenador distinto, se comunican por la red estableciendo la típica relación cliente-servidor. Dicha comunicación se lleva a cabo mediante el protocolo orientado a conexión TCP. Se supone que en la programación de las dos

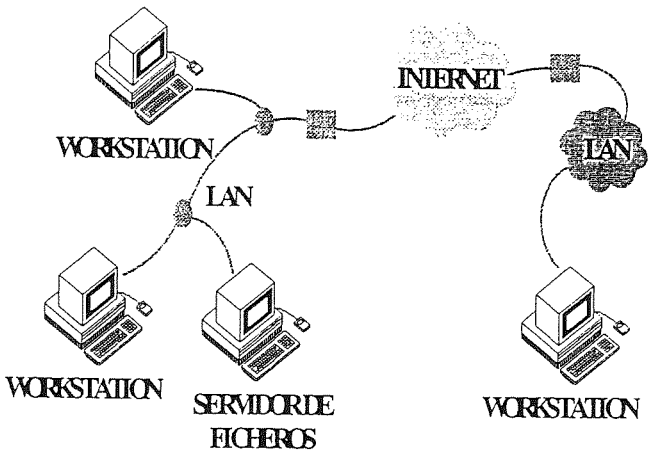


Figura 2. Ejemplo de escenario habitual

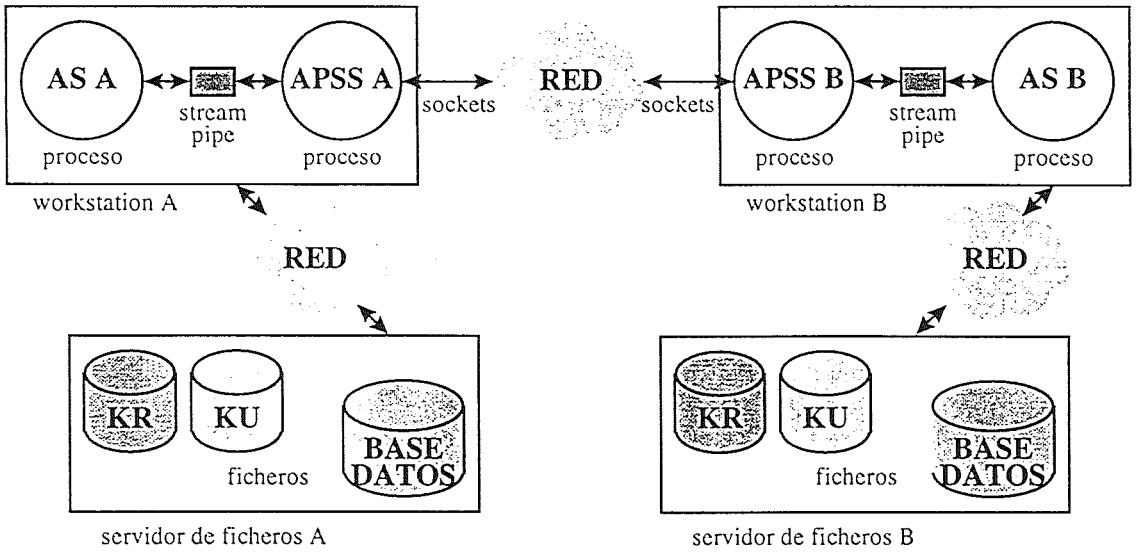


Figura 3. Arquitectura del sistema de seguridad.

aplicaciones se ha utilizado la API[1] de sockets de Berkeley.

El presente estudio considera como hipótesis de partida que la seguridad en lo que concierne al interior del ordenador y al propio entorno UNIX –passwords, permisos de acceso a ficheros, etc.– está garantizada, y en cualquier caso fuera del alcance del análisis que se realiza del problema. Sí se contemplan en cambio posibles ataques no sólo en la red a través de la cual se lleva a cabo la comunicación de dos máquinas UNIX, sino también en aquélla que conecta las máquinas con su servidor de ficheros, de acuerdo con el servicio Network File System (NFS).

3. Arquitectura del sistema

El principio básico del sistema de seguridad presentado es proporcionar un nuevo método de comunicación entre procesos (IPC) por red en un entorno UNIX que soporte los servicios de seguridad siguientes :

- Autentificación de entidad y de los datos
- Integridad
- Confidencialidad de los datos
- No repudio de origen y destino

La asociación entre las aplicaciones así establecida puede calificarse de segura.

El esquema de la arquitectura del sistema de seguridad propuesto se representa en la Fig. 3. Los elementos que intervienen son :

- **Máquinas o workstations UNIX.** Aquéllas en las que se ejecutan las aplicaciones que se comunican a través de la red. También, las que actúan como servidores de ficheros.
- **Redes.** Por un lado, la red empleada por las aplicaciones comunicantes para establecer una asociación. Por otro, las redes que interconectan los servidores de ficheros con sus clientes.
- **Aplicaciones sensibles (AS).** Aplicaciones comunicantes que solicitan al sistema de seguridad ciertos servicios. Se trata de los procesos cliente y servidor que se comunican a través de la red. Se utilizan las denominaciones **AS cliente** y **AS servidora** para la distinción de dichas aplicaciones, que se identifican alternativamente –y especialmente en las figuras– como AS A y AS B, aludiendo a la AS cliente y a la servidora respectivamente.

[1] API es el acrónimo de Application Programming Interface. Una API consiste en un conjunto de funciones disponibles para un programador de forma que éste pueda utilizar un servicio determinado. En este caso se trata de la API de sockets de Berkeley, que es una de las dos principales en sistemas UNIX para comunicación de procesos a través de una red. Una descripción y ejemplos del uso de sockets en UNIX se ofrece en [8].

Se entiende por **entidad** el usuario cuyo identificador real[2] está asociado al proceso correspondiente a cada una de las AS. Las denominaciones **entidad cliente** y **entidad servidora** permiten la distinción de las dos entidades implicadas en la comunicación. El identificador de entidad es simplemente el identificador de usuario.

- **Aplicación proveedora de servicios de seguridad (APSS).** Se ejecuta simultáneamente en las dos máquinas comunicantes. Los dos procesos resultantes se encargan de llevar a cabo los mecanismos y algoritmos criptográficos que soportan los servicios solicitados por la aplicación sensible correspondiente. Por analogía a la nomenclatura para distinguir las AS comunicantes, también son designadas las APSS como **APSS cliente** y **APSS servidora**, que alternativamente –y sobre todo en las figuras– se identifican también como APSS A y APSS B respectivamente.

- **Stream pipe.** Pipe bidireccional entre el proceso de la AS y el proceso de la APSS correspondiente.

- **Sockets.** API de E/S por red utilizada.

- **Ficheros base de datos.** Contienen información sobre los algoritmos y mecanismos criptográficos utilizados por las APSS.

- **Ficheros de parejas de claves privadas y públicas de las entidades.** (En la Fig. 3 aparecen con la abreviatura KR). Contienen la lista de las parejas de claves privadas y públicas del usuario que ejecuta la AS. Las claves privadas deben encontrarse cifradas si el ordenador en el que se ejecuta la aplicación y el servidor de ficheros no son el mismo ordenador.

- **Ficheros de claves públicas.** (En la Fig. 3 aparecen con la abreviatura KU). Contienen una lista de certificados de varios usuarios con los que el usuario que ejecuta la AS puede desear comunicarse.

La API que presenta el sistema al programador queda constituida por una ligera

[2] Cada proceso en UNIX tiene varios identificadores numéricos asociados. Entre ellos constan el identificador real de usuario y el identificador efectivo de usuario. El primero relaciona el proceso con el usuario que lo ejecuta, y es el que aparece en el fichero de passwords para la entrada en el sistema –log in. Normalmente ese valor no varía, contrariamente a lo que ocurre con el segundo identificador, utilizado para comprobaciones de acceso a ficheros

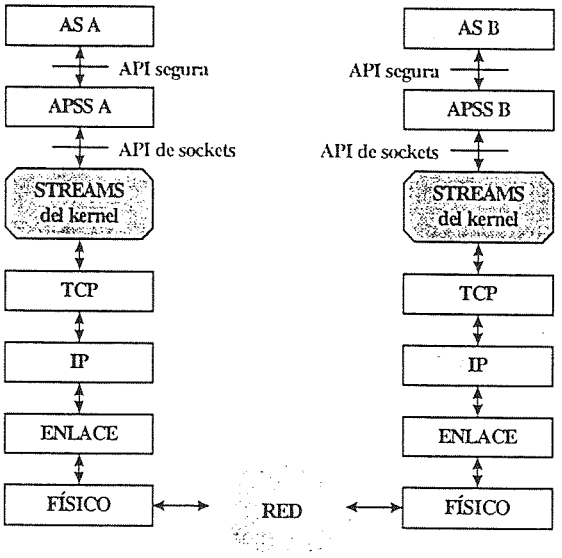


Figura 4. Sistema de seguridad desde el punto de vista de la arquitectura de protocolos.

modificación de la interfaz de sockets de Berkeley y por unas **primitivas del servicio** que permiten el control de la asociación segura y en particular la especificación del servicio solicitado. Un análisis desde el punto de vista de arquitectura de protocolos conduce a la estructura en capas ilustrada en la Figura 4.

Recuérdese que se define como STREAMS el conjunto de mecanismos del kernel que soportan el desarrollo de servicios de red y drivers de comunicación de datos. Estos mecanismos definen una interfaz estándar para la E/S de caracteres dentro del kernel y entre el kernel y el proceso a nivel de usuario.

4. Rasgos generales de funcionamiento

El servicio proporcionado por el sistema de seguridad a la AS es orientado a conexión. Así pues, se tienen los pasos : establecimiento, mantenimiento y liberación de la asociación segura.

4.1 Establecimiento de la asociación segura

La facilidad del establecimiento de la asociación segura constituye un servicio confirmado en el que se distinguen dos fases, la segunda de ellas opcional : negociación de parámetros del servicio de seguridad y gestión de claves de sesión.

4.1.1 Negociación de los parámetros del servicio

La AS cliente, o AS A en la Fig. 3, inicia el establecimiento de la asociación segura mediante una primitiva de petición enviada a la APSS cliente o APSS A, en la que especifica el servicio que

solicita. Esta petición se transmite a la APSS servidora o APSS B y es indicada a la AS servidora o AS B. Si esta última acepta el establecimiento de la asociación en las condiciones especificadas por la AS cliente, responde a la APSS servidora afirmativamente, y ésta comunica a la APSS cliente el resultado.

En este intercambio de mensajes, las APSS cliente y servidora obtienen la información necesaria para iniciar un **algoritmo de decisión conjunto** que tratará de encontrar una combinación adecuada de algoritmos y mecanismos criptográficos para atender a la solicitud de la AS cliente de establecer una comunicación segura. Dicho algoritmo de decisión utiliza además la información contenida en los ficheros enumerados anteriormente : fichero de parejas de claves privadas y públicas del propio usuario identificado por la AS cliente, fichero de certificados de clave pública de otros usuarios y fichero base de datos.

La APSS cliente, mediante el algoritmo de decisión, busca conjuntos de mecanismos que satisfagan los requerimientos de su AS local y envía cierta información sobre ellos a la APSS remota. Esta última, a su vez, realiza las comprobaciones necesarias sobre los mecanismos propuestos por la APSS cliente informándole de los resultados. El algoritmo finaliza con éxito cuando un mecanismo es aceptado por ambas partes. Para acelerar el proceso, especialmente en aquellos casos en los que se repiten con frecuencia solicitudes de servicio idénticas o muy similares, la APSS cliente dispone de una **caché** en la cual se almacenan los últimos mecanismos utilizados.

El mecanismo aceptado puede consistir en realidad en dos mecanismos :

- **Mecanismo de comunicación segura (MCS).** Mecanismo criptográfico aplicado en la fase de comunicación o intercambio de mensajes con información de las AS.

- **Mecanismo (o protocolo) de gestión de la clave de sesión (MGCS).** Mecanismo criptográfico que se utiliza, si es necesario, para obtener la clave de sesión del MCS.

4.2 Gestión de la clave de sesión

La gestión de la clave de sesión viene determinada por la fase anterior y es necesaria sólo en el caso de que el MCS requiera el uso de una clave simétrica. Sea como sea, siempre se realiza una sencilla comprobación de autentificación e integridad sobre cierta información de los mecanismos decididos que, si se lleva a cabo el MGCS, se incorpora a él. Esta comprobación tiene por objetivo contrarrestar posibles ataques contra la seguridad durante la fase de negociación del servicio.

Cada APSS informa a la AS correspondiente de la finalización del establecimiento de la asociación segura. Las AS son igualmente informadas si durante la gestión de claves se detecta una violación de la seguridad.

4.3 Mantenimiento de la asociación segura

En este estudio se entiende por mensaje una secuencia finita de datos enviados desde una AS –emisora– a otra –receptora– que desde el punto de vista de programación se traduce en una llamada al sistema *write* en emisión y otra *read* en recepción. En la fase de mantenimiento de la asociación se realiza dicho intercambio de mensajes, que se denomina comunicación segura. Las AS intercambian los mensajes que las APSS modifican mediante el MCS para garantizar la seguridad de la comunicación. La APSS correspondiente a la AS receptora realiza la transformación inversa a la realizada por la APSS correspondiente a la emisora,de modo que el mensaje viaja seguro por la red. Se llama **mensaje original** el generado por la AS emisora y recibido por la receptora, y **mensaje protegido** o **transformado** el que viaja por la red, resultado de la aplicación del MCS. La Fig. 4 ilustra este comportamiento.

Los mecanismos criptográficos empleados incluyen habitualmente comprobaciones para detectar violaciones de la seguridad. En caso de detección, las AS son informadas.

4.4 Liberación de la asociación segura

La facilidad de liberación de la comunicación segura se presenta tanto en forma de servicio confirmado como de no confirmado.

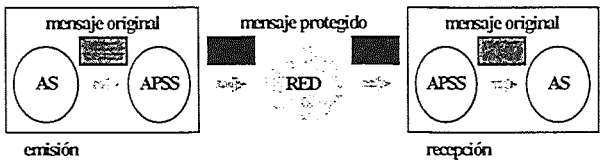


Figura 5. Comunicación segura.

5. Gestión de claves

Cada una de las AS comunicantes posee un conjunto de claves asimétricas para ser utilizadas en un protocolo de gestión de claves de sesión simétricas. La generación de estas claves asimétricas y de sus certificados es una tarea costosa que desarrolla una autoridad de certificación. Por otro lado, la obtención de estas claves asimétricas una vez generadas no es una cuestión trivial.

La solución propuesta para la gestión de las claves implicadas en los mecanismos de seguridad conduce a un escenario con una triple jerarquía de claves :

- **Claves** asimétricas **para la gestión de claves maestras** asimétricas
- **Claves maestras** asimétricas para la gestión de claves de sesión simétricas
- **Claves de sesión** simétricas

En una doble jerarquía de claves, las claves maestras son utilizadas en los protocolos de gestión de claves de sesión. Sin embargo, las propias claves maestras también requieren un procedimiento de gestión. Pues bien, en la triple jerarquía, la misma relación que guardan las claves de sesión y las maestras se establece entre las maestras y las claves del primer nivel jerárquico. De este modo, la entidad que desee obtener una nueva pareja de claves maestras lo solicitará a una autoridad de certificación. La entrega autentificada y confidencial de las claves privadas se logra mediante el uso de un protocolo con las claves de gestión de claves maestras, y la entrega de las claves públicas, en forma de certificado. La Fig. 6 esquematiza este comportamiento.

En cuanto al procedimiento de gestión de las claves de primer nivel, éste podría consistir por ejemplo en la entrega en persona de las claves, ya que se trata de claves cuyo uso se restringe a la obtención de nuevas parejas de claves simétricas, y por lo tanto el tiempo de expiración será previsiblemente largo. El método seguido para entregar las claves maestras podría basarse en el algoritmo RSA, por ejemplo, pero en todo caso estaría totalmente definido. Otra posibilidad es que cada entidad genere su propia pareja de claves privada y pública. De esta forma la autoridad de certificación no conoce la clave privada y por lo tanto se requiere una menor confianza en ella.

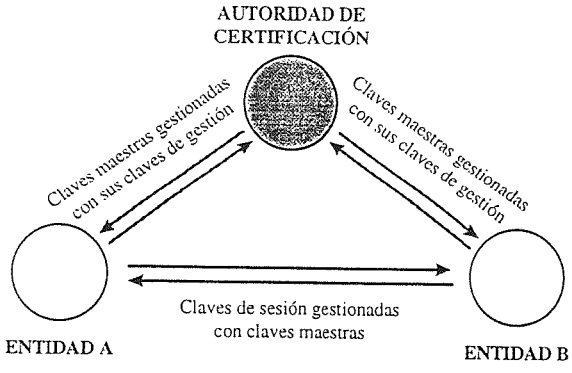


Figura 6. Gestión de claves de triple jerarquía

Una de las ventajas de este método es explotada por el sistema de seguridad, que permite trabajar con diferentes algoritmos de cifrado simétrico y firma digital en los MGCS y los MCS, y consecuentemente con diferentes tipos y longitudes de claves maestras. Si se tuviese una jerarquía doble y las obtención de claves maestras requiriese su entrega en persona, mucho más costosa, las entidades no dispondrían tan fácilmente de múltiples claves, por lo que esta flexibilidad no existiría.

En tanto en cuanto la frecuencia en el uso de las claves contribuye negativamente a su seguridad, y en caso de compromiso la obtención de nuevas claves supone un coste en absoluto despreciable, otro de los atractivos de este método radica precisamente en que las claves maestras, en caso de compromiso, se renuevan con relativa facilidad, y las claves para la gestión de claves maestras se utilizan con escasa frecuencia. Igualmente, la facilidad de obtención de claves maestras podría permitir reducir su periodo de expiración.

6. Implementación software

El sistema de seguridad se ha implementado utilizando el lenguaje de programación C++ para estaciones de trabajo Sun con sistema operativo Solaris SunOS 5.5.1 basado en UNIX System V Release 4.0. Entre las librerías de funciones se incluye la librería 3B para la compatibilidad con la versión de UNIX BSD. La implementación realizada también es compilable bajo el sistema operativo Linux 1.2.13.

7. Conclusiones

Se ha presentado un sistema de seguridad que permite proteger la información mientras es transmitida por Internet.

El sistema propuesto permite que aplicaciones sensibles transformen una conexión insegura en una conexión segura mediante la negociación de una serie de parámetros de seguridad. Una vez establecida la asociación de seguridad, el sistema es transparente para las aplicaciones, lo que lo convierte en una opción atractiva frente al desarrollo de sistemas de seguridad específicos para cada tipo de aplicación.

Una descripción más completa de este proyecto se encuentra en [9].

Referencias

[1] Comer, D. *The Internet Book*. Englewood Cliffs, New Jersey : Prentice-Hall (1995).

[2] Barret, D.J. *Bandits on the Information Superhighway*. Sebastopol, CA : O'Reilly & Associates (1996).

[3] Castro, R.M. y otros "Seguridad en sistemas UNIX". *Revista NOVÁTICA*, Julio/Agosto 1995.

[4] Medina, M. y Fernández, A. "Seguridad en Internet". *Revista NOVÁTICA*, Julio/Agosto 1995.

[5] Stallings, W. *Network and Internetwork Security : Principles and Practice*. Englewood Cliffs, New Jersey : Prentice-Hall (1995).

[6] Schneier, B. *Applied Cryptography : Protocols, Algorithms and Source Code in C*, Segunda edición. New York : John Wiley & Sons (1996).

[7] Ford, W. *Computer Communications Security : Principles, Standard Protocols and Techniques*. Englewood Cliffs, New Jersey : Prentice-Hall (1994).

[8] Stevens, W.R. *UNIX Network Programming*. Englewood Cliffs, New Jersey : Prentice-Hall (1990).

[9] D. Rebollo, J. Forné, "Seguridad en Internet. Desarrollo de una plataforma de negociación de servicios de seguridad para sistemas UNIX". Proyecto Final de Carrera, ETSETB, UPC. 1997.

Modelo de Seguridad Interno: Una alternativa para realizar Estudios de Seguridad de Pequeños Sistemas

Eduardo Jacob, Juan José Uncilla
Área de Ingeniería Telemática, Departamento de Electrónica y Telecomunicaciones
ETSII e IT de Bilbao, Euskal Herriko Unibertsitatea - Universidad del País Vasco
Alameda de Urquijo S/N, 48013 Bilbao
{jtpjatae, jtpungaj}@bi.ehu.es

Abstract:

The present article intends to explain the approach taken by the authors to study the security in Information Technology infrastructures found in small and medium sized enterprises. We propose a Security Model and a methodology that can be used both to validate an applied security policy or to find the best security policy that can be applied to a given IT system.

1. Introducción

La preocupación por la seguridad en el procesado de información y en las redes de datos está asumida ya en la actualidad por la gran mayoría de los usuarios de los mismos.

Aunque tradicionalmente el término *"seguridad"* ha hecho referencia a confidencialidad de la información tratada, hoy en día se suele incluir además la integridad de la información y la disponibilidad de los recursos hardware y software empleados.

Respecto a la preocupación o interés que despiertan los asuntos relacionados con la seguridad, podemos constatar dos realidades:

- Esta preocupación es mayor cuanto mayor es el tamaño de la empresa.
- Esta preocupación se traduce en documentos, actuaciones, normativas o recomendaciones que podemos agrupar bajo el término genérico de *políticas de seguridad*. Los procedimientos para establecer estas políticas de seguridad y para estudiar la seguridad de los sistemas, son variados y podemos decir que se atacan de una manera sistemática solo en las empresas de mayores dimensiones.

Por otra parte, en la actualidad, se dan dos nuevas circunstancias:

- Irrumpen con fuerza en empresas de todos los tamaños dos nuevas tecnologías interrelacionadas: la conexión con Internet y el concepto de Intranet.
- La disminución del tamaño medio de las empresas que acceden a estos recursos.

Esto hace que aparezca un nuevo colectivo de usuarios con necesidades de seguridad caracterizado por el empleo de una red de tamaño pequeño o mediano, con una solución de proceso de datos basada en PC's y a veces miniordenadores y con una disponibilidad limitada de recursos económicos. Además en muchos de estos casos, la infraestructura informática en uso no está diseñada ni implantada con estas necesidades en perspectiva.

Dado que un punto tipico de inflexión es la conexión a Internet, esto suele implicar entre otros, una migración desde protocolos tradicionales de red como IPX/SPX o NetBEUI a protocolos TCP-IP, la implantación de servidores con sistemas operativos tales como Unix© o NT©. Esto puede provocar que en un primer momento la experiencia de los departamentos informáticos sea pequeña o nula.

Vamos a estudiar que es lo que se ofrece en la actualidad desde el punto de vista de metodología para el estudio de la seguridad. Veremos las limitaciones que en opinión de los autores tienen estas propuestas. Se propondrá un modelo de seguridad que permite atacar a la problemática especifica de estos usuarios y lograr conocer: Cual es la mejor politica de seguridad que se puede implantar con los recursos disponibles, y cuales son los puntos débiles de su sistema.

2. Antecedentes

La seguridad en Sistemas Informáticos ha sido históricamente un motivo de preocupación para los organismos oficiales de gran parte de países. Esta preocupación se ha manifestado de varias formas, una de las cuales, ha sido la generación de los denominados *"manuales de evaluación de seguridad"*, en los que se recogen criterios, metodologías y definiciones para la evaluación de la seguridad de sistemas de tratamiento de la información.

A continuación se trace una reseña de los manuales más conocidos, estudiando la evolución de los mismos.

El primero y precursor del resto, es el *"Trusted Computer System Evaluation Criteria"* (TCSEC) o "Libro Naranja" del Departamento de Defensa de Estados Unidos de 1983. [1]. En otros países, mayoritariamente europeos, también se empiezó a trabajar en el desarrollo de diversos manuales que perseguían el mismo fin.

Más tarde, apoyándose en estos manuales nacionales surge una iniciativa por parte de Francia, Alemania, Holanda y Gran Bretaña que tiene como fruto el *"Information Technology Security Evaluation Criteria"*. (ITSEC) Este manual es adoptado en su versión 1.2 del año 1991 por la Unión Europea. [2]

Surge a continuación en Canadá un esfuerzo por armonizar los manuales TCSEC y ITSEC, y surge *"Canadian Trusted Computer Product Evaluation Criteria"* (CTCPEC) en su versión 3.1 en 1993. [3]

Paralelamente el gobierno de los Estados Unidos, crea un borrador del *"Federal Criteria for Information Technology Security"* (FC) en 1993 con objeto de combinar las aproximaciones europeas y norteamericanas en un solo documento. [4]

El último esfuerzo normalizador parte de un trabajo de los creadores de TCSEC, ITSEC, CTCPEC y FC con objeto de crear un documento que se pueda elevar a la categoría de estándar ISO. Este es el *"Common Criteria for Information Technology Security Evaluation"* que está actualmente en la revisión 1.0. [5]

Los motores de esta evolución han sido entre otros la necesidad de los gobiernos de simplificar la evaluación de sistemas con requerimientos de seguridad importantes y por otra parte la necesidad de los fabricantes de enfrentarse a un único juego de requerimientos. La evolución ha sido constante y orientada hacia la consecución de un manual de carácter global.

Todos estos manuales, persiguen un mismo objetivo: Facilitar la evaluación de las características de seguridad de productos y sistemas empleados en las Tecnologías de la Información a los compradores, desarrolladores y evaluadores de los mismos.

Ampliando un poco el objetivo, podemos decir de una manera resumida que:

- A los *compradores*, les ayuda a decidir si un producto cumple con sus requerimientos de seguridad. Estas necesidades de seguridad suelen consistir en el resultado de armonizar una política de seguridad con un análisis de riesgos. También sirve para decidir la adquisición de sistemas constituyendo una herramienta pare poder compararlos desde el punto de vista de las necesidades de seguridad.
- A los *desarrolladores*, les permite preparar los sistemas para que superen las evaluaciones de seguridad. Se describen también las acciones que un desarrollador debe realizar para presentar evidencias de que un producto cumple los requerimientos de seguridad.
- A los *evaluadores*, les detalla los criterios a seguir y metodologías a emplear

Esta aproximación a la evaluación de la seguridad de sistemas está particularmente bien adaptada a la adquisición de sistemas con requerimientos específicos de seguridad.

Sin embargo, esta metodología en opinión de los autores, puede no ser la más adaptada a la situación descrita anteriormente debido a que en estos casos:

- El costo que supone aplicar esta metodología si los sistemas no están ya evaluados a priori, puede ser muy alto.
- Hay ya una infraestructura instalada que no puede ser totalmente cambiada.
- El proceso puede ser relativamente lento.
- En muchos casos la granularidad con la que se trata la seguridad es excesiva.

También existe otro tipo de ayudas, en forma de manuales de seguridad. Pretender dar directivas de una manera más práctica para la implantación de seguridad. Son exponentes de este tipo de manual, los publicados por el National Institute of Standards and Technology (NIST) dependiente del Ministerio de Comercio de Estados Unidos. Uno de los más conocidos es *"An Introduction to Computer Security: The NIST Handbook"* [6]

Estos manuales tienen un carácter marcadamente didáctico y dan ideas prácticas sobre todo a nivel organizativo. De cara a una instalación real y estudiando la problemática a un nivel medio/bajo no dan respuesta adecuada.

En opinión de los autores, se echa en falta una metodología que permita estudiar la seguridad de sistemas informáticos con las siguientes características:

- Adaptados a instalaciones actualmente en marcha.

- Adaptados a empresas de tipo pequeño o medio con recursos económicos limitados.
- Basadas mayoritariamente en PC's y/o miniordenadores.
- Unidas por un red de área local.

3. Estudio de la Seguridad de un Sistema basándose en modelos.

En opinión de los autores, una alternativa para el estudio de la seguridad de sistemas de tratamiento de información se basa en el empleo de modelos.

Vamos a describir a continuación, que es un modelo de seguridad. A continuación describiremos el modelo de seguridad propuesto.

3.1 Concepto de Modelo de Seguridad

En el contexto que nos ocupa, un modelo es una representación esquemática de un sistema orientado al tratamiento de determinada problemática. Este modelo, debe destacar los puntos críticos a tratar.

Un modelo de este tipo es el punto de partida para estudiar la problemática de una manera sistemática y exhaustiva.

Un modelo de seguridad, es por tanto una representación esquemática de un sistema orientada al tratamiento de la problemática de la seguridad.

Dos típicos modelos de seguridad propuestos por [7] son el modelo de seguridad de red (Fig. 1) y el modelo de seguridad de acceso a red (Fig. 2).

En el primer modelo, se define la problemática de la comunicación a través de una red, y se muestran las partes implicadas y la relación entre las mismas.

Por medio de este modelo se llega a estudiar las soluciones basadas en seguridad a nivel de red, protocolo, o aplicaciones. Se llega también a las aplicaciones basadas en redes privadas virtuales.

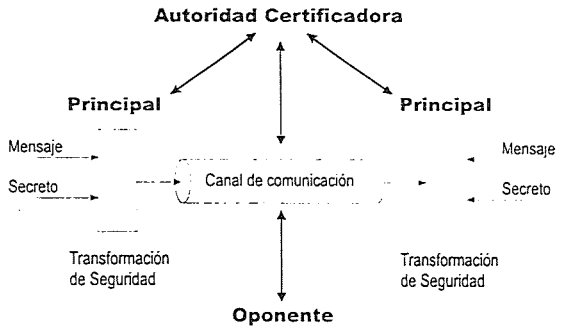


Figura 1: Modelo de Seguridad de Red

En el segundo modelo, se estudia la problemática del acceso a unos recursos informáticos desde la red. Por medio de este modelo se llega, por ejemplo, a la soluciones basadas en sistemas cortafuegos.

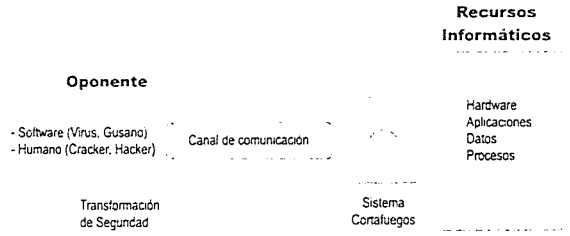


Figura 2: Modelo de Seguridad de Acceso a Red

Nuestra experiencia es que cuando se estudia la seguridad de una pequeña empresa, estos dos modelos, solo resuelven parte de la problemática. Solo se refieren a la interconexión de sistemas y se pasa por alto todo lo correspondiente a la seguridad interna. Una justificación puede consistir en que la correcta aplicación de los dos modelos anteriores, permite conseguir aislar la infraestructura interna del exterior.

Sin embargo, hoy en día, hay un sentimiento colectivo de que la seguridad interna es previa antes de plantearse la conectividad externa (a Internet, por ejemplo). Además se, considera probado que un porcentaje grande de los ataques a la seguridad (en los tres conceptos de confidencialidad, integridad y disponibilidad) parten del interior de la empresa.

Para suplir esta carencia, los autores proponemos una metodología basada en un modelo de seguridad adicional denominado *"Modelo de Seguridad Interno"* que trata de suplementar los dos modelos existentes para conseguir un estudio sistemático de la seguridad.

3.2 Modelo de Seguridad Interno

Este modelo está definido inspirándose en algunos de los principios utilizados en la definición de las capas OSI (ISO 7498). De hecho, algunas de las capas, tienen incluso nombres similares a los de alguna de las capas del modelo de referencia de Interconexión de Sistemas Abiertos, aun cuando no coincidan totalmente.

El modelo se presenta en la Fig. 3. Para explicarlo con claridad, vamos a definir el sentido de los términos empleados en el mismo.

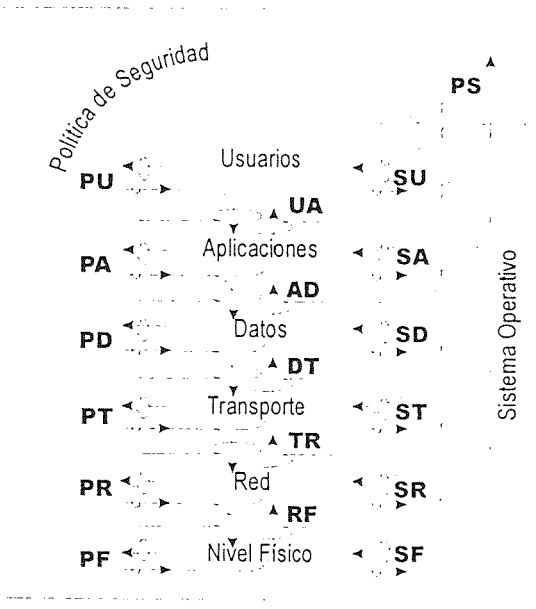


Figura 3: Modelo de Seguridad Interno

- **Política de seguridad**: Se denomina Política de Seguridad al conjunto de normas, actuaciones y recomendaciones que definen el comportamiento desde el punto de seguridad de las partes implicadas en una solución de tratamiento, transmisión, almacenamiento y recuperación de datos. La política de seguridad es algo deseado y definido por la dirección de la empresa.
- **Usuarios**: Son las personas físicas, interaccionan a través de aplicaciones que controlan en tiempo real o diferido con datos, con programas u otros usuarios.
- **Aplicaciones**: Son las herramientas a través de las cuales, los usuarios interaccionan en tiempo real o diferido con datos, otras aplicaciones u otros usuarios. No se incluyen las aplicaciones que son parte del sistema operativo.
- **Datos**: Es la información útil que manipulan los usuarios. No se incluyen los datos que son necesarios para el funcionamiento del sistema operativo como tal.
- **Transporte**: Es el sistema que se emplea para transportar la información a través de la red. Hace referencia de una manera muy clara a las características de los protocolos de transporte que se van a emplear.
- **Red**: Es el sistema que se va emplear para encaminar la información desde un ordenador a otro. Hace como en el caso anterior, refererencia muy clara al nivel de red del protocolo a emplear.
- **Físico**: Es sistema constituido por el medio físico y el protocolo de nivel físico que se va a emplear para transmitir datos de un sistema a otro. Hace referencia a los niveles físico y de enlace. Se tienen en cuenta las implicaciones que desde el punto de vista de seguridad implican los elementos activos y pasivos necesarios para formar la red.
- **Sistema Operativo**: Es el software encargado de gestionar el conjunto de recursos hardware que constituyen el ordenador (CPU, RAM, almacenamiento masivo, dispositivos de E/S), realizar el interfaz de la máquina con el usuario o usuarios, y ejecutar los programas de aplicación de los usuarios

Hay que tener en cuenta que en este modelo, las aplicaciones que son parte del sistema operativo están consideradas dentro del elemento con el que se relacionan.

Estos componentes genéricamente estarán casi siempre presentes, aun cuando, es posible que a veces, alguno de estos componentes sea casi nulo.

Se emplea dentro de la descripción del modelos también los siguientes términos:

- **Interfaz:** Es el punto de conexión entre dos elementos. Puede en principio tener tráfico bidireccional: Requerimientos y Consecuencias.
- **Requerimiento:** Es la condición que un elemento o la política de seguridad le exige cumplir a otro elemento desde el punto de vista de la seguridad.
- **Consecuencia:** Entendemos como consecuencia directa, la que resulta de la aplicación de una elección en un elemento. En el resto de los casos, podemos suponer que es un resultado no deseado (desde el punto de vista de la seguridad) de la aplicación de un requerimiento a una determinada capa. Como veremos, es el punto inicial de una realimentación.

A continuación se detallan los distintos interfaces.

3.2.1 Interfaces de Tipo P

Habitualmente a través de este tipo de interfaz, se transmiten los requerimientos de la política de seguridad de la empresa a los componentes. En estos interfaces no se pretende justificar una elección, sino transmitirla. Se supone que esta elección estará tomada desde la óptica de la seguridad. También se transmiten en sentido inverso consecuencias. Se va a describir los requerimientos en base a ejemplos.

PS: En este interfaz se define la elección de un sistema operativo. Un ejemplo de este requerimiento inicial puede ser: "*Todos los sistemas operativos a emplear deben ser de la Marca X*", *El sistema operativo debe ser Y*" o "*Nos da lo mismo el sistema operativo que se emplee*". En consecuencia, debido a esta elección, se definen también los interfaces de

tipo S. No hay que olvidar que la elección de Sistema Operativo, implica de cierta manera también la elección del hardware a muchos niveles, como puede ser a nivel de CPU, capacidades de I/O, etc. Hay que incluir también los requerimientos que permitan la recuperación de funcionalidad de sistemas en casos de averías o fallos y la instalación sistemática de parches o actualizaciones del S.O. que sean necesarios para mantener la seguridad del mismo de acuerdo a las directivas del fabricante del S.O.

PU: En este interfaz se define el normas de uso del sistema para los usuarios. Ejemplos de requerimientos de este tipo son: "*No se puede utilizar la cuenta de otro usuario*", "*No se puede dar la cuenta a nadie*", "*No se puede dejar una cuenta abierta sin estar presente*", "*No se puede instalar software de ningún tipo en el ordenador*", "*Todos los usuarios deben pertenecer a uno de los siguientes grupos:...*", "*Deben utilizarse claves de un solo uso*", "*La autentificación de usuarios debe realizarse por medio de elementos hardware*", "*Los permisos deben estar basados en ACL*", "*Debe poderse controlar y limitar el espacio ocupado en disco por usuario*"...

PA: En este interfaz se define la elección de las aplicaciones de usuario. Ejemplos de estos requerimientos son: "*Los documentos de texto deben ser generados empleando el procesador de texto Y*", "*La ofimática estará basada en la oferta de la empresa Z*", "*Las aplicaciones implicadas deben disponer de un contrato de mantenimiento*", "*El agente de usuario que se va a emplear es W*", "*Las aplicaciones implicadas no deberán ejecutarse con atributos de superusuario*", "*Las aplicaciones empleadas deben ser capaces de generar un registro de operaciones realizadas*", para el caso de que las aplicaciones implicadas sean Browsers de WWW, por ejemplo: "*No se permitirá la ejecución de aplicaciones Java o Active-X externas a la empresa*"...

PD: En este interfaz se define el manejo de los datos. Ejemplos de estos requerimientos son: "*El formato de los nombres de fichero será el siguiente:...*", "*Los ficheros de tal tipo, deben ser almacenados en tal directorio*", "*No se debe poder sacar datos de la empresa en forma de diskettes o listados*", "*El formato de los ficheros de tipo P será el Q*", "*Los datos deben guardarse encriptados de tal manera que el único acceso a los mismos sea a través de la aplicación y a los usuarios autorizados*", "*Deben realizarse copias de seguridad con una periodicidad X*", "*Los ficheros donde residen los datos de la aplicación deben estar a salvo de manipulaciones causadas por todos los usuarios, a excepción del Superusuario*", "*No van a existir unidades de almacenamiento removible en los ordenadores*"...

PT: En este interfaz se definen los protocolos de transporte disponibles y el acceso y empleo de los mismos.. Ejemplos de requerimientos que van a aparecer en este interfaz son: "*No debe ser posible utilizar la red para establecer canales de comunicación mediante protocolos no autorizados*", "*Se emplearán protocolos orientados a conexión para conexiones externas*", "*La capa de transporte debe ser capaz de ofertar un servicio confirmado*"...

PR: En este interfaz se define el tipo de interfaz de red que se hará disponible a los usuarios y que definirá el uso de los mismos. Ejemplos de requerimientos que se pueden dar aquí son: "*Los servicios de red que se deberán prestar con Tecnología IP*", "*Los datos deben ir encriptados y autenticados a nivel de Red*", "*El protocolo de red a emplear debe ser enrutable*", "*El protocolo de red, debe maximizar el rendimiento con tráfico formado por paquetes de datos de tamaño X (el que más se da con las aplicaciones en uso)*", "*Debe poderse configurar y eliminar los servicios del sistema operativo que se ofertan sobre la red*", (Nótese que las aplicaciones y servicios propios del Sistema Operativo que hacen uso de la red, se incluyen aquí en vez de en el apartado de aplicaciones").

PF: En este interfaz se definen los requerimientos correspondientes al nivel físico y de enlace de un modelo de capas OSI. Ejemplos de requerimientos típicos transmitidos por este interfaz pueden ser: "*No debe ser posible monitorizar la información que circula*", "*Todos los equipos deben tener la misma posibilidad de transmitir información*", "*Físicamente, la información no debe ser accesible a equipos no implicados en la comunicación*", ...

3.2.2 Interfaces de Tipo S

En estos interfaces, transita información en los dos sentidos: Por un lado, hay consecuencias directas fijadas por la elección del Sistema Operativo que pasan a los diversos elementos. Y por otro lado, se transmiten consecuencias desde los elementos.

SU: En este interfaz, se transmiten las consecuencias directas fijadas por la elección del sistema operativo hacia la capa de usuario. Ejemplos de los consecuencias directas de este interfaz son: "*El sistema operativo no permite perfiles de usuario*", "*El sistema operativo requieren una identificación y autenticación previa para interactuar con el mismo*"... Las consecuencias utilizarían también este interfaz pero en sentido inverso.

SA: En este interfaz, se transmiten las consecuencias directas fijadas por la elección del sistema operativo hacia la capa de aplicación. Ejemplos de los consecuencias directas que pasan a través de este interfaz son: *"Todos los procesos tienen los mismos permisos"*, *"No se pueden ejecutar varias aplicaciones simultáneamente"*... Como en el caso anterior, Las consecuencias utilizarían también este interfaz pero en sentido inverso

SD: En este interfaz, se transmiten las consecuencias directas fijadas por la elección del sistema operativo hacia la capa de datos. Ejemplos de las consecuencias directas que se transmiten a través de este interfaz son: *"Existe la posibilidad de emplear un acceso a los datos basado en ACL"*, *"No existe posibilidad de evitar el acceso a determinados datos al usuario que ostente atributos de superusuario"*... Las consecuencias utilizarían también este interfaz pero en sentido inverso.

ST: En este interfaz, se transmiten las consecuencias directas fijadas por la elección del sistema operativo hacia la capa de transporte. Ejemplos de las mismas son: *"Se dispone de protocolos orientados a conexión"*, *"Se dispone de protocolos que aseguran la confidencialidad y autenticidad"*... ,como se ha visto anteriormente, en sentido inverso, las consecuencias también podrían utilizar este interfaz.

SR: En este interfaz, se transmiten las consecuencias directas fijadas por la elección del sistema operativo hacia la capa de red. Ejemplos de esto son: *"Este sistema operativo no puede realizar enrutamiento sobre TCP-IP"*, *"Cualquier usuario puede cambiar la dirección IP de este ordenador"*, *"Cualquier usuario que emplea la máquina puede crear sockets con cualquier valor de puerto origen"*... También son bidireccionales como los anteriores.

SF: En este interfaz, se transmiten las consecuencias directas fijadas por la elección del sistema operativo hacia la capa Física. Ejemplos de los consecuencias directas son: *"Cualquier usuario puede cambiar la dirección ethernet de su adaptador de red"*. Hay que tener en cuenta que estas consecuencias tienen en cuenta aspectos concretos provenientes del hardware tanto a nivel del adaptador, como de los elementos activos y pasivos en uso. También tiene bidireccionalidad.

3.2.3 Interfaces Internos

Existen en los interfaces entre capas del modelo. En realidad la información que se transmite a través de estos interfaces en los dos sentidos, son producto de la actuación de los requerimientos de tipos P y S sobre las capas interesadas y sus consecuencias.

UA: En este interfaz se transmiten requerimientos y consecuencias entre la capa de usuario y la de aplicación

AD: En este interfaz se transmiten requerimientos y consecuencias entre la capa de aplicación y la de datos.

DT: En este interfaz se transmiten requerimientos y consecuencias entre la capa de datos y la de transporte.

TR: En este interfaz se transmiten requerimientos y consecuencias entre la capa de transporte y la de red.

RF: En este interfaz se transmiten requerimientos y consecuencias entre la capa de red y la física.

4. Metodología de Aplicación

Este modelo, permite variando ligeramente la metodología, obtener resultados para tres casos diferentes:

- Por un lado permite estudiando una determinada instalación, obtener cuáles son los puntos en los que la política de seguridad de la empresa no son respaldados por la instalación y sugerir cambios.
- Por otro lado, permite estudiar a partir de una determinada instalación cual es la mejor política de seguridad que se puede implantar.
- Permite por último planificar una instalación de tal manera que a partir de una política de seguridad determinada se pueda obtener la infraestructura hardware y software necesaria para el cumplimiento de la misma.

La metodología de uso de este modelo está en estos momentos en fase de mejora y definición formal. Lo que se desarrolla a continuación son los fundamentos.

Se aplicará el modelo para cada uno de los dispositivos distintos que nos podemos encontrar. Un dispositivo es igual a otro cuando todos los elementos que lo componen son iguales.

El modelo se completa con información de la situación inicial que se extrae de la instalación. Se estudia el sistema operativo de PC's y microordenadores, los sistemas operativos de red instalados en servidores, las aplicaciones, los datos en juego, la implantación de la red de área local, el hardware empleado en el equipamiento tanto informático como de comunicaciones. A continuación, se detallan las consecuencias directas en los interfaces de tipo S, las consecuencias y requerimientos en los interfaces internos.

Si, por ejemplo, el objetivo que se pretende es el primero, se estudia la política de seguridad implantada y a continuación, se completan los requerimientos de los interfaces P.

A partir de este momento, empieza un procedimiento iterativo. Cuando un punto de la política de seguridad genera un requerimiento, este va a parar al elemento. En este elemento, confluyen las consecuencias que provienen de su elemento superior e inferior más las que le llegan a través del interfaz S correspondiente. Si todas las informaciones que confluyen en el elemento, son compatibles, no se generan consecuencias, si esto no es así, se generan consecuencias, bien sobre los nivels adyacentes o sobre el elemento S a través de los interfaces S o sobre la política de seguridad directamente. Tenemos que ir modificando el valor de los elementos en juego, hasta que todos los elementos son capaces de absorber los requerimientos y consecuencias sin generar nuevas consecuencias.

El elemento es inmutable en función de nuestras necesidades. Por ejemplo: Podemos cambiar una aplicación si buscamos cumplir la política de seguridad relativa a esa aplicación; si no queremos o podemos cambiar la aplicación, tendremos que modificar la política de seguridad para que el requerimiento generado previamente desaparezca y la política refleje lo que le puede entregar el sistema en cuanto a seguridad.

Cuando llegamos a un punto en el que al cambiar un elemento, podemos escoger entre varias opciones, se deberían usar criterios adicionales tales como el económico, el seguimiento de un estándar, etc.

Este modelo interno tal como lo hemos presentado es un modelo genérico adaptable a cualquier dispositivo. Este modelo puede simplificarse en función del dispositivo al que se aplique.

Por ejemplo, los *clientes* de usuario, basados en sistemas operativos monousuario, simplifican en gran medida el estudio ya que los interfaces de tipo S, no tienen gran capacidad de configuración. Sin embargo en los *servidores* es de singular relevancia el capítulo de la definición de los interfaces de tipo S. Esto es debido a que por un lado son sistemas multiusuario o bien tienen la capacidad de aceptar conexiones de usuario, por otro lado son objetivo prioritario de los atacantes y para acabar, el sistema operativo tiene gran capacidad de configuración y multitud de opciones. También se podrían estudiar los aparatos implicados en las comunicaciones, tales como cortafuegos, routers y pasarelas.

5. Conclusiones y trabajo futuro

En este momento el modelo y la metodología son muy nuevos, sin embargo, en opinión de los autores y a la luz de las primeras aplicaciones de este modelo y metodología asociada a la resolución de problemas reales el método es válido.

En estos momentos, el trabajo se centra en varios frentes: estudiar si el modelo se adapta a todos las posibilidades de requerimientos y consecuencias, dar al modelo y metodología asociada un tratamiento más formal, definir un sistema que permita definir de una manera normalizada los requerimientos y consecuencias, personalizar el modelo para componentes estándar de una infraestructura. Todo esto a la luz y de manera compatible con los publicaciones que tratan la evaluación de la seguridad en sistemas de tratamiento de la información como el CC.

Referencias

[1] U.S. Department of Defense, "Trusted Computer Systems Evaluation Criteria (TCSEC)", US DoD 5200.28-STD, Diciembre 1985.

[2] Office for Official Publications of the European Communities, "Information Technology Security Evaluation Criteria (ITSEC)", Version 1.2, Junio 1991.

[3] Canadian System Security Centre, Communications Security Establishment, Government of Canada, "Canadian Trusted Computer Product Evaluation Criteria (CTCPEC)", Version 3.0, Enero 1993.

[4] National Institute of Standards and Technology and the National Security Agency, "Federal Criteria for Information Technology Security (FC)", Draft Version 1.0, (Volumes I and II), US Government, Enero 1993.

[5] The National Institute of Standards y Technology National Security Agency, National Security Agency de USA, Communications Security Establishment de Canada, UK IT Security and Certification Scheme, Bundesamt für Sicherheit in der Informationstechnik de Alemania, Service Central de la Sécurité des Systèmes d'Information de Francia y Netherlands National Communications Security Agency de Holanda, "Common Criteria for Information Technology Security Evaluation (CC)", Version 1.0, Enero 1996

[6] NIST, Technology Administration, U.S. Department of Commerce, "An Introduction to Computer Security: The NIST Handbook", Draft, , Junio 1994.

[7] Stallings, Williams, "Network and Internetwork Security principles and practice", New Jersey: Prentice Hall, 1995

Sistema Seguro de Correo Electrónico

Gonzalo Alvarez Marañón y Fausto Montoya Vitini
DEPARTAMENTO DE TRATAMIENTO DE LA INFORMACIÓN Y CODIFICACIÓN
INSTITUTO DE FÍSICA APLICADA, CSIC
SERRANO 144, 28006 MADRID
Correo electrónico: gonzalo@iec.csic.es

Abstract:

Both in the business and science world there is a growing use of e-mail as a means of transference of messages and information, within the corporate or University intranet and over Internet. However, as a consequence of the lack of security tools in most messaging systems, the threat to confidentiality and integrity of data inhibits the free use of Internet with commercial purposes or for secret information interchange over e-mail. The Secure E-Mail System presented in this work represents a big improvement in the security features of e-mail clients, supplying the necessary cryptographic tools to ensure privacy in interchanged messages, thus stimulating the widespread use of e-mail services.

1. Introducción

Desde las pequeñas empresas hasta las grandes compañías, pasando por las universidades y usuarios particulares, el uso de la mensajería electrónica como medio rápido y económico de intercambio de información está creciendo espectacularmente. Sin embargo, se trata de una de las aplicaciones más vulnerables sobre Internet, que no ofrece ninguna seguridad, lo que ha coartado su crecimiento, inhibiendo el uso de Internet para las comunicaciones y comercio electrónico, puesto que los mensajes de e-mail pueden ser leídos, alterados o borrados sin conocimiento del emisor o receptor.

El sistema descrito en este trabajo es un *plug-in* para *Microsoft Exchange*, ofreciendo a los usuarios un conjunto de herramientas para crear un sistema de correo electrónico seguro, tanto en la Intranet como a través de la Internet.

El sistema, integrado en la bandeja de entrada (*inbox*) de *Windows 95* o *Windows NT (Windows Messaging Service)*, extiende las barras de herramientas y menús del cliente, poniendo a disposición del usuario funciones de seguridad como cifrado transparente del cuerpo del mensaje y firma digital de los documentos.

Como algoritmo de clave secreta, utiliza una variante de *Akelarre* [1], diseñado en este laboratorio, mientras que como algoritmo de clave pública se sirve de una mejora de Blum, Blum y Shub, descrita en [2], también desarrollada en este laboratorio. Como algoritmo de resumen utiliza el propio *Akelarre* con clave secreta conocida.

2. El cliente *Microsoft Exchange*

2.1 ¿Por qué *Microsoft Exchange*?

Ante el desafío de desarrollar un cliente de correo seguro, se plantea en primer lugar la posibilidad de escribir la aplicación a medida completamente desde cero o bien considerar la ampliación de un cliente ya existente.

Durante años se viene trabajando en este departamento con los sistemas operativos *Windows 95* y *Windows NT*. Después de examinar varios clientes de correo electrónico, mejor o peor adaptados a *Windows*, se descubrió que *Microsoft Exchange* presenta indudables ventajas frente a otros sistemas, puesto que fue diseñado como una aplicación extensible y adaptable según las necesidades del usuario (*customizable*). Se puede modificar y mejorar gradualmente el funcionamiento por defecto del cliente *Exchange* y añadir nuevos comandos sobre sus barras de menús y de herramientas, permitiendo así integrar aquellas características que faltan en él o modificar las que no se adapten a nuestros deseos [3].

Por estos motivos, crear una aplicación de seguridad sobre el cliente *Exchange* ofrece muchas ventajas inmediatas tanto al ingeniero de software como al usuario. La más obvia es que suministra el marco para la interface de usuario de la aplicación como un componente acabado y listo para ejecutar, sin necesidad de un esfuerzo adicional de desarrollo. Además, proporciona a los usuarios de la aplicación criptográfica un entorno y una interface familiares. Por último, puesto que el cliente *Microsoft Exchange* viene incluido en *Windows 95*, la extensión criptográfica puede ejecutarse sobre cualquier ordenador cuyo sistema operativo sea *Windows 95*, no siendo necesario que el usuario adquiera un producto separado para usarla bajo *Windows 95*.

2.2 ¿Qué hay debajo de *Microsoft Exchange*?

El cliente *Exchange* consiste en un único proceso que gestiona la operación de múltiples ventanas independientes, que pueden ser de dos clases: vistas y formularios, los cuales constituyen el fundamento de la arquitectura vista/formulario característica del cliente *Exchange* y de todas sus aplicaciones derivadas [4].

El cliente es una aplicación creada sobre MAPI (*Messaging Application Programming Interface*), una arquitectura para mensajería que permite a múltiples aplicaciones interactuar con múltiples sistemas de mensajería de forma integrada sobre una amplia variedad de plataformas. MAPI se compone de un conjunto de interfaces de

programación de aplicaciones, empleadas para crear y acceder diversas aplicaciones y sistemas de mensajería desde un entorno uniforme, y de un componente de bibliotecas dinámicas (DLL), que contiene el subsistema MAPI para gestionar la interacción entre las aplicaciones cliente y los proveedores de servicio [5].

MAPI se apoya en el estándar *Microsoft* de COM (*Component Object Model*), que define e implementa mecanismos para permitir que componentes software como aplicaciones, objetos de datos, controles y servicios interactúen como *objetos*. Un objeto software comprende un cuerpo de datos, junto con una o más funciones, llamadas *interfaces*, para acceder y usar esos datos [6].

3. Seguridad incorporada a *Microsoft Exchange*

3.1 Servicios de seguridad

Hasta la fecha se han incluido los siguientes servicios de seguridad, integrados en el sistema de correo:

- *Confidencialidad*: asegura que la información intercambiada sólo puede ser accedida por las entidades autorizadas.
- *Integridad*: asegura que la información no ha sido modificada por entidades no autorizadas. La modificación incluye escritura, cambio, borrado, creación y reactuación de los mensajes transmitidos.
- *Autenticación*: permite identificar correctamente al origen del mensaje, asegurando que la entidad no es falsa.
- *No repudio*: ofrece protección a un usuario frente a que otro usuario niegue posteriormente que en realidad se realizó cierta comunicación.

3.2 Mecanismos de seguridad

Con el fin de proveer los servicios anteriormente enumerados, se han implementado los siguientes mecanismos de seguridad:

3.2.1 Cifrado

Garantiza que la información no es inteligible para entidades no autorizadas, esto es, garantiza la confidencialidad de la información. Consiste en transformar un texto en claro mediante un mecanismo de cifrado en un texto cifrado, gracias a una información secreta o clave de cifrado.

Cifrado simétrico

Cuando se emplea la misma clave en las operaciones de cifrado y descifrado, se dice que el criptosistema es simétrico o de clave secreta. Estos sistemas son mucho más rápidos que los de clave pública, y resultan apropiados para el cifrado de grandes volúmenes de datos.

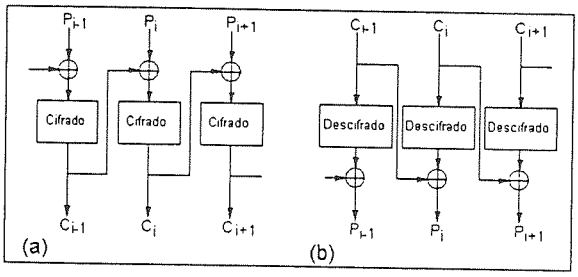


Figura 1. Modo de cifrado en bloque CBC: a) proceso de cifrado; b) proceso de descifrado.

Ésta ha sido la opción utilizada para cifrar el cuerpo del mensaje. Para ello se ha empleado el cifrado en bloque con una variante del algoritmo *Akelarre*, descrito en la sección 5.1.

También se emplea este algoritmo a la hora de generar resúmenes a partir de un documento, cifrando en modo CBC (*Cipher Block Chaining*), mostrado en la Fig. 1. En este modo de cifrado en bloque, el resultado de cifrar los bloques previos se realimenta al cifrado del bloque en curso. En otras palabras, cada bloque se emplea para modificar el cifrado del bloque siguiente, de modo que cada bloque de texto cifrado depende no solamente del texto en claro que lo ha generado, sino también de todos los bloques de texto en claro anteriores a él. Por consiguiente, aprovechando esta propiedad del CBC, si se toma como resumen del documento el último bloque resultante del cifrado de todo el documento con una clave conocida, se habrá obtenido un resumen unidireccional del mismo (cf. 5.3).

Cifrado asimétrico

Por otro lado, cuando se utiliza una pareja de claves para separar los procesos de cifrado y descifrado, se dice que el criptosistema es asimétrico o de clave pública. Una clave, la privada, se mantiene secreta, mientras que la segunda clave, la pública, es conocida por todos. De forma general, las claves públicas se utilizan para cifrar y las privadas, para descifrar. El sistema posee la propiedad de que a partir del conocimiento de la clave pública no es posible determinar la clave privada ni descifrar el texto con ella cifrado. Los criptosistemas de clave pública, aunque más lentos que los simétricos, resultan adecuados para los servicios de autenticación, distribución de claves de sesión y firmas digitales, como se explicará posteriormente.

En este sistema, el cifrado asimétrico se emplea para cifrar las claves de sesión utilizadas para cifrar el documento, de modo que puedan ser transmitidas sin peligro a través de la red junto con el documento cifrado, para que en recepción éste pueda ser descifrado. La clave de sesión se cifra con la clave pública del destinatario del mensaje, que aparecerá en una libreta de claves públicas (cf. 4.3).

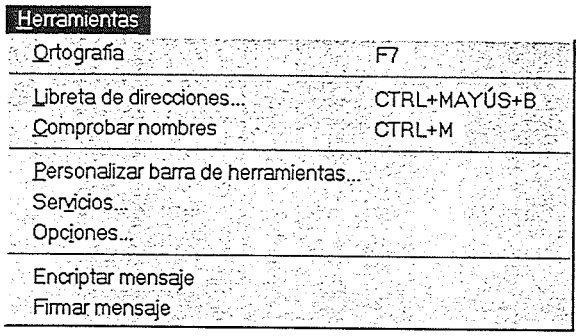


Figura 2. Menú de herramientas extendido de la ventana "Mensaje nuevo" de Microsoft Exchange.

El cifrado asimétrico se emplea también para firmar documentos y autenticar entidades, como se describe a continuación.

3.2.2 Firma digital

En principio, basta con cifrar un documento con la clave privada para obtener una firma digital segura, puesto que nadie excepto el poseedor de la clave privada puede hacerlo. Posteriormente, cualquier persona podría descifrarlo con la clave pública, demostrándose así la identidad del firmante. En la práctica, debido a que los algoritmos de clave pública son muy ineficaces a la hora de cifrar documentos largos, los protocolos de firma digital se implementan junto con funciones unidireccionales de resumen (*hash*), de manera que en vez de firmar un documento, se firma un resumen del documento. Este mecanismo implica el cifrado, mediante la clave privada del emisor, del resumen de los datos, que serán transferidos junto con el mensaje. Éste se procesa en el receptor, para verificar su integridad. Por lo tanto, los pasos del protocolo son:

1. *A* genera un resumen del documento.
2. *A* cifra el resumen con su clave privada, firmando por tanto el documento.
3. *A* envía el documento junto con el resumen firmado a *B*.
4. *B* genera un resumen del documento recibido de *A*, usando la misma función unidireccional de resumen. Después descifra con la clave pública de *A* el resumen firmado. Si el resumen firmado coincide con el resumen que él ha generado, la firma es válida.

De esta forma se ofrecen conjuntamente los servicios de no repudio, ya que nadie excepto *A* podría haber firmado el documento, y de autenticación, ya que si el documento viene firmado por *A*, podemos estar seguros de su identidad, dado que sólo él ha podido firmarlo. En último lugar, mediante la firma digital se garantiza asimismo la integridad del documento, ya que en caso de ser modificado, resultaría imposible hacerlo de forma tal que se generase la misma función de resumen que había sido firmada.

4. Extensión criptográfica del cliente *Exchange*

Gracias a la extensión criptográfica del cliente *Exchange*, el usuario dispone ahora de forma totalmente integrada de las siguientes dos opciones:

Cifrado del documento

En el menú "Herramientas" de la ventana "Mensaje nuevo" (Fig. 2) se añade la opción "Encriptar mensaje", representada en la barra de herramientas por el botón . Cuando se presiona este botón o se elige la opción en el menú, el

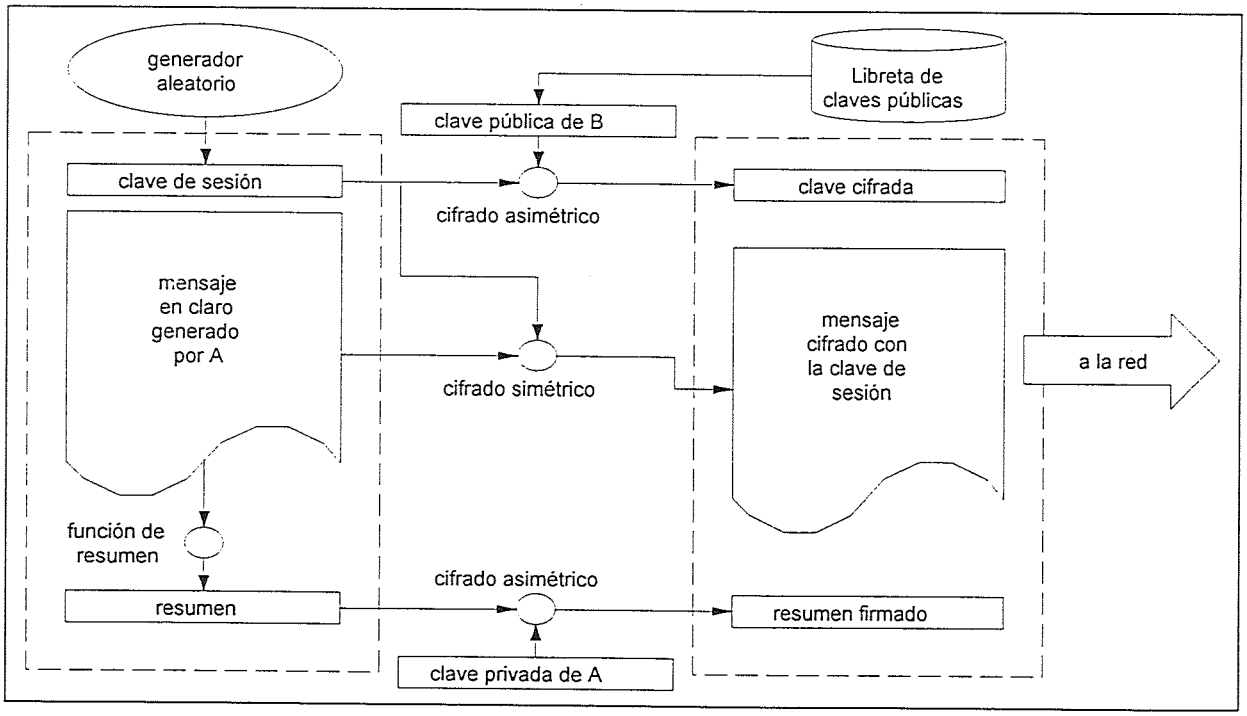


Figura 3. Diagrama de bloques del procedimiento de envío de mensajes cifrados y/o firmados.

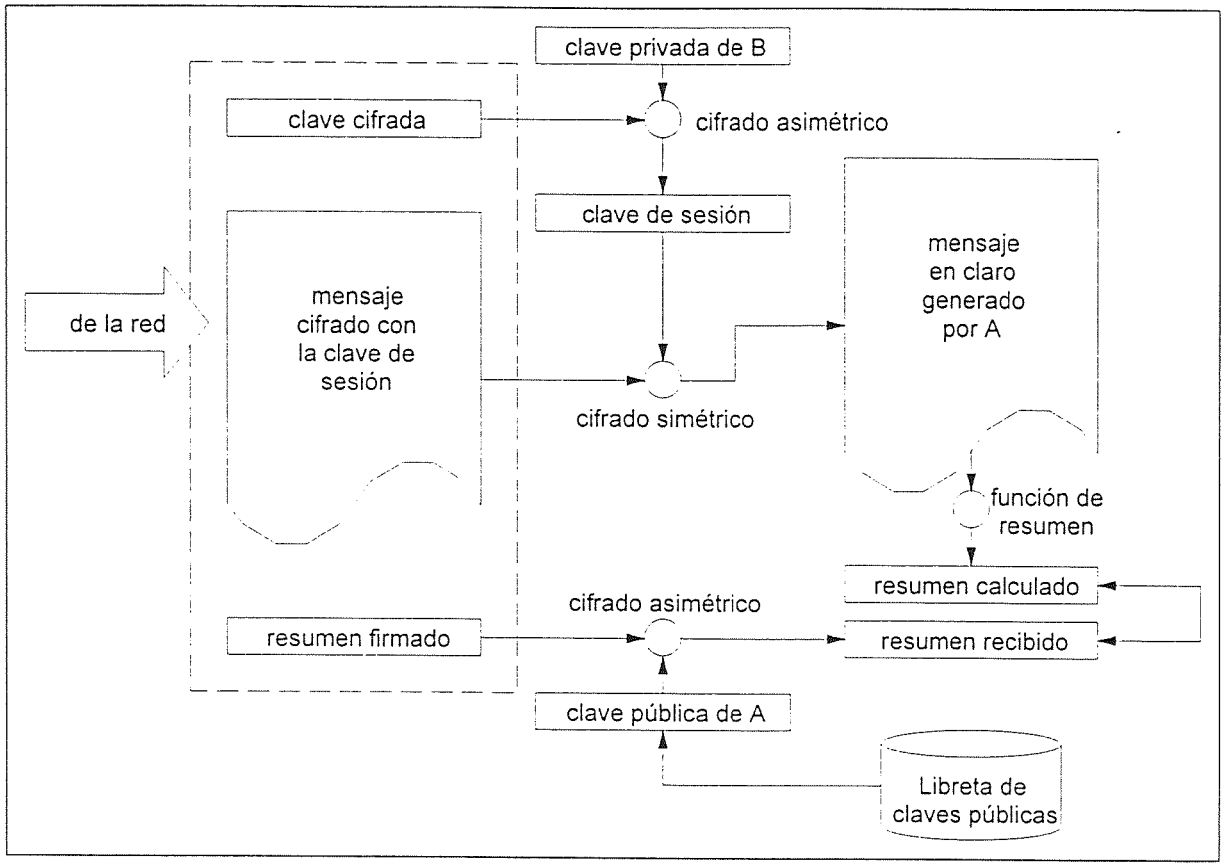


Figura 4. Diagrama de bloques del procedimiento de recepción de mensajes cifrados y/o firmados.

mensaje será cifrado con una clave de sesión elegida aleatoriamente y acompañado de una firma digital.

Firma digital del documento

En el menú "Herramientas" de la ventana "Mensaje nuevo" (Fig. 2) se añade también la opción "Firmar mensaje", representada asimismo en la barra de herramientas por el botón . Cuando se presiona este botón o se elige la opción en el menú, el mensaje se acompañará de una firma digital, aunque no será cifrado.

A continuación se describen paso a paso las acciones emprendidas por la extensión añadida al cliente *Exchange* para incorporar al sistema los mecanismos de seguridad descritos en la sección anterior.

4.1 Envío de mensajes cifrados y/o firmados

Cuando el usuario selecciona la opción de cifrar el mensaje o firmarlo, tiene lugar el siguiente proceso, representado en la Fig. 3:

- Se lee la dirección de e-mail del destinatario.
- Se lee la clave pública del destinatario en la libreta de claves públicas.

a) Si sólo se desea firmar el mensaje:

- Se carga en memoria el cuerpo del mensaje.
- Se calcula el resumen criptográfico cifrando con clave conocida 0000... en modo CBC.
- Se cifran los últimos 256 bits anteriores con la clave privada del emisor, obteniéndose así la firma digital del documento.
- Se codifica según el estándar definido en RFC1421 [7], para poder transmitir a Internet.
- Se copia la firma digital al final del mensaje
- Se asigna a la propiedad PR_SECURITY el valor "firmado" (SECURITY_SIGNED).
- Se asigna a la propiedad PR_CONTENT_LENGTH el valor de la longitud del cuerpo del mensaje sin la firma.
- La propiedad PR_KEYWORD se deja sin rellenar, puesto que no hay clave de sesión.

b) Si se desea cifrar el mensaje en vez de firmarlo solamente:

- Se genera una clave de sesión aleatoria (utilizando la hora del reloj del sistema como semilla aleatoria).
- Se cifra con la clave pública del destinatario.
- Se carga en memoria el cuerpo del mensaje.
- Se calcula el resumen criptográfico cifrando con clave 0000... en modo CBC.
- Se cifran los últimos 256 bits anteriores con la clave privada del emisor, obteniéndose así la firma digital del documento.
- Se cifra el cuerpo con la clave de sesión aleatoria.
- Se añade la firma digital al final del cuerpo cifrado.
- Se codifica todo con el estándar RFC1421.

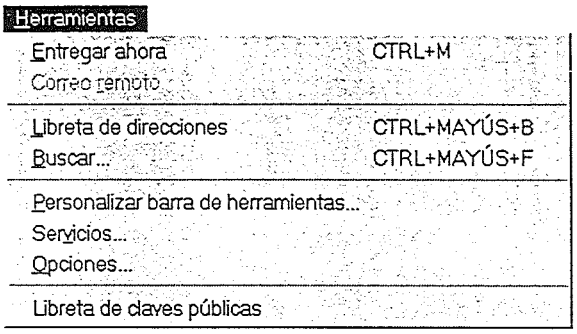


Figura 5. Menú de herramientas extendido de la ventana principal de Microsoft Exchange.

- Se asigna a la propiedad PR_SECURITY el valor "cifrado" (SECURITY_ENCRYPTED).
- Se le asigna a la propiedad PR_CONTENT_LENGTH el valor de la longitud del cuerpo del mensaje sin la firma.
- Se asigna a la propiedad PR_KEYWORD el valor de la clave de sesión cifrada con la clave pública del destinatario.

4.2 Recepción de mensajes cifrados y/o cifrados

Cuando el usuario recibe un mensaje que está cifrado o ha sido firmado, tiene lugar de forma totalmente transparente el siguiente proceso, representado en la Fig. 4:

- Se lee el objeto que se ha recibido.
- Se leen los valores de las siguientes propiedades: longitud del cuerpo del mensaje; el tipo de procesamiento de seguridad: firma digital o cifrado; la clave de sesión cifrada y la dirección de correo del remitente.

a) Si el mensaje recibido está cifrado:

- Se descifra la clave de sesión usando la propia clave privada.
- Se carga en memoria el cuerpo del mensaje recibido.
- Se decodifica con el estándar RFC1421.
- Se extrae la firma digital.
- Se descifra el cuerpo del mensaje empleando la clave de sesión recuperada y se copia en el cuerpo del mensaje.
- Se calcula el resumen del mensaje en claro cifrando con clave 0000... en modo CBC.
- Se extrae el resumen de la firma digital y se comprueba si coincide con el resumen calculado.

b) Si el mensaje recibido sólo ha sido firmado:

- Se carga en memoria el cuerpo del mensaje recibido.
- Se decodifica con el estándar RFC1421.
- Se extrae la firma digital.
- Se calcula el resumen del mensaje en claro cifrando con clave 0000... en modo CBC.
- Se extrae el resumen de la firma digital y se comprueba si coincide con el resumen calculado.

4.3 Libreta de claves públicas

En el menú "Herramientas" de la ventana principal se añade el comando "Libreta de claves públicas" (Fig. 5), representada en la barra de herramientas por el botón . Cuando se presiona este botón o se elige el comando en el menú, aparece la ventana "Libreta de claves públicas" (Fig. 6), que permite añadir nuevas claves públicas (asociadas a direcciones de e-mail), modificar las ya existentes o borrarlas.

5. Descripción de los algoritmos utilizados

Todos los algoritmos utilizados en el sistema seguro de correo electrónico han sido diseñados en este laboratorio a lo largo de los últimos años.

5.1 Cifrado simétrico

Para el cifrado simétrico se utiliza el algoritmo de cifrado en bloque *Akelarre* [1]. Se trata de un cifrador en bloque de clave secreta de gran flexibilidad en su nivel de seguridad, permitiendo la modificación vía software de parámetros tales como el número de vueltas de encriptación, el tamaño de palabra y la longitud de la clave. Su fuerza criptográfica reside en el uso intensivo de rotaciones dependientes de los datos y en la mezcla de operaciones aritméticas de grupos algebraicos diferentes. Los algoritmos de cifrado y descifrado son idénticos y de fácil implementación tanto en hardware como en software.

Para esta implementación se ha escogido una palabra de 32 bits de longitud, 512 bits de clave y 2 vueltas de encriptación.

5.2 Cifrado asimétrico

El método de cifrado asimétrico se basa en una mejora del generador de residuos cuadráticos inventado por Blum, Blum y Shub, que denotaremos por BBS*. Ellos demostraron en [8] que las secuencias generadas por la función x^2 mod n son criptográficamente seguras y las propusieron para ser usadas en la criptografía de clave pública. Sin embargo, en la práctica surgen tres problemas: 1) encontrar qué módulos $n = p \cdot q$, producen las órbitas

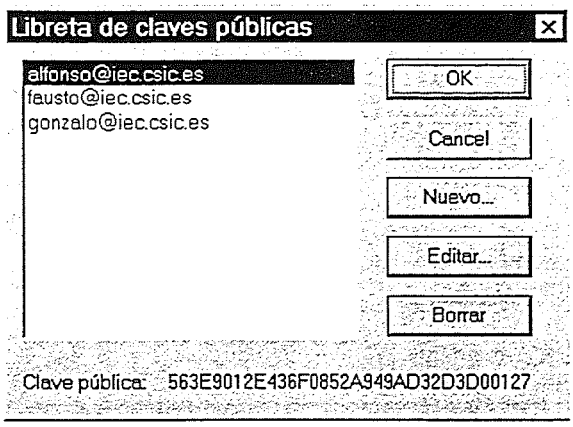


Figura 6. Libreta de claves públicas.

de período máximo, 2) determinar con qué semillas se alcanzan los períodos máximos, y 3) calcular de forma rápida y eficiente la semilla de la órbita durante la operación de descifrado, cuando se utiliza el generador para implementar un criptosistema de clave pública. La solución de estos problemas, así como el uso práctico del generador BBS* para clave pública y firmas digitales, se detalla en [2].

5.3 Función de resumen

Una función de resumen es una función, matemática o de otro tipo, que toma una cadena de entrada de longitud variable (llamada pre-imagen) y la transforma en una cadena de salida de longitud fija (más corta que la entrada).

Una función unidireccional de resumen es una función de resumen que funciona en una sola dirección: a partir de la pre-imagen resulta sencillo calcular el resumen, mientras que por el contrario resulta muy difícil generar una pre-imagen que se resuma en un valor dado. Además una buena función de resumen debe estar libre de colisiones, es decir, será muy difícil generar dos pre-imágenes con el mismo resumen.

Las funciones de resumen son públicas, ya que su seguridad descansa en su unidireccionalidad. Si están bien diseñadas, el cambio de un solo bit en la pre-imagen cambia, en promedio, la mitad de los bits del resumen.

El algoritmo de resumen utilizado consiste en el cifrado en bloque del documento mediante *Akelarre* en modo CBC, usando una clave conocida, como se describió en 3.2.1. El último bloque cifrado constituye así el resumen criptográfico de todo el documento.

6. Líneas futuras de trabajo

En próximas versiones se introducirá la posibilidad de elegir entre diferentes algoritmos de cifrado, tanto de clave secreta, para el cifrado del grueso de la información, como de clave pública, para el cifrado de claves de sesión y para algoritmos de firma digital. De esta forma el usuario podrá elegir libremente el algoritmo criptográfico al que confía su seguridad.

Queda aún por determinar un asunto primordial: el procedimiento para la generación y distribución de claves públicas, así como la elección de los mecanismos de certificación. En principio, se barajan dos posibilidades: introducir en el sistema un algoritmo de generación de parejas de claves públicas y permitir que los usuarios se certifiquen entre ellos, según un esquema de confianza mutua; o bien, delegar la responsabilidad en una autoridad central de certificación. Estas cuestiones son actualmente objeto de estudio para perfeccionar el sistema seguro de correo en el futuro.

7. Conclusiones

Se ha presentado un sistema de correo seguro basado en *Microsoft Exchange*, con la capacidad de cifrar y firmar documentos, que ofrece los servicios de confidencialidad, autenticación, integridad y no repudio.

En esta línea, este sistema representa una mejora considerable de las capacidades de la mensajería electrónica, al aportar herramientas criptográficas que aseguran la privacidad de los mensajes intercambiados, impulsando así el uso generalizado del servicio de correo electrónico.

Agradecimientos

Esta investigación ha sido financiada por la Comunidad de Madrid, Beca de Formación de Personal Investigador, y por la CICYT, TIC95-0080, proyecto "Servicio criptográfico de protección de datos para red digital de servicios integrados RDSI".

Referencias

[1] Álvarez Marañón, G., Guía Martínez, D., Montoya Vitini, F. y Peinado Domínguez, A., "Akelarre: Nuevo Algoritmo de Cifrado en Bloque", *Actas de la IV Reunión Española sobre CRIPTOLOGÍA*, 93-100 (1996).

[2] Hernández Encinas, L., Muñoz Masqué, J., Montoya Vitini, F., Álvarez Marañón, G. y Peinado Domínguez, A., "Algoritmo de Cifrado con Clave Pública mediante una Función Cuadrática en el Grupo de los Enteros módulo n". *Actas de la IV Reunión Española sobre CRIPTOLOGÍA*, 101-108 (1996).

[3] *Microsoft* Corporation, *Microsoft Exchange Client's Extensions Programmer's Reference*, *Microsoft*® Win32® Software Development Kit (SDK), (1996).

[4] Ben Goetter, *Developing Applications for Microsoft Exchange with C++*, *Microsoft* Press, (1996).

[5] *Microsoft* Corporation, *Messaging Application Programming Interface (MAPI) Programmer's Reference*, *Microsoft*® Win32® Software Development Kit (SDK), (1996).

[6] *Microsoft* Corporation, *OLE Programmer's Reference*, *Microsoft*® Win32® Software Development Kit (SDK), (1994-1996).

[7] J. Linn, "Privacy Enhancement for Internet Electronic Mail", *RFC1421*, (1993).

[8] L. Blum, M. Blum y M. Shub, "A simple unpredictable pseudo-random number generator", *SIAM J. Comput.*, **15**, 364-383 (1986).

Telegestión Segura de Cuentas de Proyectos de I+D Mediante el Uso de la Tarjeta Inteligente del Personal Investigador

José Luis Zoreda Bartolomé (1), Justo A. Carracedo Gallardo (2), Angel Redondo F.-Rebollos (1), David Cerezo Quesada(1).
(1) Grupo Universitario de Tarjeta Inteligente. Dpto. Tecnología Fotónica. E.T.S.I.Telecomunicación. U.P.M.
(2) Dpto. de Ingeniería y Arquitecturas Telemáticas (DIATEL). E.U.I.Telecomunicación. U.P.M.

Abstract:

The main objective of this project is to develop a Security Application based on Smart Cards. With this application, a user (Researcher of the OTRI) owning a smart card will be able to access from his personal computer to a remote service for consulting and transferring funds in a secure way through open networks.
Technologycally, the secure mechanims are based on the capability of the smart cards in store keys, manage the public and privated keys involved in this project and encrypt/decrypt data using simmetrics and asimmetrics algorithms.

1 Introducción

El objetivo último de este proyecto es la definición y el desarrollo de un servicio telemático (teleservicio), mediante el cual los investigadores puedan gestionar sus cuentas de proyectos de Investigación y Desarrollo (I+D) en el entorno de las Oficinas de Transferencia de los Resultados de Investigación (OTRI) de cualquier universidad. Para ello, se dotará a cada investigador de una tarjeta inteligente, mediante la cual podrá acceder, de forma segura y personal, a los servicios telemáticos, incluidos los de telegestión de cuentas, que la correspondiente OTRI proporcione, a través de la Red Iris (Internet España). Para proporcionar este servicio, en el proyecto, se desarrollará un nuevo modelo de gestión, en colaboración con el personal de la Oficina de Transferencia Tecnológica (OTT), que es la denominación de la OTRI de la Universidad Politécnica de Madrid (UPM), teniendo en cuenta la incorporación de la tarjeta como un elemento más del correspondiente sistema de información, y como parte del sistema que proporciona y gestiona los necesarios mecanismos de seguridad. En este modelo se tendrán en cuenta todos aquellos procesos susceptibles de ser automatizados, tanto de forma local como remota, mediante el uso de la tarjeta.

Para la materialización del modelo de telegestión de cuentas, en objetivos específicos, se desarrollarán los correspondientes elementos tecnológicos. La mayoría de estos desarrollos se obtendrán en base a la integración de productos que existen en la actualidad, tanto comerciales como desarrollados previamente por los grupos participantes. También, se incorporarán nuevos desarrollos, realizados a lo largo del proyecto, que cubrirán necesidades específicas.

Además de los objetivos específicos del proyecto, tanto los tecnológicos como el propio teleservicio, se pretende que el empleo de la Tarjeta Inteligente del Personal Investigador (TIPI) dentro del entorno de las OTRIs sea el primer paso para la progresiva introducción de un modelo de tarjeta inteligente universitaria, que tenga en cuenta el marco de competencias y obligaciones incluidas en la Ley de Reforma Universitaria (LRU).

2 Servicios de Telegestión de Cuentas

Entre las Entidades Promotoras y Observadoras (EPOs) de este proyecto, se cuenta con la OTT de la UPM. Su participación es esencial, ya que su personal actúa como asesores para la definición de un modelo realista de telegestión de cuentas de investigación. En el desarrollo del modelo, entre otros aspectos, se tendrá en cuenta el correspondiente marco funcional y legislativo, con el objeto de seleccionar, entre los actuales, aquellos procedimientos administrativos que sean susceptibles de realizarse de forma automática (en modo local o remoto) y, al menos, con las mismas garantías. Es decir, para la definición de los servicios telemáticos, se tendrá en cuenta si se necesita o no la firma del investigador y/o de la OTT, para que tenga valor el procedimiento administrativo equivalente (Figura 1).

Además, a lo largo de todo el proyecto, el personal de la OTT evaluará que los resultados desde el doble punto de vista de la entidad que presta los servicios como de uno de los usuarios del sistema (servicios telemáticos).

Dentro de los servicios que con carácter inmediato podrían ofrecerse a los investigadores a través de la

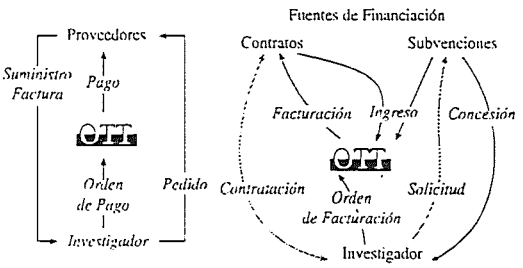


Figura 1. Flujos de Información

infraestructura de RedIRIS existente, este proyecto contempla los siguientes:

* La consulta del estado de cuentas de los proyectos gestionados por el investigador.
* La transferencia de fondos entre cuentas del investigador o investigadores, previamente autorizadas por la OTT.
* Ordenes de pago a suministradores, peticiones de cantidades a justificar (p.e. viajes), etc.
* Otros procedimientos administrativos, en los que se necesita la orden y/o autorización del investigador.

3 Modelo de la Tarjeta Inteligente del Personal Investigador (TIPI)

El modelo de la TIPI es un caso particular de un modelo de tarjeta inteligente universitaria, algunos de cuyos elementos han sido previamente desarrollados por los investigadores de este proyecto. Tal como hemos indicado, en nuestro modelo se han tenido en cuenta diversos aspectos de la LRU, que delimitan aspectos funcionales y tecnológicos de la tarjeta. Por ejemplo, hemos considerado que:

* Las Universidades son organizaciones jerarquizadas, con una clara delimitación del marco de responsabilidades y competencias.
* Para cumplir sus objetivos de formación e investigación, las Universidades deben prestar múltiples tipos de servicios.
* Existe una diversidad de miembros de la comunidad universitaria (docentes, investigadores, PAS, becarios, alumnos, ...).
* El carné universitario es un documento oficial, que identifica a cada miembro de la comunidad universitaria.

En base a estas y otras consideraciones, en nuestro modelo de tarjeta inteligente universitaria hemos considerado que:

* La tarjeta inteligente universitaria debe ser un documento oficial de identificación y llave de acceso a los distintos servicios, que presta la propia Universidad.
* Existen distintos tipos de tarjetas en función del tipo de usuario.
* La Universidad debe ser el emisor principal de la tarjeta, permitiendo que existan múltiples emisores secundarios (Escuelas o Facultades, Departamentos, OTRIs, etc.).
* Las tarjetas deben tener capacidad de proporcionar multiplicidad de aplicaciones (tarjetas multiaplicación), o diversos tipos de tarjetas en función del uso (aplicación).

La Tarjeta Inteligente del Personal Investigador (TIPI), en nuestro modelo, servirá como llave de acceso, soporte de identificación y/o elemento de certificación de todos aquellos procesos que realiza el investigador con su OTRI y son susceptibles de ser realizados a través de las redes de comunicación.

4 Objetivos Tecnológicos

Desde el punto de vista tecnológico, el desarrollo de este proyecto está estructurado en tres grandes bloques: Entorno Seguro del Investigador, Entorno Seguro del Gestor de la Aplicación Servidora y Autoridad de Certificación (figura 2).

Para llevar a cabo la comunicación en el sistema propuesto se están definiendo e implementando protocolos seguros entre las diversas entidades involucradas.

El transporte de la información se realizará a través de los servicios ofrecidos por RedIRIS, a los que se tiene acceso tanto desde las dependencias de la OTT como desde los centros que participan en el proyecto, tanto en su desarrollo como en su evaluación.

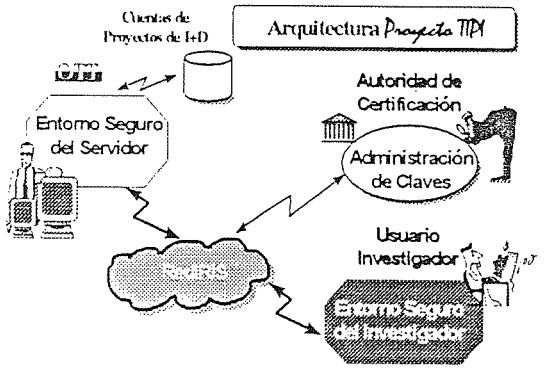


Figura 2. Arquitectura del Proyecto TIPI

4.1 Entorno Seguro del Investigador

Este entorno lo hemos definido bajo la siguiente óptica: cualquier investigador puede acceder desde cualquier terminal, del sistema, a los servicios telemáticos que proporciona la OTT. Por ello, en el entorno se distinguen tres elementos (entidades):

Tarjeta Inteligente del Personal Investigador (TIPI)

Como soporte técnico de la TIPI se utilizarán tarjetas inteligentes con criptoprocesador de clave pública. Estas tarjetas, de última generación, permiten que funcionalmente la TIPI sea un elemento activo de los mecanismos de seguridad, simplificando enormemente los otros elementos tecnológicos (tanto hardware como software) que se deban incorporar al terminal. Además, la tarjeta servirá como soporte físico de identificación. Las tarjetas que se están utilizando en el proyecto son el modelo CryptoFlex de Schlumberger, que las proporciona como EPO del proyecto.

Unidad de Lectura Escritura (ULE) de tarjetas.

Para soportar la funcionalidad de la tarjeta, en este proyecto, se están adaptado e integrando paulatinamente distintos tipos de ULEs (Unidad de Lectura Escritura), algunas de ellas comerciales y otras desarrolladas previamente por el Grupo Universitario de Tarjeta Inteligente (GUTI). En una primera fase se le adaptará e integrará una ULE a un terminal, a través de un puerto serie. De esta forma la tarjeta podrá servir como llave de acceso al mismo, medio de identificación del usuario y certificación de transacciones.

Entre los tipos de ULEs comerciales se están utilizando algunos los modelos, por ejemplo la Reflex 60 fabricada por Schlumberger.

Además de los terminales basados en ordenador, en el proyecto se pretende utilizar otros tipos de terminal. Para ello se utilizará el terminal telefónico PT100, de Philips Communications & Processing Services (EPO del proyecto), como un modelo de terminal del investigador.

Aplicación Cliente

Esta entidad local es la encargada del intercambio de información y diálogo con los otros elementos del sistema. Concretamente, ofrece un interfaz cómodo al usuario para operar remotamente sobre su cuenta. Por un lado, dialoga con la tarjeta inteligente, para verificar que el poseedor es su propietario legítimo y en base a esta identificación permitir el acceso a sus datos internos, tanto para almacenar como para leer, credenciales de usuario. También, dialoga con la Autoridad de Certificación para obtener las credenciales necesarias para presentarlas ante la Aplicación Servidora. Por último, es la encargada de dialogar con la Aplicación Servidora para que en función de una correcta identificación permita la manipulación de la cuenta del investigador.

En el proyecto se pretende llegar a un compromiso entre la cantidad de software a instalar en el terminal con la velocidad de realización de las transacciones. Además, se pretende que los desarrollos puedan utilizarse en un amplio tipo de terminales. Por estas razones se está utilizando el lenguaje "Java" y lenguajes nativos como soporte de la aplicación cliente. Esta posibilidad reduciría la necesidad de instalación de software y evitaría el desarrollo de distintas versiones de la misma aplicación, en función del tipo de terminal. El desarrollo de los elementos de la aplicación lo está realizando investigadores del GUTI. Algunos de estos elementos son nuevas versiones, y otros son resultado de la integración de desarrollos previos.

4.2 Entorno Seguro del Gestor de la Aplicación Servidora

Para este entorno se definirá una tarjeta específica, la Tarjeta del Profesional de la OTRI (POTRI). Esta tarjeta tiene análogas funciones que la TIPI, entre ellas destacamos su uso como sistema de control de acceso al Servidor.

Servidor de Atención al Investigador (SAI)

Esta entidad es la encargada de gestionar las distintas aplicaciones que se pongan a disposición del investigador para facilitarle el intercambio de información con la OTT (figura 3). Este Servidor, físicamente situado en las dependencias de la OTT, está controlado por el gestor autorizado del SAI a través de una tarjeta inteligente. Deberá realizar las siguientes funciones:

* Dialogar con la tarjeta inteligente, para verificar que el poseedor es su propietario legítimo y en base a esta identificación permitir el acceso a la operación del SAI.
* Dialogar con la Autoridad de Certificación para verificar las credenciales presentadas por el usuario para la obtención de un servicio.
* Dialogar con la Aplicación Cliente para que en función de una correcta identificación permita la realización de la operación solicitada.

El acceso al servidor se realiza a través de un navegador WEB, con el objeto de simplificar la aplicación cliente cara al usuario.

En este servidor se situará una versión reducida del sistema de bases de datos de la OTT, cuyos datos son de carácter experimental, con el objeto de evaluar el sistema de telegestión.

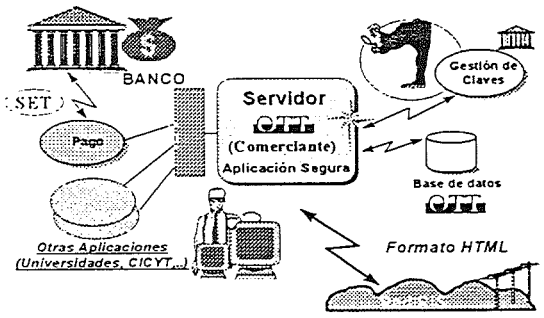


Figura 3. La OTT es el Servidor Seguro de la Aplicación

4.3 Servidor de Seguridad (Autoridad de Certificación)

En este proyecto se proporcionarán estos servicios de seguridad mediante el empleo de algoritmos asimétricos. Estos algoritmos se basan en la existencia, para cada usuario, de dos claves, una pública (y, por tanto, disponible para el resto de los usuarios) y otra privada, en este caso se hace necesario que *alguien* de confianza ratifique que la clave pública empleada por un usuario para la obtención de un determinado servicio es la clave vigente. Esta entidad recibe el nombre de *Autoridad de Certificación* (CA).

La Autoridad de Certificación es la entidad encargada de garantizar, en todo momento, la validez de las claves públicas manejadas, emitiendo y verificando las credenciales necesarias para la utilización del *Servidor de Atención al Investigador*.

DIATEL dispone de un desarrollo previo de un Servidor de Seguridad (SecServer), con funcionalidad de Autoridad de Certificación, realizado como parte de las tareas encomendadas a DIATEL dentro del proyecto EDISE (acción PASO 1993-1995), que serviría como punto de partida para la instalación de la CA.

Actualmente, este SecServer recibe las peticiones de servicio a través de correo seguro (PEM), codificadas en el cuerpo del mensaje según un formato preestablecido. Asimismo, una vez ejecutada la petición, el SecServer devuelve la respuesta mediante mensajes PEM. Proporciona las siguientes facilidades a sus usuarios: registro automático de los usuarios que desean usar los servicios del SecServer, certificación de la clave pública de usuario, información de claves públicas de usuarios y CA, baja de usuarios, información de la Lista de Certificados Revocados.

En este proyecto se propone actualizar este SecServer tanto desde el punto de vista de la aplicación cliente, como de la aplicación servidora del investigador (SAI), con el objeto de incorporar nuevos tecnológicos y funcionalidades, necesarios en el marco de nuestro proyecto. Cocretamnete, se contempla el desarrollo e implementación de nuevo protocolo seguro de acceso al Servidor de Seguridad, basado en una estructura cliente/servidor, así como adaptar la funcionalidad del mismo a las particularidades del usuario final.

Otra de las funcionalidades, de la Autoridad Certificadora, es la personalización, emisión y mantenimiento de las tarjetas. Para ello se desarrollará el correspondiente sistema, en base a desarrollos previos del GUTI, compuesto por el correspondiente software que gestionará una modelo especifico de máquina de personalización, porporcionado por SAETIC, como EPO del proyecto.

5 Conclusión

En estas líneas se han presentado los aspectos más relevantes del Proyecto TIPI, fundamentalmente aquellos aspectos que van a suponer dos novedades importantes, relacionadas entre sí:

* La utilización de la tarjeta inteligente como sistema integral de soporte y gestión de la seguridad en las transacciones a través de redes abiertas, lo que permite la telegestión segura.

* Una nueva forma de gestionar las cuentas de los proyectos de I+D en las Universidades.

Por otro lado, es importante destacar una serie de entidades y empresas que están participando en el proyecto (figura 4), bajo la figura de lo que dentro del Plan Nacional de I+D se denominan EPOs (Entidades Promotoras Observadoras).

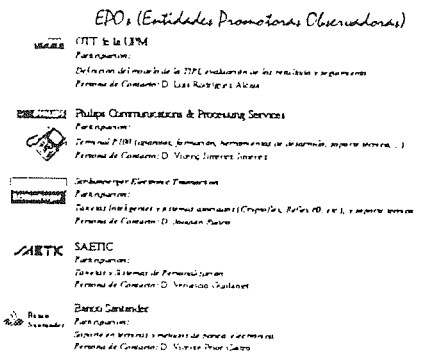


Figura 4. EPOs y Entidades Colaboradoras

Agradecimientos
Proyecto financiado por el Plan Nacional de I+D. Comisión Interministerial de Ciencia y Tecnología (CICYT). Referencia TEL96-1322.

Mecanismos de Seguridad en Internet mediante Tarjetas Inteligentes

José Luis Zoreda, Angel Redondo, David Cerezo, Jaime de Pereda, Raúl Sánchez.
Grupo Universitario de Tarjeta Inteligente (GUTI). Dpto. Tecnología Fotónica. E.T.S.I.Telecomunicación. U.P.M.

Abstract:

In this document we analize the need for a new security system on the Internet. This system will use a technology created to provide security, Smart Cards. They provide more portability and usability to existing security models. Smart Cards are good managing passwords and even making complex cryptographic algorithm. A security model for Internet transactions must provide privacy, authentication, integrity and nonrepudiation. The cryptography has the solution to this requirements but must be helped by Smart Cards as we introduced at the end of this paper.

1. Introducción

Nos encontramos en unos años de cambio en los medios de comunicación de masas. La red, Internet, se ha convertido en el modo más rápido, cómodo y barato de comunicación entre cualquier parte del mundo. Su uso está extendiéndose a un ritmo imparable, superando ya su número de usuarios los 45 millones.

Esta rápida expansión de la utilización de Internet se debe sobre todo a la proliferación de servidores http (*HiperText Transfer Protocol*). Este servicio, debido a su facilidad de uso y a lo atractivo que hace la presentación de la información, a propiciado el que esta tecnología alcance ambientes menos técnicos.

El desarrollo histórico de Internet, tal y como ahora la conocemos, partió de pequeños grupos de ordenadores interconectados entre sí y situados en diversas Universidades en los Estados Unidos. Poco a poco fueron conectándose otras Universidades y organismos oficiales, y más tarde empresas comerciales que veían en este medio una forma de presentarse al mundo.

El modo en que ha crecido la red ha sido bastante anárquico, sin ningún plan preestablecido. Esto ha motivado que, al haber alcanzado el tamaño mundial que hoy mantiene, hayan surgido numerosos problemas de prestaciones y eficiencia en las comunicaciones y los servicios.

Uno de los problemas con que nos encontramos en Internet es la baja velocidad de las comunicaciones. La inclusión día a día de nuevos servicios cada vez más sofisticados y con mayores necesidades de ancho de banda impone desesperantes esperas a los usuarios que intentan beneficiarse de estas nuevas posibilidades.

Pero el problema más importante con que se está topando Internet para llegar a ser el medio de comunicación por excelencia es la seguridad. Para los servicios utilizados hasta ahora los sistemas de seguridad que existían eran medianamente eficaces, pero las empresas comerciales quieren explotar el filón que supone un posible mercado de millones de personas. Para poder implantar estos nuevos servicios habrá que desarrollar nuevos sistemas de seguridad que mejoren las actuales e insuficientes medidas.

2. Condiciones de Seguridad

Para que una transacción realizada entre dos usuarios pueda ser considerada segura debe de cumplir las siguientes condiciones:

Privacidad de los datos que se envíen, evitando que un tercero pudiera, en caso de interferir la comunicación, entender lo que en ésta se transmite. Es la forma de protección más básica y más antigua en la historia de la criptografía y, hoy en día, es necesario completarla con otras para conseguir un nivel de seguridad adecuado.

Autentificación de los participantes en la comunicación, de forma que tanto el emisor como el receptor de la transacción sean quienes dice ser. Este tipo de comunicación que no establece un contacto directo entre los participantes permite el ataque externo en forma de suplantación de la identidad de uno de ellos.

Integridad de los datos que se transmiten. El paquete de datos puede ser atrapado en uno de los nodos por los que pasa, modificado y vuelto a transmitir para beneficio de un tercero. Tratándose de datos de una transacción podría pensarse en cambiar el beneficiario de un cobro o el importe recibido.

No repudio de la transacción por ninguna de las partes. Tiene que existir un mecanismo que proporcione la certeza de que ninguna de las partes va a echarse atrás en una transacción pasado un punto de aceptación mutua.

3. Mecanismos de Seguridad

El medio más común de proporcionar seguridad informática es el uso de algoritmos criptográficos. La criptografía ha evolucionado paralelamente a los ordenadores aprovechándose de los avances en programación y, sobre todo, en potencia de cálculo. Existen numerosos algoritmos que proporcionan

diferentes niveles de seguridad y protección frente a los distintos ataques posibles.

3.1 Privacidad y Autenticidad

Para conseguir asegurar la privacidad de los datos se utilizan técnicas de cifrado de datos mediante el uso de claves secretas. Estas técnicas se basan en ocultar el significado de los datos utilizando un conjunto de operaciones que requieren una clave que, posteriormente, podrá ser usada para poder entenderlos. Así, como se puede ver en la figura 1, un usuario cifraría sus datos con la clave secreta antes de enviarlos y, posteriormente, en el otro extremo se llevaría a cabo la operación inversa volviendo la información inteligible. La clave usada en el cifrado y el descifrado es la misma y por eso estos algoritmos son llamados de clave simétrica.

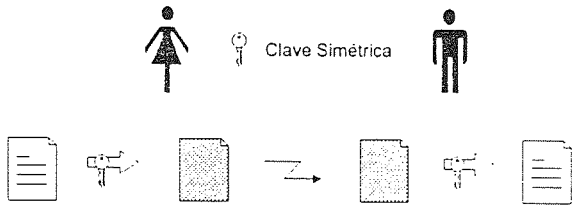


Figura 1. Cifrado con clave simétrica.

Existe otro tipo de algoritmos de cifrado que utiliza claves asimétricas, es decir dos claves diferentes cada una de las cuales realiza la operación inversa a la otra. De estas claves una permanece en poder del usuario (privada), mientras que la otra (pública) puede ser difundida entre los posibles receptores. Este sistema permite dos tipos de juego. El más obvio consiste en cifrar la información que se quiere enviar a un destinatario con su clave pública, que previamente él se ha preocupado de difundir. Esto garantiza que el único que puede leer la información es el destinatario ya es él quien posee la clave privada que la descifra (Fig. 2).

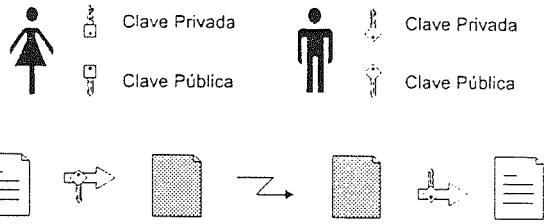


Figura 2. Cifrado con clave asimétrica.

Otra forma de usar los algoritmos de clave asimétrica es para firmar la información. Al firmar un mensaje estamos asegurando la identidad de quien lo envía. Si los datos que se van a transmitir se cifran previamente con la clave privada esto significa que únicamente podrán ser descifrados con la clave pública correspondiente, o lo que es lo mismo, si un mensaje puede ser descifrado mediante la clave pública de un usuario podemos estar seguros de que ha sido él quien lo envió cifrándolo con su clave privada (Fig. 3).

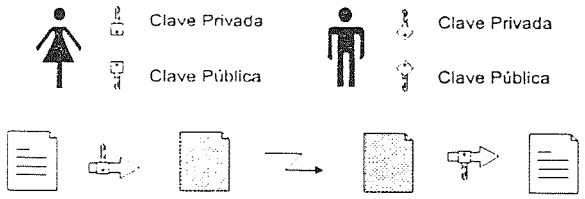


Figura 3. Firma con clave asimétrica.

3.2 Gestión de las claves

Un problema importante que poseen estos algoritmos es la gestión de las claves, tanto su difusión como su almacenamiento.

Al utilizar un sistema de cifrado basado en el secreto de las claves es importante que éstas no sean descubiertas por un usuario hostil. Tanto en algoritmos de clave simétrica como de clave asimétrica existe un peligro de seguridad en el sistema al difundir las claves, aunque por diferentes causas.

En un algoritmo de clave simétrica el secreto de la clave es esencial, por lo que es imprescindible encontrar un canal seguro por el que se haga llegar ésta al destinatario del mensaje. Debido a la fragilidad del sistema ante capturas de claves lo usual es utilizar una clave de sesión que se genera aleatoriamente en el momento de cifrar los datos y que se envía por el canal seguro al destinatario de esos datos. Este canal seguro puede conseguirse mediante otros mecanismos criptográficos como puede ser, por ejemplo, el uso de un algoritmo de claves asimétricas con el que se cifre la clave de sesión.

En un algoritmo de clave asimétrica el problema viene al difundir la clave pública. Esta clave pública identifica a un usuario que firme sus datos. Si durante la difusión de la clave pública de un usuario ésta es cambiada por otra, el destinatario, no tiene modo de saber que esta suplantación ha tenido lugar, lo que provocará que identifique erróneamente al remitente en posteriores mensajes.

Este problema se ha solucionado incluyendo otro participante en la transacción. Éste se encargará de emitir certificados que identifiquen a los usuarios del sistema. Los certificados incluyen, además de la clave pública, información personal del usuario que impida la suplantación de la clave, y todo ello va cifrado con la clave privada de la entidad certificadora. Cuando se realice una transacción, por tanto, lo que se difunde es el certificado en lugar de la clave pública y el remitente podrá obtener dicha clave sin más que descifrar el certificado con la clave pública de la entidad. Esto conduce a establecer unas jerarquías de credibilidad entre

entidades certificadoras que deberá ser cuidadosamente estructurado.

El problema del almacenamiento de las claves en sistemas de clave simétrica ya se ha visto solucionado. Basta con no almacenarlas, sino generarlas en el momento en que se necesiten. El problema, entonces, se encuentra en los sistemas de clave asimétrica en los que realizar esto requiere un coste excesivo de cálculo y el no poder tener previamente difundidas las claves públicas. La solución pasa entonces por encontrar un lugar seguro para las claves privadas. Los ordenadores pese a los sistemas de seguridad de que disponen pueden ser sometidos a múltiples y repetidos ataques que podrían llegar a atravesar sus defensas. Lo ideal sería que el propio usuario pudiera transportar las claves. Memorizarlas podría ser un ejercicio complicado ya que este tipo de claves deben tener una longitud superior a los 1024 bits. Es necesario, por tanto, utilizar un *hardware* adicional.

Una tarjeta chip parece el soporte adecuado, su transporte es sencillo y posee la electrónica necesaria para almacenar las claves. Pero, si almacenamos las claves en la tarjeta y ésta las pasa al ordenador para realizar el cifrado, pueden ser capturadas por un programa que esté a la escucha en ese momento. Una solución más segura consistiría en que fuera la tarjeta quien realizara el cifrado.

Convertir la tarjeta en un elemento activo del proceso de cifrado aumenta en gran medida la seguridad del sistema. Las nuevas tarjetas criptográficas son capaces de realizar cifrados asimétricos de pequeñas cantidades de información, pero todavía no tienen la suficiente potencia como para cifrar todo el mensaje en un tiempo razonable. De todos modos, esto no es necesario. El cifrado del mensaje se puede realizar por medio de un sistema de clave simétrica (mucho más rápido) y esta clave se transmitiría cifrada con la clave pública del destinatario (Fig. 4). Las tarjetas chip son también capaces de realizar cifrados simétricos, aunque para grandes cantidades de información aún son algo lentas debido a que el cifrado se tiene que efectuar en bloques pequeños.

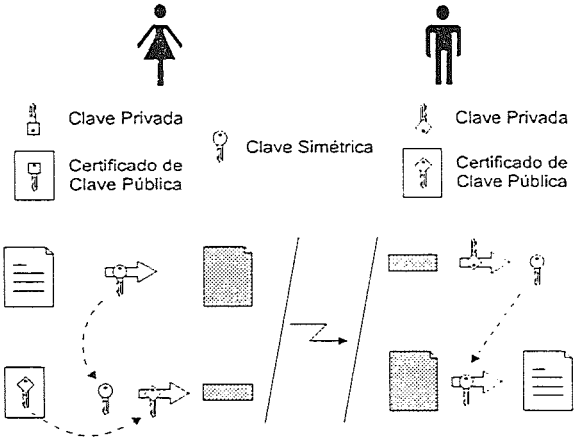


Figura 4. Cifrado asimétrico de clave simétrica.

3.3 Integridad

Otra de las condiciones que se deben cumplir en el sistema es la integridad de los datos que se transmiten. En una transacción puede no ser tan importante el que los datos sean visibles como el que puedan ser modificados y retransmitidos sin que se note.

Para poder saber si los datos se han modificado o no se recurre al uso de algoritmos que proporcionan una huella digital de dichos datos. Esta huella digital, como la huella de un dedo, caracteriza su origen pero a partir de ella es imposible reconstruirlo. La huella digital suele ser una cadena de bits de longitud fija resultado de un tipo de función llamada *hash*. Esta huella se transmite junto con el mensaje del que proviene, de forma que en el destino se puede volver a realizar la huella y compararla con la enviada. Como la función *hash* tiene solución única, si ambas huellas no coinciden los datos han sido modificados (Fig. 5).

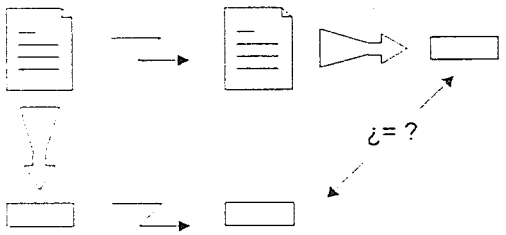


Figura 5. Huella digital.

La huella digital se puede transmitir cifrada con la clave privada, proporcionando así además, una firma del remitente. En un futuro se espera que las tarjetas también incorporen este tipo de funciones.

3.4 No repudio

Para evitar el repudio de una transacción acordada se utilizan varios métodos. Tiene que haber un tercer participante que haga de mediador y registre el acuerdo al que se ha llegado. Además se pueden utilizar sistemas de dobles firmas por los que la transacción quede firmada por las partes participantes antes de poder aprobarse.

4. La solución

Combinando todos estos procesos podemos llegar a una solución que proporcione suficiente seguridad en las transacciones. Esta solución realizará un paquete seguro de los datos que se quieren transmitir en el emisor, es decir, no hará uso intensivo de las comunicaciones sino que éstas se limitarán al momento de enviar el mensaje protegido y de recibir una confirmación.

El hecho de utilizar una tarjeta inteligente para el almacenamiento de las claves y para realizar el cifrado

de clave asimétrica de la información proporciona un nivel de seguridad superior a los sistemas actuales. Otra importante ventaja que se deriva de esto es la movilidad que proporciona. Al no estar almacenadas las claves en un determinado ordenador sino que se encuentran en un sistema portátil como es la tarjeta se podría utilizar cualquier terminal al que se tuviera acceso. Es decir, se tenderá a que el terminal soporte cada vez menos peso al proporcionar la seguridad en favor de la tarjeta. Si el servicio al que se accede es independiente de la arquitectura en la que se ejecute, el terminal podría ser cualquier tipo de plataforma (PC, Mac, Unix...), lo que aumentaría aun más la flexibilidad y difusión de los servicios ofrecidos.

Agradecimientos

Proyecto financiado por el Plan Nacional de I+D. Comisión Interministerial de Ciencia y Tecnología (CICYT). Referencia TEL96-1322.

UPM por su ayuda en la gestión de la pantente: "Sistema seguro de gestión de transacciones en redes abiertas mediante el uso de tarjetas con circuito(s) integrado(s)". Referencia UPM/PAT-9705.

SISTEMA JERARQUICO DE ADMINISTRACION DE CLAVES PUBLICAS PARA EL CORREO ELECTRONICO

Lucía Pino, Antonio Maña, Juan J. Ortega, Javier López
Dpto. Lenguajes y Ciencias de la Computación
E.T.S. Ingeniería Informática
Tfo: (95) 2131327; Fax: (95) 2131397
e-mail: jlm@lcc.uma.es

ABSTRACT

Electronic mail is one of the reasons for the continuos increment of Internet users. Not only data but resources become vulnerable once the user is connected to the rest of computers in the network. Messages travel open and available and the use of encipherment techniques is the best approach to provide security, and public-key cryptosystems are highly used in broad computer networks. In this paper we propose a new hybrid system for public keys management over Internet that avoids synchronisation problems between key servers and reduces network traffic rate.

1. INTRODUCCION

En la actualidad, cerca de 40 millones de usuarios de 159 países de todo el mundo utilizan Internet de forma habitual. Se calcula que existen 5 millones de páginas Web, 45000 redes interconectadas y 56000 dominios registrados. La tasa de crecimiento anual de Internet es del 180%, de forma que en 1997, si se cumplen las previsiones del instituto de investigación Forrester, la red mundial contará con 126 millones de personas conectadas. Estas cifras ilustran de forma clara el alcance de este fenómeno que es Internet. Todo el mundo habla de esta red de redes.

En España, de momento, nos movemos en cifras mucho más modestas. Contamos con alrededor de 30000 usuarios profesionales con conexión propia y no llegan a 150000 los que tienen una dirección reservada Internet. Hace unos años, en nuestro país, el mundo Internet sólo era moneda corriente en los ámbitos académicos y de investigación. Subidos al carro internacional de la explosión Internet, desde 1992 - y muy especialmente durante 1995 - los tentáculos de la red han ido traspasando poco a poco los muros universitarios hasta llegar al resto de la sociedad española a un ritmo de crecimiento superior al 100% anual, en línea con los del resto de Europa.

2. IMPORTANCIA DEL CORREO ELECTRONICO

Internet dobla su tamaño cada año, y el correo electrónico es una de las razones principales para ese crecimiento. Es uno de los productos con más amplio impacto en los medios de comunicación humana y es, de hecho, la aplicación más utilizada en entornos distribuidos, disponiendo los usuarios de un buzón donde acceder para la transmisión de documentos, gráficos o programas informáticos.

Dada la importancia que las empresas y los individuos están dando al correo electrónico, el crecimiento anual de buzones de correo (mailboxes) es espectacular, estando previsto que se supere la cifra de 100 millones en 1996. Varias tendencias clave han motivado este crecimiento:

1. Aumento del número de ordenadores personales.
2. Adopción generalizada de interfaces gráficos de usuarios.
3. Integración del correo electrónico con el sistema operativo y las aplicaciones.
4. Afirmación de la informática C/S.
5. Crecimiento de Internet.

La importancia del mercado de e-mail seguirá su carrera ascendente dentro del negocio global del software. Según datos de un estudio de IDC, en 1996 el mercado mundial de software de e-mail se aproximará a los 1600 millones $. Para el año 2000 se prevé que la

facturación de software total supere los 2600 millones, lo que representa un ratio de crecimiento anual del 19%.

Pero existe un grave problema ya que la mayor parte del correo electrónico es vulnerable [1]. Cuando se realiza una conexión a Internet tanto los datos (información que se guarda en los ordenadores) como los recursos (ordenadores en sí mismos) quedan expuestos al riesgo.

3. REQUISITOS, OBJETIVOS Y TECNICAS DE SEGURIDAD

Desde el punto de vista de la seguridad el mundo del correo electrónico es similar al mundo de las postales. Los mensajes viajan de máquina a máquina de forma abierta y disponible, como los mensajes escritos en la parte trasera de una postal. Con los mensajes moviéndose por la red, sus usuarios tienen la posibilidad de acceder a los mensajes. Puede que esos usuarios no hagan uso de esa posibilidad o puede que, por restricciones de acceso impuestas por los sistemas operativos de red, tengan difícil acceder a los mensajes. En cualquier caso la única seguridad existente tiene como única base la honestidad, la indiferencia o la ignorancia de los usuarios de cualquiera de los puntos intermedios por los que tiene que viajar el mensaje. Más aún, ni emisor ni receptor del mensaje tienen control sobre cuáles son esos puntos intermedios.

Adquirir cierto grado de seguridad en el envío de una postal no es difícil. Algo tan simple como meter la postal en un sobre le proporcionará protección y evitará que se pueda ver su contenido de forma fácil. Pero no existe la posibilidad de introducir un mensaje de correo electrónico dentro de un sobre; aunque sí tenemos una herramienta que puede funcionar de forma análoga, el cifrado, proporcionado por los programas de seguridad de correo electrónico.

Para ser conscientes de los muchos peligros que corre el más insignificante de los mensajes de correo electrónico hay que considerar que le pasa al mensaje desde que abandona el extremo emisor hasta que llega al extremo receptor. En primer lugar y como ya se comentó anteriormente, cuando el mensaje viaja a través de la red cualquiera de los ordenadores a lo largo del camino puede hacer una copia del mismo. En segundo lugar, cuando el mensaje llega a su destino espera hasta que el receptor lo recoge, y durante este tiempo el mensaje es vulnerable. Y por último, dependiendo de cómo lee el receptor el mensaje, éste es susceptible de ser interceptado mientras se traslada desde el servidor del sistema hasta la estación de trabajo del usuario receptor del mensaje. Es por ello que existe una creciente demanda para complementar el servicio básico de correo electrónico mediante servicios de autenticación y confidencialidad. Estos servicios se proporcionan mediante la utilización de técnicas de cifrado.

Cuando se van a utilizar las técnicas de cifrado como método de protección de la información hay que ser consciente de que en realidad se ha de adoptar una estrategia completa de seguridad para que el esfuerzo adicional de cifrar los datos no sea contrarrestado por otros problemas. Así pues, se persiguen varios objetivos:

· Privacidad, porque uno de los objetivos es mantener privados los documentos, aunque a veces esto también se puede conseguir mediante la utilización de passwords o mediante sistemas de control de acceso.

· Integridad, es decir, asegurarse de que los datos o aplicaciones no son modificados sin consentimiento.

· Autenticidad, asegurando que la persona que utiliza el sistema es realmente quien dice ser, lo cual ayuda a mantener la privacidad y la integridad en el sistema previniendo que personas no autorizadas utilicen el sistema.

· Disponibilidad, es decir, que el ordenador y los datos que contiene estén disponibles para el usuario cuando este los necesite.

Como es bien sabido, las diferentes técnicas de cifrado se pueden agrupar en dos categorías, criptosistemas de clave privada y criptosistemas de clave pública (fig. 1). En los sistemas de clave privada, dos individuos comparten una única clave secreta, y ambos utilizan esa clave tanto para cifrar como para descifrar. En los sistemas de clave pública se generan dos claves matemáticamente relacionadas para cada individuo; un mensaje

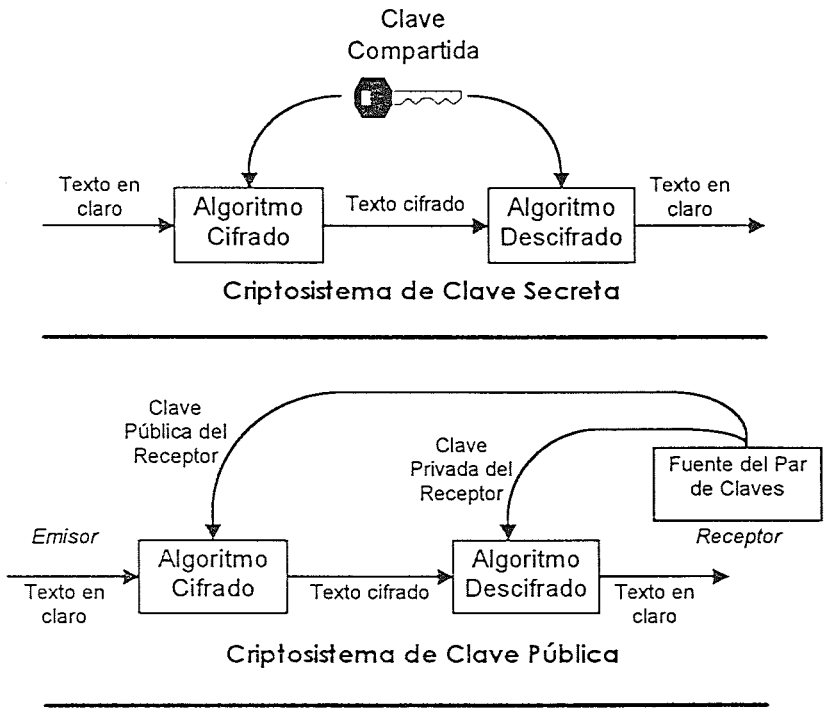


Fig 1. Técnicas Criptográficas

cifrado con la primera clave (clave pública) sólo se puede descifrar con la segunda (clave privada).

La mayor ventaja de la criptografía de clave pública respecto a la criptografía de clave privada es el incremento de seguridad; en la primera de ellas la clave privada nunca necesita ser transmitida o revelada a nadie, mientras que en la segunda siempre existe la posibilidad de que alguien descubra la clave secreta mientras es transmitida. Otra ventaja de los sistemas de clave pública es que proporcionan un método para implementar firmas digitales. Y por último, en grandes sistemas de red, los sistemas de clave pública no requieren la gran cantidad de claves que necesitan los sistemas de clave privada [2].

Los criptoanalistas, a menudo atacan tanto a los sistemas de clave pública como a los sistemas de clave privada a través de su administración de claves ya que el diseño de algoritmos y protocolos criptográficos seguros no es fácil, pero mantener secretas las claves es mucho más difícil [3]. A menos que a las claves se les de el mismo nivel de protección que a los datos, aquellas serán un punto débil. Aunque el algoritmo de cifrado sea computacionalmente imposible de romper, el sistema entero puede ser vulnerable si las claves no son adecuadamente protegidas.

Para la administración de claves son extremadamente importantes métodos seguros. En la práctica la mayoría de los ataques en los sistemas de clave pública se realizan en el nivel de administración de claves, más que en el algoritmo criptográfico en sí. Los usuarios deben ser capaces de obtener de forma segura un par de claves acorde con sus necesidades de eficiencia y seguridad. Debe existir una forma de ver las claves públicas de los demás y de publicar la propia. Los usuarios deben confiar en la legitimidad de las claves públicas de los demás; de otra forma, un intruso puede o bien cambiar claves públicas listadas en un directorio, o bien impersonar a otros usuarios.

4. ESQUEMAS DE ADMINISTRACION DE CLAVES

Se han propuesto distintas técnicas para la administración de claves públicas. Todas estas propuestas se pueden dividir en dos esquemas generales: el esquema centralizado y el esquema distribuido. Ambos esquemas presentan ciertos problemas y nuestra propuesta va encaminada a subsanar estos problemas

mediante un esquema híbrido de los dos anteriores.

En el esquema centralizado [4] aparece el concepto de Centro de Distribución de Claves (KDC) que es el receptor de las peticiones de claves públicas de los usuarios (fig 2). Este KDC suele convertirse en un cuello de botella en el sistema central ya que cada vez que un usuario desea contactar con cualquier otro, necesariamente tiene que pedir la clave pública al sistema. Además, existe una única persona como autoridad superior sobre todo el esquema de seguridad.

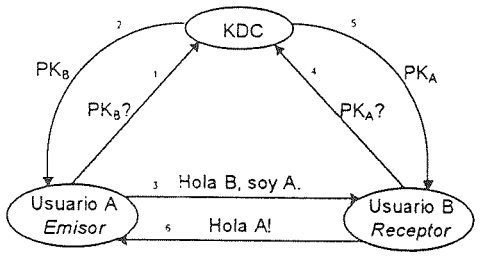


Fig 2. Sistema Centralizado de Administración de Claves

Si como medida de protección ante fisgones la tabla de claves públicas está cifrada con la clave secreta del KDC, y esta clave se ve comprometida, entonces todo el sistema de seguridad será vulnerable, quedando completamente bajo el control del oponente y pudiendo éste sustituir una clave falsa por la original, impersonando a cualquier usuario del sistema y fisgoneando en todos los mensajes que se envíen.

Dentro del esquema distribuido [5] existen dos enfoques:

- El primero (fig. 3a) contempla la existencia de servidores de claves con versiones replicadas de ficheros de claves públicas a lo largo de toda la red Internet, es decir, todos esos servidores sincronizan sus claves. Los servidores de claves en Internet son un intento de solución del problema fundamental de cómo conseguir la clave pública de alguien con quien se desea comunicar. La comentada sincronización entre servidores llega a ser un problema intratable ya que crece exponencialmente al aumentar el número de usuarios del sistema de correo electrónico. Cuando se hace uso de un servidor de claves, es importante darse cuenta de que realmente no hay forma de verificar cuándo la clave de una persona es legítima ya que los servidores, al menos en PGP, no chequean la autenticidad de las claves que almacenan. También es importante darse cuenta que la mayoría de los servidores de claves son gestionados por individuos; los servidores de claves no son ofrecidos como servicios oficialmente soportados por las universidades y empresas donde residen.

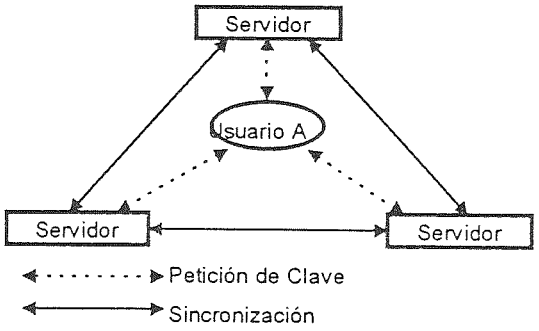


Fig 3.a Servidores de Clave Sincronizados

- El segundo (fig. 3b) no utiliza servidores de claves. En este caso la clave se solicita directamente al usuario con el que se desea realizar la comunicación. El inconveniente de este sistema distribuido es que un usuario particular no puede verificar la validez de la clave pública de todos los demás usuarios. Si se recibe la clave pública de cualquier otro usuario, no se tiene garantía de conocer a alguno de los que firman o certifican dicha clave pública (llamados introducers) y, por tanto, no se puede confiar en la validez de la misma.

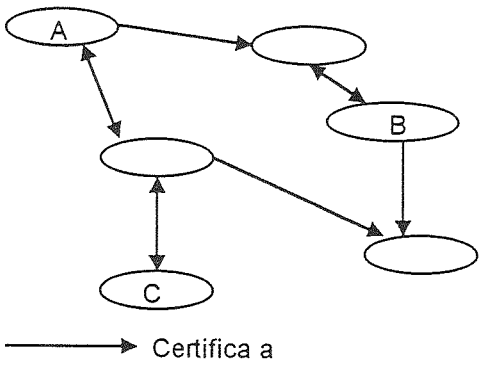


Fig 3.b Distribución sin Servidores

5. UN NUEVO SISTEMA DE SEGURIDAD EN CORREO ELECTRONICO

Una vez vistas las carencias de los esquemas centralizado y distribuido, nuestro objetivo ha sido encontrar un sistema que además de evitar esas carencias incremente la seguridad y la simplicidad de los anteriores.

La elevada tasa de crecimiento anual de usuarios de Internet y la creciente necesidad de seguridad en el correo electrónico que estos utilizan obliga a que uno de los objetivos principales de nuestro esquema sea el de la simplicidad. Hay que tener en cuenta que, de esa gran cantidad de usuarios, la mayoría no tiene un gran dominio de las diversas herramientas informáticas, entre ellas la del correo electrónico, por lo que el sistema ha de ser extremadamente simple en su manejo para que no sufra rechazo por parte de dichos usuarios. Asimismo, y debido a ese gran volumen de usuarios, es una exigencia prioritaria que el sistema propuesto no genere una alta tasa de tráfico en la red.

En cuanto a la seguridad, es absolutamente necesario que aquellos que firman las claves no sean simples usuarios sino verdaderas autoridades certificadoras y, a diferencia de los esquemas basados en la confianza de una única autoridad de certificación, nuestra propuesta recoge la existencia de un grupo de ellas que pueden funcionar de forma independiente. Además, para una mayor confianza, es necesaria la existencia de una sola copia de cada clave pública de usuario.

En nuestro esquema, cada grupo de usuarios registra sus claves públicas en una base de datos localizada en una Unidad de Servicios de Claves (KSU), donde reside su estafeta de correo electrónico. Además, por cada una de estas KSU existe una Autoridad de Certificación que es responsable del mantenimiento de las claves y de la integridad del sistema. Por lo tanto, cada clave pública sólo se almacena en la estafeta de correo más próxima al usuario definida por su dominio y no es necesaria la sincronización requerida en el esquema distribuido, por lo que el cambio de claves es muy simple.

De la misma forma cada grupo de Unidades registra sus claves en una base de datos de una KSU superior de la jerarquía (fig. 4) basada en la estructura de dominios de la red [6], cuya Autoridad de Certificación autentifica esas claves. Al contrario de los servidores de claves del esquema distribuido, los cuales no autentificaban las claves que le llegaban, en nuestro sistema las claves de cada base de datos están certificadas por la correspondiente Autoridad.

Cuando la clave pública de algún usuario, digamos A, es necesitada por algún otro usuario de Internet, digamos B, éste solicita dicha clave a su Unidad de Servicio de Claves, KSU_B , la cual se comunica con la Unidad de Servicio de Claves de A, KSU_A; ésta envía a KSU_B la clave pública de A y un timestamp de creación. Este es el caso normal (fig. 5), como se ha visto que operan los servidores de claves del PGP.

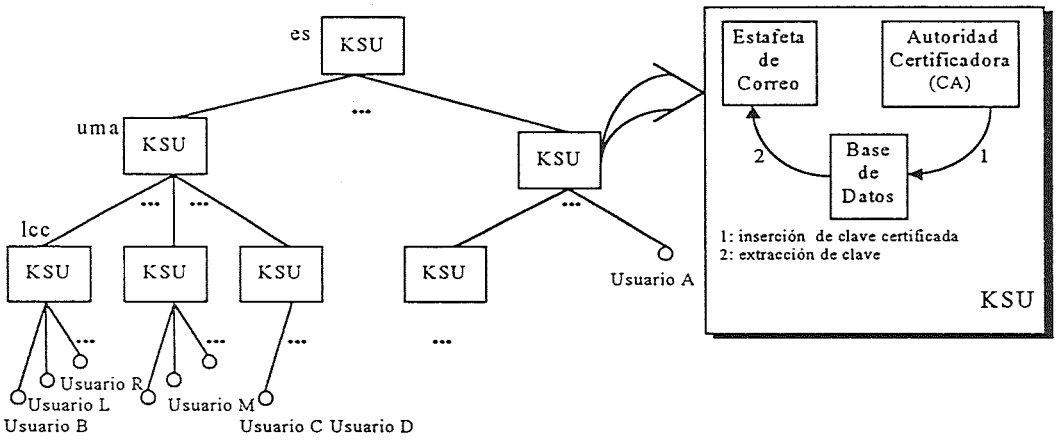


Fig 4. Jerarquía de nodos y componentes del KSU

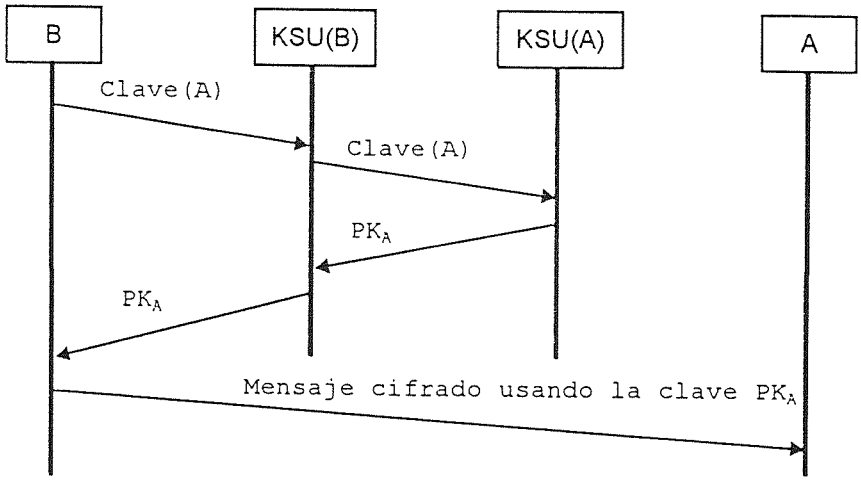
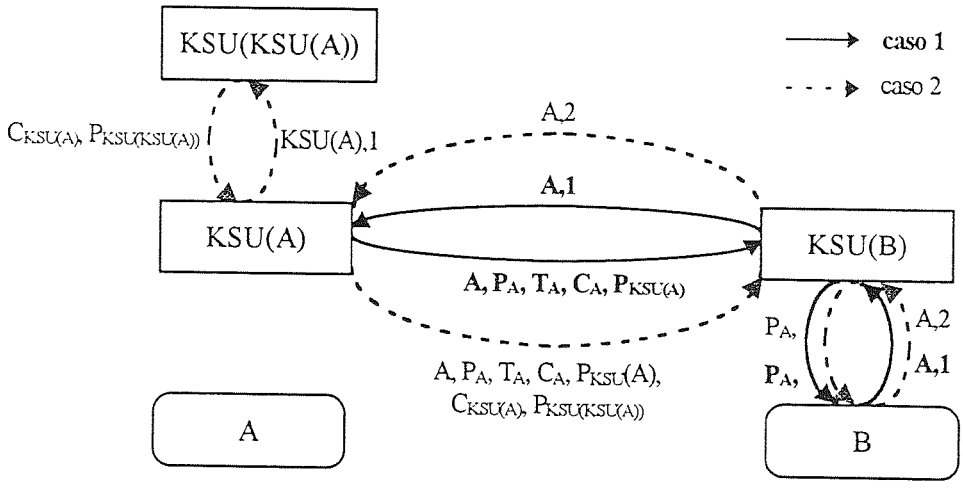


Fig 5. Petición simple de clave (sin certificación)

Pero nuestro esquema va más allá, porque considera que el usuario B puede exigir ciertas garantías de autenticación de la clave de A solicitando, no sólo esa clave, sino además la certificación de la correspondiente Autoridad de Certificación (fig. 6). En este caso KSU_A envía a KSU_B , además de la clave pública de A y su certificado, la clave pública de la Autoridad de KSU_A para verificar dicho certificado. En general, el usuario B puede elegir, en su petición, el nivel de profundidad de certificación dentro de la jerarquía de Unidades a la que el usuario A pertenece.

Obviamente, la existencia de una verdadera Autoridad de Certificación en cada una de las KSUs hace que la verificación de autenticidad de las claves sea menos necesaria por el acceso directo al KSU más próximo al usuario y por la inherente seguridad del lugar donde las claves están almacenadas, ya que es la única entidad que interactúa con la base de datos de claves públicas, pero aún así, y como se ha visto, se puede alcanzar un nivel de verificación tan profundo como se quiera.

Anteriormente se comentó que era una exigencia prioritaria que el sistema no generara una alta tasa de tráfico en la red. Una forma de conseguirlo es evitando continuas transferencias de la clave pública de un mismo usuario hacia un mismo KSU de la red. Para ello dotamos a cada KSU de un proxy con las últimas claves solicitadas durante un determinado periodo.



P_X : Clave pública de X
T_X : Timestamp de creación de P_X
C_X : Certificación de P_X
X,i : Dirección de e-mail del usuario X, Nivel de Certificación

Fig 6. Flujos de datos para uno y dos niveles de certificación

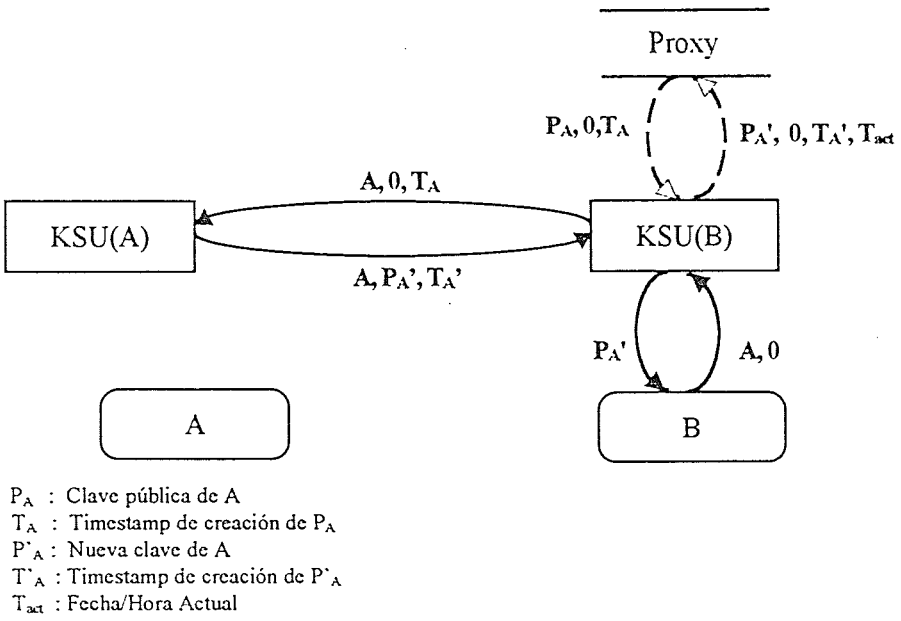


Fig 7. Uso del Proxy cuando la clave está presente pero ha expirado

El proxy almacenará grupos de: clave pública, certificado, nivel de verificación de certificación, timestamp de creación y fecha de incorporación al proxy. De forma que cuando el usuario B le pide a KSU_B la clave de un determinado usuario de Internet, ésta consulta el proxy para ver si existe esa entrada (fig. 7). Si no existe esa entrada o existe con un nivel de profundidad de verificación inferior al solicitado en ese momento por B, KSU_B realiza la petición como se explicó con anterioridad. En otro caso, en la petición se incluye el timestamp de creación para que sea verificado por KSU_A; si los timestamps coinciden, entonces KSU_A envía simplemente una confirmación y, en caso contrario, envía la nueva clave.

También se mencionó que era deseable un sistema simple en su manejo para evitar el rechazo de utilización por parte de los usuarios. Como se puede observar, la totalidad del trabajo recae directamente sobre la KSU, ya que el usuario sólo se limita a pedir la clave y el nivel de profundidad de certificación que desea, quedando descargado de cualquier otra tarea.

Asimismo, el esquema propuesto tiene la posibilidad de interactuar con otros servicios de claves. Esto es totalmente transparente al usuario porque el sistema es el responsable de conseguir las claves que le son solicitadas por sus usuarios.

Los servicios de claves más usados son:

- servidores sincronizados de claves PGP
- acceso vía finger
- mensaje de corre electrónico indicando "get key" en el campo "Subject:"

6. CONCLUSIONES Y TRABAJO FUTURO

Se ha presentado un sistema para el servicio de claves públicas, que proporciona seguridad en el correo electrónico para los usuarios de Internet. Se ha mostrado que los diseños existentes hasta el momento no son completamente seguros, y cómo un sistema híbrido puede solucionar estos problemas y mejorar el rendimiento de la red evitando una alta tasa de tráfico. Además de este hecho, el sistema es fácil de usar, transparente al usuario y compatible con la mayoría de los servicios de claves ya existentes. Este sistema está siendo estudiado para probar futuras mejoras tanto en el rendimiento como en su aceptación por parte del usuario. También se estudia cómo mejorar la interacción entre un KSU y los superiores en su jerarquía.

REFERENCIAS

[1] Schneier, B. *E-mail Security: How to Keep your Electronic Messages Private.*
John Wiley & Sons, Inc., 1995

[2] Diffie, W. *The First Ten Years of Public-Key Cryptography.*
Proceedings of the IEEE 76, 1988

[3] Fumy, S.; Landrock, P. *Principles of Key Management.*
IEEE Journal on Selected Areas in Communications, June 1993

[4] Popek, G.; Kline, C. *Encryption and Secure Computer Networks.*
ACM Computing Surveys, December 1979

[5] Garfinkel, S. *PGP: Pretty Good Privacy.*
O'Really & Associates, Inc., 1995

[6] System and Network Administration. *Administering DNS.*
SUN Microsystems

Aplicaciones e Impacto Social

Monitorizacion de Datos Usando Tecnologías Internet/Intranet

Jon Barandiaran Landin
Dpto. de Tecnologías de la Información
LABEIN
Parque Tecnológico, 101. 48170 - ZAMUDIO
Correo Electrónico: jonb@labein.es

Abstract

During the last years conectivity between computers has increased to a point that allows the development of applications that some years ago were inconceivable. The explosive growth of Internet and the browsers usability allows to every computer connected to a telephone to obtain information from everywhere in the Net. This posibility can be used by utilities to extract information from the RTUs and to present that information to anybody in the company who can make good use of it. This is more important nowadays when utilities are facing deregulation and have to decrease costs without affecting quality

1. Introducción

El aumento de la conectividad entre ordenadores ha originado que la informática distribuida haya ido abriéndose paso progresivamente en el campo de las tecnologías de la información.

Inicialmente, los ordenadores se unían dentro de las empresas por medio de redes de area local, LAN, o entre las distintas sedes de las empresas por medio de redes de banda ancha, WAN. Hoy en día es posible acceder a ordenadores en todo el mundo utilizando la red Internet.

En los últimos años se ha producido un espectacular crecimiento de la red Internet, tanto a nivel de número de usuarios como en cuanto al número de ordenadores conectados a la misma. Este espectacular crecimiento crea un sin fin de nuevas oportunidades de aprovechamiento de la accesibilidad que proporciona la red, entre las que se encuentra la monitorización de datos en tiempo real.

En la actualidad, la monitorización y el control de las redes de transporte y distribución en empresas de distribución de electricidad, gas y agua se realiza a tres niveles. Los datos, entradas digitales o analógicas, contadores, ..., son captados por las remotas situadas en campo, normalmente en nodos de la red, subestaciones en el caso de la red eléctrica y estaciones de regulación y medida en el caso de redes de agua o gas. Las remotas, que supervisan los distintos equipos están conectadas a un control integrado que agrupa las señales captadas y las envía a continuación a los despachos de control de las compañías, donde se realizan tareas de supervisión y control. El transporte se realiza usando redes propiedad de las empresas y utilizando una gran variedad de medios físicos como son: la onda portadora sobre red eléctrica, las microondas, las líneas punto a punto o la red telefónica conmutada y protocolos propietarios de cada fabricante. La información es enviada a los despachos de supervisión y control donde como tarea principal se controla que los distintos parámetros de la red no se salgan de los límites de seguridad preestablecidos para la explotación, además, se ejecutan una serie de programas que procesan la información recogida y la suministran a los distintos servicios de la empresa: gestión de mantenimiento, previsión de la demanda, ...

Las compañías no tiene acceso en sus despachos de control a toda la información que se genera en los puestos remotos. En la mayoría de los casos, parte de la información se agrupa antes de su envío para reducir los altos costes de transmisión, el ciclo de *polling* y el ancho de banda utilizado.

Otro problema que presenta el actual esquema de recogida y transmisión de la información es que la información recibida está confinada en ciertos servicios de la empresa, los despachos de control, y en muchos casos el acceso a ella por el resto del personal de la empresa queda reducida a una serie de informes periódicos que cubren ciertas necesidades pero no todas. Esta concentración de los datos en un único lugar provoca que se produzca una distribución mucho menor de la información que la que sería deseable para la explotación óptima de los sistemas de transporte y distribución. Aunque esta arquitectura ha satisfecho hasta ahora la mayoría de las necesidades de los usuarios, presenta inconvenientes como son el alto coste de cada punto captado y enviado, el alto coste de mantenimiento de las bases de datos y de las redes de transmisión, la necesidad de protocolos específicos y el hecho de que sea necesario disponer de unos puestos clientes específicos preconfigurados para el acceso a la información. La información sólo es recibida en directo por aquellas personas que dispongan de una consola del sistema de supervisión, el resto de los destinatarios de la información deben esperar a un segundo ciclo de distribución y recibirán la información elaborada por el sistema de supervisión y control que llegará a ellos en forma de informes o en ciertos casos, muy específicos, directamente a través de la red corporativa de la empresa.

En la actualidad, la recogida y transmisión de la información tiene como fin prioritario la supervisión de la explotación de la red y en en segundo lugar la ejecución de ciertos procesos con el objetivo de conseguir la optimización de la explotación de las redes de transporte y distribución. Sin embargo el proceso de liberalización de mercados, la necesidad de elevar la competitividad de las empresas mejorando su productividad y la exigencia de mantener o elevar la calidad de servicio exigen que la información esté disponible en todos aquellos lugares en los que sea de utilidad. Adicionalmente hay servicios de la empresa que necesitan en muchos casos la información tal y como se produjó y no según una visión que puede ser adecuada para tareas de supervisión pero no así para otro tipo de tareas como puede ser el mantenimiento, la previsión de la demanda, la información al cliente ...

En este articulo se describirá la oportunidad que nos ofrece la tecnologia Internet/Intranet para acceder a la información recogida en las remotas desde cualquier puesto de la empresa.

El conjunto de funcionalidades que se describen a continuación tienen como objetivo complementar las ya existentes en el esquema tradicional de supervisión y control y no pretenden en ningún caso sustituir a dicho sistema. Los actuales servicios de supervisión y control tienen unos requisitos de seguridad, garantía de fiabilidad y garantía de tiempo de respuesta que en ningún caso pueden ser igualados por la tecnología que se describe. La aplicación que se describe ofrece la oportunidad de utilizar la infraestructura ya existente para acercar la información a un mayor número de usuarios de forma que la empresa se beneficie de las oportunidades que ofrece el mayor uso de la información.

2. Un poco de historia

El origén de la Internet se remonta a 1.970 con la ARPANET, una red experimental creada por el departamento de defensa de Estados Unidos. A mediados de los años ochenta comenzó un periodo de crecimiento al conectarse a la red las primeras agencias gubernamentales, instituciones académicas, laboratorios de investigación e individuos lo que originó un rápido crecimiento de la red y su expansión por todo el mundo.

La existencia de la red creo la posiblidad de su utilización para el intercambio de información entre los ordenadores que estabán conectados a ella. Sin embargo, aunque la conexión a nivel fisico existia no habia ninguna forma estándar de intercambiar datos ni protocolos para utilizar eficientemente la infraestructura existente.

La necesidad que tenían los usuarios de la red de optimizar el uso que hacían de ella originó la aparición de las primeras aplicaciones para ampliar la utilización de la red:

- Telnet, ofrece la posibilidad de acceso remoto a los ordenadores, evitando desplazamientos, conociendo únicamente la dirección del ordenador al que se quiere acceder un nombre de usuario y la palabra de paso en el ordenador al que se deseaba acceder.

- Ftp, protocolo de transferencia de ficheros de gran utilidad para transportar información de un ordenador a otro pero con el incoveniente de que requiere saber donde está la información y el nombre del fichero a recuperar.

- USENET, un tablón de anuncios enorme dividido en distintos foros donde se puede encontrar información y participar en discusiones sobre casi cualquier tema.

- El correo electrónico que permite la difusión de la información persona a persona.

- Las listas de distribución que permiten la difusión de la información a una serie de personas a las que les interesa un tema y que se han subscrito previamente a la lista.

- Gopher, usado principalmente por universidades para proporcionar información sobre el campus.

- WAIS, una herramienta muy potente para la busqueda y recuperación de documentos.

El inconveniente de cada una de las herramientas mencionadas, alguna de las cuales sigue siendo muy utilizada hoy en día, era que cada una de ellas tenía una interfaz diferente y exigía la instalación y el conocimiento de protocolos diferentes. Otro de los inconvenientes que se producía en el intercambio de documentos era la multitud de formatos diferentes: texto, postscript, LaTex, troff, RTF y los formatos producidos por los distintos procesadores de texto de los ordenadores personales. El mismo problema se produce al intercambiar fichero gráficos debido a la multitud de formatos o con la información almacenada en bases de datos.

En 1.989, en el CERN, se inició el proyecto de la World Wide Web con el objetivo de crear un sistema que ayudará a los cientificos que allí trabajaban a buscar información y a intercambiarla usando Internet. El objetivo del proyecto era que la comunidad científica dispusiera de un único programa, un navegador, que les sirviera de interfaz comun a la multitud de protocolos y servidores de información que ofrecia Internet. La presentación de la información utilizando el hipertexto

se convirtió con rapidez en una parte central del proyecto. Los documentos relacionados contarían con enlaces hipertexto y seleccionando dichos enlaces se navegaría de un documento a otro. Dichos documentos no tenían por que estar situados en las mismas máquinas, el navegador se encargaría transparentemente de recuperarlos y presentarlos a los usuarios.

La World Wide Web comenzó a utilizarse internamente en el CERN en 1.991 y rápidamente se convirtió en un medio muy popular para el intercambio de artículos y resultados de experimentos científicos. A comienzos de 1.992 la iniciativa se hizo pública y comenzó a usarse como medio de intercambio de información, en primer lugar entre los laboratorios de física de alta energía, y posteriormente entre todo tipo de laboratorios y centros académicos.

Un hecho fundamental en el desarrollo de la Web fue la aparición en febrero de 1.993 del navegador Mosaic para X-Windows, producido por la NCSA en Estados Unidos. El Mosaic era un navegador con un interfase gráfico muy amigable que era capaz de incorporar en la misma página gráficos, texto y sonido. El NCSA proporcionó también un servidor Web de dominio público para Unix que contribuyó al rápido crecimiento del número de servidores web existentes en el mundo. A finales de 1.993 se presentaron las versiones Mosaic de Microsoft Windows y Apple Macintosh, con lo que a partir de ese momento se disponía de navegadores para las plataformas más populares.

La aparición del Mosaic originó un crecimiento exponencial de la red, pasándose de 500 servidores Web registrados en el CERN a finales de 1.993 a más de 4.600 a mediados de 1.994. Las estimaciones más recientes situan el número de servidores Web en más de 500.000 estimandose que la cifra se dobla cada seis meses.

3. Situacion actual de los centros de control

Los sistemas de control de las empresas suministradoras de gas, agua y electricidad han estado basados en arquitecturas cerradas con protocolos y aplicaciones propietarias. En los últimos años se está tendiendo a una migración hacia sistemas más o menos abiertos que permitén una mayor facilidad en la introducción de aplicaciones de ayuda a la operación que utilicen las bases de datos y los datos recogidos por los sistemas de supervisión y control de la red.

Hasta la fecha, las empresas han estado trabajando en un entorno de negocio regulado, pero esta situación está cambiando, y en un breve plazo de tiempo se va a pasar a un negocio desregulado en el que las empresas van a tener que competir entre ellas, proporcionando el servicio al menor coste posible sin bajar o incluso aumentando la calidad.

Aquellas empresas que ofrezcan una mejor relación calidad precio aumentarán su cuota de mercado en detrimento del resto. La aparición de la competencia va a exigir la optimización de los costes. Por otra parte, debido a presiones medioambientales cada vez va a ser más difícil la construcción de líneas eléctricas o gaseoductos, por lo que las infraestructuras existentes deberán usarse al límite, lo que exigirá una mayor supervisión y una mayor optimización en la explotación de la red. Por otra parte, el Estado va a exigir unos niveles de calidad de servicio que deberán ser respetados por las empresas si no quieren enfrentarse a graves sanciones.

Los condicionantes descritos anteriormente conllevarán unas mayores necesidades de análisis de la información generada por las remotas que hoy en día no es utilizada, pero cuya adecuada utilización puede convertirse en una gran ventaja competitiva.

La información podrá ser utilizada por varios tipos de usuarios:

- Las brigadas de mantenimiento podrán acceder a la información sin desplazarse hasta las instalaciones y realizar un primer análisis y diagnóstico de los problemas en función de la información recibida.

- El personal de la oficina técnica podrá realizar tareas de configuración de las remotas, análisis de incidentes, planificación de labores de mantenimiento desde sus propias oficinas.

- El personal de supervisión y control dispondrá de una información complementaria a aquella que le proporciona el SCADA accediendo directamente a las instalaciones. Podrán ver la totalidad de la información y no unicamente la información prefiltrada y agrupada que reciben en la actualidad.

- El personal de planificación de red podrá estudiar en función de los datos recibidos las inversiones necesarias para reforzar la red en aquellos lugares en las que se encuentre cercano al límite técnico de explotación.

- El personal de atención a clientes podrá conectarse directamente a las instalaciones para proporcionar a sus clientes información directa sobre los incidentes en la red, también podra estudiar patrones de utilización del servicio por parte de grandes clientes para ofrecer a éstos condiciones ventajosas en casos de interrumpibilidad por razones de mantenimiento o averia.

La gran evolución de las telecomunicaciones en los

últimos años, y sobre todo, la popularización de la tecnología Internet/Intranet permiten hoy en día ofrecer una solución de bajo coste para la distribución de la información. La información llegará a los usuarios de una forma más flexible aunque menos jerarquica con un coste casi nulo de los puestos clientes. Este alto grado de disponibilidad permitirá que todos los colectivos anteriormente citados y aquellos otros que puedan sacar un beneficio de la información puedan acceder a la información casi sin ninguna dificultad aprovechando las infraestucturas ya existentes en la empresa.

La facilidad de acceso a la red, sólo es necesario un modem, un PC y una línea telefónica o bién un PC conectado a la red corporativa de la empresa, así como la sencillez del interfaz de acceso a la información hacen que esta tecnología sea el puente que acerque la información a todos los usuarios que puedan utilizarla.

La información se ofrecerá usando como interfase de usuario los navegadores web. Los navegadores son fáciles de utilizar y casi universales. Hoy en día están disponibles de manera gratuita o a un coste mínimo en la mayoría de las plataformas y sistemas operativos más habituales en las empresas.

Las aplicaciones basadas en navegadores web ofrecen las siguientes ventajas:

- Multiplataforma, se simplifica el proceso de desarrollo y la curva de aprendizaje. Los fabricantes pueden abarcar un mercado más amplio y los usuarios pueden pasar de una plataforma a otra sin la necesidad de aprender un nuevo entorno.

- Escalables, se pueden dar servicios dentro de la empresa, Intranet, a ciertos usuarios externos Extranet o a todo el mundo, Internet.

- Facilidad de creación e instalación, existen multitud de herramientas que ayudan al desarrollo, por otra parte los clientes son casi universales por lo que se reducen los costes de formación y soporte.

4. Funcionalidades deseables

Las instalaciones con acceso Internet dispodrán de un servidor http que permitirá la conexión de los usuarios remotos con las instalación. En dicho servidor estarán almacenadas las páginas estáticas y dinámicas que proporcionaran la información.

La solución anteriormente descrita ofrecerá a los usuarios debidamente identificados y autorizados las funcionalidades que se describen a continuación.

Los usuarios podrán acceder a la instalación que deseán monitorizar seleccionando sobre su nombre en una lista de instalaciones o sobre su localización en un mapa.

Una vez que se ha accedido a la instalación, que estará definida como una URL, los usuarios deberán identificarse con su nombre de usuario y una palabra de paso. El sistema comprobará si la identificación de los usuarios es correcta y permitirá únicamente el acceso al resto de las funcionalidades a aquellos que estén autorizados. La aplicación podrá distinguir varios perfiles de usuario teniendo cada grupo acceso a una serie de funcionalidades en forma jerárquica como puede ser: supervisión, configuración, que incluye supervisión, y mando, que incluye a los dos anteriores.

Los usuarios podrán acceder a las bases de datos de las remotas y, si su perfil de usuario se lo permite, realizar las operaciones usuales en bases de datos: altas, bajas, modificaciones y consultas. Entre otros ejemplos de aplicación se podrán comprobar los límites de alarma de cada medida, los factores de conversión entre los datos leídos y las magnitudes reales, la medida, estado digital, contador o mando que corresponde a los distintos contactos de la remotas y cualquier otro dato que sirva para configurar la remota.

Los usuarios podrán visualizar los datos que está captando la remota, estos datos serán medidas que pueden corresponder a intensidades o tensiones en el caso de la red eléctrica o caudales y temperaturas en el caso de una red de gas. Se podrá comprobar el estado de los equipos, si un interruptor o una válvula está abierta o cerrada. Existirá la posibilidad de "navegar" dentro de una instalación o por las distintas instalaciones usando el ratón, al cambiar de una instalación a otra se comprobará si el usuario está autorizado a acceder a la nueva instalación.

Los usuarios tendrán la posibilidad de programar tendencias de forma que puedan monitorizar la evolución de una medida durante un periodo de tiempo predeterminado.

Los eventos que se produzcan podrán ser visualizados remotamente. Se visualizarán todas aquellas situaciones de alarma programadas previamente en la remota límites: de magnitudes, presencia de intrusos en la instalación, detección de humos ....

Si el usuario está debidamente autorizado se podrá realizar la configuración remota de la instalación con el consiguiente ahorro en tiempo y viajes que esta actividad supone a las empresas.

Otra funcionalidad que se puede implementar, pero que hoy en día no se considera por la reticencia de las empresas, es el mando remoto de forma que el operador debidamente identificado y autorizado pueda abrir o cerrar válvulas o interruptores. La tendencia actual es

que el mando remoto quede localizado en aquellos operadores del SCADA, cuya misión es la supervisión y la adecuada configuración de la red para que se cumplan los criterios de seguridad y óptimo de explotación de la red.

El usuario podrá solicitar un informe sobre alguno o varios de los valores almacenados en las bases de datos de históricos de la instalación. Podrá ver el informe por la pantalla o acceder a él a través de la impresora, ya sea directamente si su navegador se lo permite o indirictamente recibiéndolo por correo electrónico si por razones de seguridad no tiene la posibilidad de imprimir datos remotos en su navegador.

Los servidores instalados en las instalaciones remotas generarán informes periódicos automáticamente y los enviarán por correo electrónico a los usuarios previamente registrados. Dichos informes podrán contener información sobre medidas que puede interesar al personal que planifica la red o sobre operación de los distintos equipos o automatismos que puede ser usado por el personal que planifica el mantenimiento de la red o sobre los consumos de los clientes que interesaría al departamento comercial.

El sistema puede enviar también avisos ante la detección de situaciones anormales en la instalación al personal encargado de su mantenimiento, el tipo de información que se puede enviar es baterias bajas, error en base de datos, ...

Los usuarios podrán realizar sobre las instalaciones simulaciones estáticas o dinámicas que podrán usarse para análisis del comportamiento de la red en condiciones extraordinarias o para realizar previsiones sobre como se comportará la red ante aumentos significativos de la demanda.

Otra utilidad de este tipo de aplicaciones combinando la visualización de datos en tiempo real con la simulación es la formación de los operadores en un entorno muy similar al que van a trabajar. La formación puede ser orientada tanto a nuevos operadores, enseñandoles los mecanismos habituales de trabajo, como a operadores expertos, situandoles en situaciones de emergencia y analizando como reaccionan.

Las funcionalidades anteriormente citadas hay que proporcionarlas integradas de la forma más sencilla posible con los sistemas ya existentes.

Como requisitos añadidos podríamos citar los siguientes: la información existente en los puestos remotos no debe ser duplicada y los servidores deberían autoconfigurarse al conectarlos a los controles ya existentes de forma que aprovechen la información ya existente en las posiones remotas para su inicialización.

5. Puntos débiles de Internet para monitorización

El protocolo http es un protocolo basado en transacciones sin noción de estado. El hecho de que el protocolo no tenga noción de estado implica que el servidor no almacena ninguna información sobre el cliente ni sobre sus peticiones. El hecho de que el protocolo esté basado en transacciones significa que la interacción se implementa mediante transacciones simples en las que el cliente abre una conexión con el servidor, manda su petición, recibe la respuesta a su petición o un error si la petición no puede ser atendida o está incompleta y cierra la transacción.

Para implementar la clase de funcionalidades que se describen en este artículo se requiere que los servidores tengan noción de estado, de forma que si se está monitorizando una posición con una serie de datos el servidor envie sólo aquellos que hayan variado y unicamente al usuario que los está viendo. El mayor problema que se presenta en la monitorización remota usando esta tecnología es implementar la noción de estado sobre el protocolo http.

Se puede implementar el estado de tres formas en una aplicación cliente-servidor:

- El cliente almacena el estado, en este caso cada petición debe incorporar como parametro la descripción completa del objeto que la realiza.

- El servidor almacena toda la información sobre el estado y el objeto tiene unicamente que pedir la información indicando la identificación del estado a que se refiere.

- El servidor almacena unicamente el estado actual del objeto, olvidando sus estados previos de forma que cualquier petición se refiere al estado actual.

Para el tipo de aplicaciones que se describen en este artículo se cree que la segunda solución es la más apropiada de forma que el cliente sea un simple navegador. El servidor podrá recibir peticiones simultáneas de varios usuarios, cada uno de ellos analizando una versión diferente del proceso. La posibilidad de que exista varios usuarios monitorizando distintas variables del proceso exije que el servidor necesite mantener, de forma independiente, la historia de las peticiones de cada usuario durante la sesión para proporcionarle correctamente la información requerida en cada momento.

El refresco de la información dinámica que estén monitorizando los usuarios es otro aspecto importante en este tipo de proyectos. Usando técnicas actuales como puede ser el CGI, *commom gateway interface*, los servidores generarían una página nueva cada vez que

fuera necesario refrescar un dato. Sin embargo para reducir la carga en el servidor y para aumentar la velocidad de respuesta se deben estudiar otras alternativas. Como alternativa los *applets* que están incluidos en la página html inicialmente suministrada por el servidor, pueden refrescar sus datos dinámicos usando al protocolo de transmisión de ficheros, ftp, o bién abriendo una comunicación dedicada mediante un socket TCP/IP.

El sistema de control y supervisión que existe en la actualidad es prioritario respecto a la solución aqui propuesta por lo que a la hora de realizar el diseño y la instalación de este tipo de aplicaciones hay que tener en cuenta que no deben interferir con las aplicaciones tradicionales de supervisión y control. El sistema debe permitir que las remotas realizen las tareas que tienen asignadas y que atiendan a está segunda via de acceso a la información en los tiempos muertos.

Otro aspecto a tener en cuenta es la seguridad. La aplicación de monitorización puede permitir ofrecer un mayor servicio si se permite el acceso a una serie de usuarios autorizados desde fuera de la empresa, por ejemplo desde grandes clientes o desde proveedores por lo que esté tipo de aplicaciones necesitarán de los adecuados controles de seguridad que eliminen la posibilidad de acceso a los usuarios no autorizados.

6. Conclusiones

El trabajo que se describe en este articulo demuestra que es posible la utilización de la Internet para acercar la información a todos los servicios de la empresa de forma que desaparezcan las islas de información y de forma que cada usuario pueda acceder de la forma más rápida posible a la información que necesite. El valor añadido de la información va a aumentar ya que cada vez va a poder utilizarse más rápido y por más usuarios con un escaso aumento de los gastos.

La utilización de está tecnología es relativamente barata respecto a los beneficios que produce por lo que se cree que se extenderá su utilización y cada vez será más facil acceder a información on-line allá donde se necesite sin tener que recurrir a los canales tradicionales de petición de información off-line.

Los servicios de supervisión y control van a seguir existiendo ya que la tecnología aqui descrita no puede ofrecer ni la seguridad ni el tiempo de respuesta que requieren dichos procesos.

Este tipo de aplicaciones van a permitir la colaboración más estrecha entre los distintos servicios de la empresa y va a permitir que tanto los suministradores como los grandes clientes accedan a la información con el consiguiente beneficio.

Referencias

[1] Ed Tittel, Mark Gaither, Sebastian Hassinger, Mike Erwin, "Fundamentos de programación con HTML & CGI". Anaya Multimedia, 1.996.

[2] Lincoln D. Stein, "How to setup and maintain a Web Site. Lincoln D. Stein. Addison-Wesley, 1997.

[3] Bill Ackerman, Chuck Adamson, David Joy, "Electrical Utility benefits form acces machine technology". *Utility Automation*, January-February 1997 56-60

[4] Robert J. Thomas, "The Internet, Java and Objects for Netwok-Centered Computing and Communications with Applications for Power Systems". *IEEE PICA* (1997)

Aplicación de la Telemática en las PYMEs. Proyecto TRANSMETE

Armando Ferro, Mikel Olabe y Juanjo Uncilla
{jtpfevaa|jtpolbam|jtpungaj}@bi.ehu.es
Telematic Networks Group
Escuela Técnica Superior de Ingenieros Industriales y de Telecomunicaciones de Bilbao
September 15th, JITEL'97

Abstract

The TRANSMETE project has developed and delivered courses on telematic applications for SMEs in Greece, Spain, Sweden and Finland. The project's main objective is to provide SMEs with information, consultancy services, and training about the benefits of using telematics in their day-to-day activities. The TRANSMETE project has developed and delivered courses on telematics applications. The training programme consists of training courses and an innovative business game which will raise the awareness of SMEs and will assist SMEs to experience by themselves the benefits of telematic applications. About 1000 people coming from SMEs (managers and personnel) will attend the 16 hours course and will participate to the 10 hours business game, at five training sites. During the course, the participants will gain the knowledge of the most commonly user telematic applications, but in an "SME activity- oriented" way. At the end of the course, participants will be able to perform, using the applications, the operations required for each one of the activities of an SME. The business game will support the training as an extended practical exercise, where the participants will be able to use the applications by themselves.

1. Introducción.

TRANSMETE es un proyecto parcialmente financiado por la Comisión Europea en el marco de trabajo del Programa de Aplicaciones Telemáticas para Educación y Formación.

La meta principal del Programa de la Comisión es ampliar la labor de investigación en Telemática para Educación y Formación y así equipararse con los avances de otros países desarrollados haciendo uso de los continuos progresos en comunicaciones multimedia vía redes de banda ancha o satélites y obtener avances en aplicaciones de simulación interactiva o sobre entornos virtuales.

Los subsectores y tareas que están más relacionados con estos objetivos son:

- Aplicaciones para Educación y Formación.
- Herramientas y aplicaciones innovadoras.
- Fuentes de apoyo específicas para la Telemática.

El Programa donde se enmarca TRANSMETE corresponde a una actividad de la Dirección General XIII de Telecomunicaciones conocida como Mercado de Información y Explotación de Investigación de la Comisión Europea.

TRANSMETE es un proyecto de cooperación europeo que se inicia en Octubre del año 1995 y que actualmente sigue en fase de realización. El consorcio está constituido por empresas expertas en formación, consultoras, institutos de educación, así como organizaciones y cámaras de comercio cuya actividad principal les hace estar en permanente contacto con las PYMES.

Este proyecto dispone de 4 centros de formación e información situados en Grecia, España, Suecia y Finlandia. Los socios participantes en España son Carsa, Instituto Catalá de Telemática Aplicada y NEXTEL.

El objetivo principal de TRANSMETE es ayudar a las PYMES a beneficiarse de las ventajas proporcionadas por la evolución de las aplicaciones y servicios telemáticos, por lo que este proyecto asistirá a las PYMES objetivo mediante acciones de formación, información y asesoramiento en la utilización provechosa de la telemática en las actividades diarias de la empresa.

2. Proyecto TRANSMETE

TRANSMETE es un proyecto europeo que tiene por objeto ayudar a las pequeñas y medianas empresas (PYMES) en la incorporación de las nuevas aplicaciones y servicios telemáticos. El proyecto pretende proporcionar a las empresas apoyo en las siguientes áreas:

- Formación
- Información
- Asesoramiento

Esta labor está enfocada a la aplicación de la telemática para resolver los problemas que la empresa tiene en su actividad diaria.

¿Por qué, empresas de pequeño y mediano tamaño?

Porque hay 15 millones de PYMES en el sector primario que emplea más de 68 millones de personas en la Unión Europea, lo cual es más de dos tercios del empleo comunitario. Las PYMES generan más de dos tercios del retorno de la comunidad europea y entre el 65% y el 85% del valor añadido en aquellos países de cuyos datos se dispone[1].

La comisión europea reconoce que las PYMES cumplen un papel crucial en el eslabón entre crecimiento y empleo. Por otro lado existe además una inquietud creciente por parte de las autoridades públicas en relación a las dificultades de estas empresas para mantener su competitividad en un entorno de mercado y competencia internacional.

Los estudios revelan que uno de los aspectos clave es el grado de adquisición y adaptación de las PYMEs a las nuevas formas de actuación que resultan del avance tecnológico en el campo de la información y las comunicaciones.

La Comunidad Europea quiere por tanto, disponer de una estrategia de respaldo diseñada para facilitar a los negocios, particularmente a las PYMEs, que se adapten a los nuevos requerimientos de competitividad y asegurar que los factores económicos puedan movilizarse adecuadamente para soportar el crecimiento, la competitividad y el empleo[2].

2.1 Objetivos del proyecto

El objetivo principal del proyecto TRANSMETE es proporcionar a las PYMEs suficiente información, servicios de consultoría y entrenamiento práctico sobre los beneficios que le supone la aplicación de la telemática en sus actividades del día a día. Para ello se ha desarrollado y distribuido cursos sobre aplicaciones telemáticas para las empresas en Grecia, España, Suecia y Finlandia.

2.2 Fases en la realización del proyecto

La realización del proyecto tiene definidas las siguientes fases:

1. Análisis de requerimientos. Empresas PYMEs representativas de sus sectores han sido analizadas[4] para identificar las actividades que realizan periódicamente y son susceptibles de mejora aplicando soluciones telemáticas (Ejemplo: actividades finanieras, de marketing, etc).
2. Especificación de las soluciones. Identificación de los servicios telemáticos que resuelven de forma satisfactoria las necesidades detectadas y, más importante, la definición de las metodologías que la empresa debe de utilizar para proceder a la incorporación de estos servicios de forma productiva.
3. Desarrollo del material de formación y divulgación. Los contenidos estarán más orientado a la actividad de la empresa que a la mera descripción técnica.
4. Divulgación hacia las PYMEs de los resultados obtenidos demostrando así la validez de las metodologías. Se realizarán cinco WorkShops en los diferentes países con más de 300 PYMEs.
5. Acciones de formación. Las metodologías de implantación desarrolladas formarán la base de un curso de formación a empresas de 16 horas y de un curso de entrenamiento (business game) de 10 horas.
6. Evaluación de los resultados del proyecto. Se evaluará el material desarrollado, la calidad de la labor pedagógica, el acierto y mejora de los procedimientos de enseñanza, los aspectos socio-económicos introducidos y la aceptación por parte de los usuarios PYMEs de los resultados del proyecto.

En la actualidad están finalizada todas las fases del proyecto y se está a la espera del análisis definitivo de los resultados recogidos en la fase de Evaluación de resultados. El consorcio sin embargo mantendrá las labores de difusión y formación por un periodo más largo en busca de asegurar los resultados obtenidos.

2.3 ¿Quiénes se ven beneficiados por la acción de este proyecto?

El aspecto más sobresaliente que ofrecen los servicios y aplicaciones telemáticas es la oportunidad de abrir nuevos caminos para romper las barreras locales y para darse a conocer en el mercado europeo e internacional que cada vez es más extenso.

Los agentes que principalmente se ven beneficiados por la acción de este proyecto son los siguientes:

- *Las PYMEs* ya que tienen la posibilidad de participar en sesiones de formación particularizadas a sus necesidades y obtienen toda la información y asesoramiento necesarios para mejorar la eficiencia de sus actividades cotidianas.

- *Los proveedores de servicios y aplicaciones* telemáticas ya que amplian así sus oportunidades de mercado y aumentarán sus servicios en el sector de PYMES.
- *Las organizaciones de formación* que les permite conocer un enfoque innovador en la formación que busca obtener una mejor personalización de los materiales de formación y una herramienta valiosa para adaptarse mejor a las necesidades de los usuarios, horarios desplazamientos, etc.

Las PYMEs necesitan adaptar y utilizar nuevos desarrollos tecnológicos según estos surgan. Las Aplicaciones y Servicios Telemáticos ofrecen especialmente nuevas formas de eliminar las barreras tanto para darse a conocer en los mercados locales como en los internacionales.

De este modo las PYMEs serán capaces de competir en una base de mayor igualdad con las grandes compañías. Las relaciones cautivas contratante-suministrador se debilitarán. Las empresas serán más competitivas, crecerán más rápido y se crearán más empleos. La relación con las administraciones será más simple y productiva [3].

La comisión europea, los gobiernos y los proveedores de tecnología reivindican que los servicios y aplicaciones telemáticas, ayudarán a las empresas en su esfuerzo para incrementar su productividad, reducir sus costes y ampliar su base de mercado. De esta forma se espera fortaler el crecimiento de empleo en Europa.

2.4 ¿Cómo se proporcionan estos servicios?

De todas formas en la mayoría de las ocasiones estos servicios y aplicaciones se presentan a la PYME con un carácter muy técnico y muy general sin demostrar el beneficio específico para cada una de ellas y sin mostrar ejemplos concretos y reales de su modo de operación y beneficios alcanzados.

El proyecto TRANSMETE busca romper el vacío de comunicación entre los proveedores de tecnología y las PYMEs. Mostrando la mejor imagen posible de las nuevas formas de hacer negocio en estas empresas mediante la aplicación de las nuevas tecnologías. Se atenderán a aspectos como los enumerados seguidamente:

- ¿cómo realizan sus ventas y actividades de mercado?
- ¿cómo se comunican y hacen frente a la competencia?
- ¿cuales son sus miedos y visiones para el futuro?.

Otro asunto a tratar será como las nuevas tecnologías y las aplicaciones y servicios telemáticos se pueden integrar en las actividades del día a día en las PYMEs y cuál es el mejor enfoque a seguir cuando se introducen estas tecnologías en algunos medios empresariales.

La clave está en conseguir una visión general acertada sobre la forma de trabajar de las PYMES. Esta visión general mostrará cómo las modernas tecnologías y las aplicaciones y los servicios telemáticos pueden integrarse en las actividades diarias de las PYMEs, y cual es el camino de acercamiento pedagógico a seguir más apropiado cuando se introducen estas tecnologías en el campo de los negocios.

3. Metodología del proyecto.

La metodología del proyecto[5] está enfocada hacia tres direcciones:

- Productos y servicios orientados hacia el usuario
- Entrenamiento sobre aplicaciones Telemáticas
- Aumentar la participación de las PYMEs.

Estos son los aspectos fundamentales en la metodología del proyecto TRANSMETE. El resultado final es un programa de formación en aplicaciones telemáticas, diseñado y desarrollado para cubrir las necesidades específicas de las PYMEs. El programa quiere demostrar cómo la telemática puede ser fácilmente integrada en la estructura de las PYMEs, y cómo éstas pueden aumentar su productividad y mejorar su competitividad utilizando estos servicios.

3.1 Acciones de divulgación del consorcio

Un aspecto clave en la metodología del proyecto es incrementar el conocimiento de las PYMEs en relación a las aplicaciones y servicios telemáticos proveyéndoles con información apropiada. Para ello se han abierto las siguientes acciones de divulgación:

- *Workshops* en Grecia, España, Suecia y Finlandia, con presentaciones de los propios usuarios y de los proveedores de la tecnología sobre las ventajas de las aplicaciones telemáticas, de tratamiento, y de proceso de la información, acompañados de exhibiciones de las aplicaciones en las actividades profesionales tradicionales de la vida real.

- *Publicaciones Regionales trimestrales* (TRANSMETE News) producidas en Grecia, España, Suecia y Finlandia, proporcionando a las PYMES una información práctica sobre las acciones del proyecto planeadas, así como noticias regionales de las PYMEs.

- *Un CD-ROM demostración,* diseñado, desarrollado y distribuido para las PYMEs, como una única herramienta de formación standard y en diferentes lenguajes (inglés, español, griego, etc.), proporcionando información útil sobre el proyecto y otras materias de ayuda.

- *Información WWW* (TRANSMETE WebSite) incluyendo información general sobre el proyecto y sus actividades, noticias, publicaciones, y con enlaces a las paginas Web de los socios involucrados.

Estos son los canales de comunicación regulares que el consorcio utiliza para proporcionar a las PYMEs información útil.

3.2 Acciones de formación e información

Este servicio se ofrece a los participantes en los cursos desde los centros de formación TRANSMETE.

TRANSMETE desarrolla y distribuye un programa completo de formación para Pymes compuesto de:

- *16 horas de formación* de como usar los servicios y aplicaciones telemáticas para desarrollar el negocio de sus empresas. Mil directivos de empresas serán objeto del seminario.
- *Business game*. Durante el juego los usuarios serán capaces de establecer una presencia en internet, hacer negocios, intercambiar información y comunicarse usando la telemática.

Además de formación e información, TRANSMETE proporciona servicios de asesoramiento que buscan superar las barreras de entendimiento entre la PYME y las empresas desarrolladoras de aplicaciones mostrando la forma mejor de utilizarlas en el desarrollo de oportunidades de negocio, mejora del servicio del cliente, en la cooperación con los proveedores, etc.

Durante el periodo de formación, las empresas pueden resolver gran parte de las incógnitas que les surgen en cuanto a la forma de aplicar las tecnologías de la información en su propio beneficio. Entre otras, las cuestiones más frecuentes que aparecen y son resueltas en estos cursos suelen ser:

- ¿Cuáles son las aplicaciones y servicios telemáticos más extendidos?
- ¿Qué es y cómo se puede acceder al mercado electrónico?
- ¿Cómo preparar productos y servicios para el mercado electrónico?
- ¿Cómo establecer y facilitar las cooperaciones internacionales?
- ¿Cómo acceder a información útil para la empresa?
- ¿Cómo preservar la seguridad de la información?
- ¿Cómo facturar?

Durante las sesiones de formación las PYMEs se familiarizan con algunas de las aplicaciones telemáticas más extendidas y tienen la posibilidad de explorar internet y descubrir allí ejemplos reales de empresas con negocios en internet.

3.3 Business game

El business game muestra como se llevan a cabo nuevos caminos de cooperación entre empresas a través de naciones y culturas utilizando la telemática para acelerar los procedimientos de producción de desarrollo y añadir valor a todas las empresas participantes.

En el transcurso de los cursos de TRANSMETE los directivos de las empresas pueden realizar una serie de ejercicios prácticos reales (business game) que les ayudará a la hora de decidir el aprovechamiento de estas tecnologías para reforzar la actividad de su negocio. Así ellos mismos pueden:

- experimentar por su cuenta el uso de aplicaciones telemáticas,
- establecer una presencia y establecer contactos con otros centros de formación que participan en TRANSMETE
- intercambiar información y explorar el mercado usando herramientas y servicios telemáticos

En cada uno de los centros de formación TRANSMETE de cada país se definirán dos empresas virtuales que van a actuar como empresas piloto del juego. Estas empresas estarán constituídas por las PYMEs que participan en el curso de entrenamiento. El consorcio TRANSMETE actuará como empresa consultora encargada de la

coordinación, planificación, soporte técnico y asesoramiento.

El objetivo del juego es utilizar las herramientas telemáticas descubiertas para resolver las problemáticas que van a ir surgiendo en el desarrollo del mismo. Estas estarán orientadas a las siguientes actividades:

- Organización y formación del equipos de trabajo entre empresas ubicadas en diferentes países.
- Obtención y captura de información entre socios y utilizando recursos externos
- Desarrollo de la creatividad en las diferentes actividades de negocio
- Control de proyectos multinacionales y planificación
- Desarrollo de productos de forma cooperativa

Las empresas participantes asumirán un papel similar al de su actividad real bien sea en asuntos económicos, ingeniería mecánica, construcción, marketing, electrónica, etc. Esto permitirá familiarizar a cada empresa con los beneficios concretos que le aportan las nuevas tecnologías de forma directa. Las empresas virtuales del juego actuarán como empresas subcontratadas por un cliente común.

De esta forma los participantes pueden mejorar su creatividad y conocimiento poniéndose en contacto con personas de diferentes naciones y culturas para realizar negocios de forma conjunta.

El juego se desarrolla en los centros de formación los cuales están dotados de la infraestructura telemática adecuada como: líneas telefónicas, telefax, odenadores, acceso a Internet, servidores Web, videoconferencia.

Este ejercicio estimulará a los participantes en el uso de nuevas técnicas para la comunicación y establecimiento de relaciones más allá de las fronteras entre países y culturas. Además podrán intercambiar experiencias profesionales con otros participantes lo cual les mostrará las similitudes y diferencias entre empresas de diferentes países.

Una experiencia interesante para los participantes es el estudio del impacto de estas tecnologías en la evolución de la empresa. Por ejemplo las PYMEs podrán evaluar la posibilidad de contratación de empleados virtuales a lo largo de diferentes países con la aportación de diferentes perfiles de conocimiento y que repercusión puede tener eso en su negocio.

4. Contenidos técnicos del material divulgativo.

Dentro de la metodología se ha diseñado un curso de formación[6] específico para orientar a las empresas hacia la implantación en su entorno de las soluciones telemáticas más adecuadas.

Se ofrece a los participantes formación básica respecto a los recursos telemáticos que pueden estar a su alcance. Se les proporciona conocimiento de la infraestructura telemática de su entorno, así como de las herramientas software disponibles en el mercado. Esta formación está enfocada a cubrir las necesidades reales de su negocio y se les enseña como pueden utilizarlas para mejorar la capacidad de su empresa.

El curso diseñado inicialmente ofrece información básicamente en dos entornos diferenciados; soluciones INTERNET y soluciones EDI.

Dentro del mundo de Internet se familiariza a los directivos de las empresas con las aplicaciones más extendidas y que pueden ofrecer mayor cobertura a sus negocios como pueden ser:

- Correo electrónico (E-mail)
- Servidores de noticias (NEWS)
- Servidores de ficheros (FTP)
- Navegadores (Browsers WWW)

Se presta especial hincapié en las soluciones INTRANET y las posibilidades de futuro de esta tecnología como agente integrador del resto de tecnologías para ofrecer soluciones fáciles de implementar y económicas.

En cuanto a las soluciones EDI se muestra a los participantes aquellos aspectos más interesantes de la tecnología y como puede cambiar la forma de hacer negocios de determinados sectores (como la banca, automóvil, etc) en el futuro según se vaya realizando su implantación progresiva. Para ello se realiza un estudio de implantación de un proyecto EDI atendiendo a los siguientes aspectos:

- Análisis del negocio
- Definición de necesidades
- Implantación del proyecto
- Evaluación de los beneficios

En los contenidos del seminario de formación también se ofrece información sobre soluciones de videoconferencia mostrando a los directivos las aplicaciones posibles, el estado actual de la tecnología y las implicaciones de su implantación.

En todos los aspectos se les proporciona asesoramiento y se les ofrece una serie de consejos útiles para la toma de decisiones en cada implantación y los pasos que deben de realizar para su consecución.

5. Relación con otros proyectos.

El consorcio está interesado en cooperar con otros proyectos dentro del Programa de la Comisión Europea de "TELEMATICS APPLICATIONS" manteniendo un intercambio razonable de recursos y de información no-confidencial. Especialmente se está cooperando activamente con proyectos del área de "Support Action Projects" que pretenden difundir y realizar actividades en el marco de la Comunidad para desarrollar estudios de impacto económico y social en Europa.

Otros proyectos que se han realizado o se están realizando actualmente y con los que se tiene relación a través de los diferentes socios del consorcio son:

1. Erasmus Co-operation: Con la Universidad de Abertay Dundee, la Escuela politécnica de Kemi-Tornio.(Finlandia), La Universidad de Lulea (Suecia) y la Escuela Fachhochschule de Augsburg (Alemania).
2. Cursos de formación realizados en el programa COMETT: El Instituto Oulu (Finlandia) y la Politécnica de Kemi-Tornio (Finlandia) realizaron dos propuestas de cursos dirigidos a las PYMEs para el programa COMETT que fueron aceptadas; International Driver's Licence for Information Highways and Network Service Designer/Developer for Information Highways.
3. El programa de formación europeo LEONARDO DA VINCI con el cual se han establecido vínculos de interés.
4. Es de destacar también los proyectos PRONET, IDEALS y AGORA así como el proyecto ABUITSS del programa Esprit IV.

6. Aportaciones del proyecto

6.1 Aportaciones a la Comunidad Europea

El proyecto RANSMETE ayudará a disminuir el aislamiento de las PYMEs en las regiones europeas menos favorecidas permitiéndoles familiarizarse con las tecnologías alrededor de la información más modernas (E-mail, transferencia de ficheros, EDI, telecomercio, videoconferencia, teleaprendizaje, etc). De esta forma las PYMEs pueden desarrollar una capacidad de competencia potencial que les permita mejorar sus productos, reorganizar sus recursos y encontrar nuevos y más amplios mercados. La descentralización de los mercados es uno de los logros que se alcanzarán en un futuro próximo con la aparición del mercado común unificado en Europa.

La política asumida por la Unión Europea sobre la creación de entornos unificados y de no competencia en la Sociedad de la Información en Europa está perfectamente respetada dentro del proyecto TRANSMETE dado su carácter abierto y el uso de aplicaciones Telemáticas estándarizadas y de uso general en las PYMEs. El beneficio de esta política es doble. Por una parte se favorece el acceso directo de las empresas PYMEs a la información, por ejemplo a la publicada por la Unión Europea (EU), sus regulaciones y actividades mientras que anteriormente las empresas sólo accedían a esta información a través de informes o publicaciones específicas. En segundo lugar como resultado de las acciones y actividades del proyecto TRANSMETE las PYMEs adquirirán conocimiento de como funciona el mercado europeo de colaboración entre socios comerciales de diferentes países lo que facilitará la política europea de eliminación de barreras regionales y promoción de la cooperación entre empresas dentro de la Unión.

6.2 Impacto económico y Social

La aparición de servicios telemáticos de valor añadido incorporados por las PYMEs permitirá a los Gobiernos locales, las Cámaras de Comercio y las Asociaciones de Empresas ofrecer a las PYMEs programas integrados de desarrollo de carácter local o regional proporcionándoles los servicios requeridos por las empresas así como promoviendo campañas de información y aprendizaje utilizando el material desarrollado en el ámbito de este proyecto.

Las PYMEs podrían obtener los siguientes beneficios de la labor promotora de estos organismos:

- Información sobre los recursos existentes que mejor les orienten en su mercado regional
- Navegar por las bases de datos del promotor y obtener información de sus mercados, competidores, productos, de las actividades de la EU, etc.
- Orientación sobre la forma de introducirse en los mercados exteriores.
- Mejor organización de los recursos propios y mayor facilidad de establecer relaciones con clientes, proveedores, etc.
- Una información tecnológica en su sector perfectamente actualizada y que le permite

mantener la posición de liderazgo en el mercado
- La incorporación de soluciones sectoriales como el EDI que le pueda permitir nuevas áreas de negocio y la simplificación de los procesos administrativos.
- Estar en contacto con los centros de Investigación de su entorno e incluso exteriores. De esta forma se puede paliar en parte las deficiencias crónicas de las PYMEs en cuanto a su evolución en tecnologías y metodologías aplicadas en I+D.
- Tener una herramienta fácil y asequible que les permita estar informados de los productos y tendencias del mercado.

Las empresas proveedoras de servicios aumentarán la base de sus clientes potenciales con la incorporación de las PYMEs dentro del mercado de consumo telemático. Además con la incorporación en las PYMEs de infraestructura telemática, surgirán nuevas necesidades y servicios que es necesario cubrir.

Las empresas de formación podrán utilizar los resultados del proyecto TRANSMETE para cubrir las necesidades de sus nuevos clientes. La metodología del proyecto les permitirá un acercamiento a los requerimientos de las empresas para cubrir sus demandas.

7. Evaluación de resultados

Dentro del marco del proyecto se ha realizado una prospección del mercado de las PYMES en los diferentes países para estimar su grado de aceptación, el nivel de preparación y el tipo de aplicaciones cuyos contenidos sean más susceptibles de integración.

Durante esta fase de evaluación se ha seguido una estrategia previamente planeada[7] y cuyos criterios de evaluación se han establecido en busca de los siguientes retornos:

- La calidad pedagógica de los seminarios.
- La efectividad del procedimiento de enseñanza
- El impacto socio-económico
- La aceptación por parte de los clientes

Esta prospección inicial se ha realizado en base a los seminarios de difusión que se han llevado a cabo en la fase de demostración sobre una población muestra en cada uno de los países.

Los resultados recopilados en la fase de demostración se emplearán para realizar un análisis cualitativo y estadístico.

Los datos recogidos se han obtenido en base a unos cuestionarios que los participantes han rellenado después de atender a la labor de difusión realizada en el marco del proyecto generalmente finalizada con un seminario de formación específico diseñado en base a los resultados de investigación del proyecto TRANSMETE.

Los aspectos que se han considerado interesantes de recopilar se pueden agrupar en varios aspectos:

- Información sobre la realización de la acción de difusión. Esto permitirá enfocar de una forma más correcta los mecanismos de acercamiento hacia la empresa.
- Información sobre los beneficios directos generales y parciales obtenidos por las empresas
- Información sobre las inquietudes que se originan en las empresas en la aplicación de las nuevas tecnologías descubiertas.

En la fecha de presentación de este artículo sólo se cuenta con los datos recogidos en la labor de difusión realizada en España, y más concretamente sobre una población muestra del País Vasco.

Seguidamente se expresarán los resultados más interesantes recogidos de las encuestas realizadas a los asistentes a las jornadas de demostración que se han realizado en el País Vasco a grupos de PYMEs.

La asistencia a los seminarios generalmente se ve motivada por decisión propia (85%) y en algunos casos de la empresa (15%).

Los contenidos del seminario son adecuados para una gran mayoría (80%). Un porcentaje reducido (20%) cree que parcialmente.

El 70% de los asistentes entiende que han sido satisfechas sus pretensiones previas de una forma completa, y el 30% parcialmente.

El 87% entiende que podría aplicar los conocimientos adquiridos en su empresa y un 13% lo cree pero parcialmente.

La documentación recibida en las jornadas es de buena calidad (94%) en cuanto a contenidos y en suficiente cantidad (87%) para la gran mayoría.

La calidad de las presentaciones ha sido excelente sin ninguna duda.

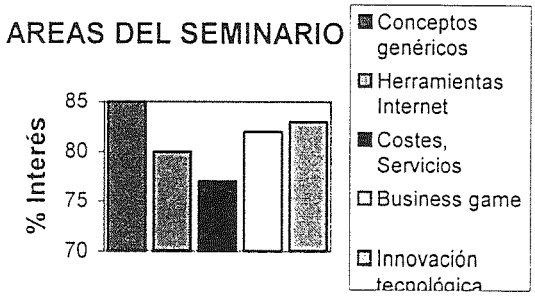
Los asistentes no han tenido dificultades de seguimiento en un 72% y parcialmente en un 28%. Los contenidos han sido cercanos a las necesidades

de las empresas para un 66% y solamente de forma parcial para un 34%.

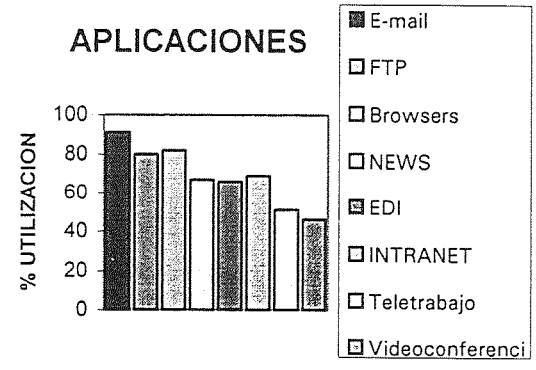
La transferencia de conocimiento hacia las empresas ha sido alta para un 39% e intermedia para el 61%. Los contenidos han sido muy teóricos para un 28% y el 72% entiende que han sido equilibrados.

La duración del seminario se considera adecuada para un 51% y demasiado corta para un 49%. En todos los casos los asistentes recomiendan la asistencia al seminario a terceras personas.

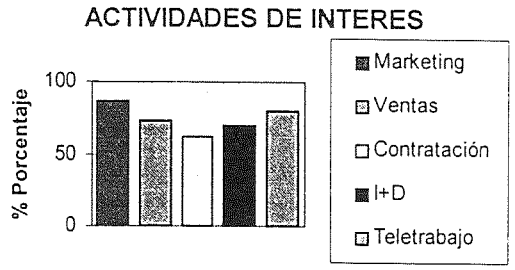
Las áreas del seminario que más interés han despertado y por orden decreciente han sido; Conceptos genéricos sobre Telemática (85%), innovación tecnológica (83%), Business game (82%), herramientas internet (80%) y los costes y servicios de las soluciones telemáticas (77%).



En cuanto a las aplicaciones telemáticas que se han mostrado en los seminarios y el interés que los participantes podían tener en incorporarlas en sus organizaciones para proporcionar valor añadido a sus negocios los resultados han sido; E-mail (91%), FTP (80%), Navegadores (82%), News (67%), soluciones EDI (66%), soluciones INTRANET (69%), aplicaciones de teletrabajo (51%) y videoconferencia (46%).



Por último, las actividades de la empresa donde la aplicación de la telemática demuestra mayor interés según los resultados de los seminarios son; el Marketing (86%), el apoyo a Ventas (73%), la contratación con clientes y proveedores (62%), las labores de I+D o actualización del conocimiento de la empresa (70%) y el teletrabajo (80%).



8. Conclusiones

TRANSMETE es un programa liderado por PYMES, que proporciona a éstas, a las organizaciones relacionadas y a los proveedores de tecnología, la información y conocimientos requeridos para una satisfactoria introducción de la telemática en sus actividades operativas diarias.

El curso cubre la formación en el acceso al mercado internacional, preparación de productos hacia el mercado exterior, desarrollo de una estrategia de marketing, cooperación internacional, venta y facturación. Estas materias constituyen el armazón principal bajo el cual se presentan los servicios telemáticos y sus aplicaciones.

TRANSMETE complementa sus actividades de formación e información con un servicio de asesoramiento. Este servicio tiene por objeto permitir un mayor acercamiento entre las PYMES usuarias y los proveedores de tecnología y aplicaciones telemáticas.

Los puntos de formación y los grupos de usuarios son, en resumen, oficinas de ayuda para todas las PYMES regionales, ofreciéndoles formación y asesoramiento.

Referencias

[1] Comunidad Europea , "Informes económicos de la Comisión Europea"

[2] Comunidad Europea , "Libro Blanco de la Competitividad, Crecimiento y Empleo"

[3] Informe Bangemar , "Servicios Telemáticos para las PYMES"

[4] Consorcio TRANSMETE , "Especificación de los requerimientos telemáticos de las PYMES"

[5] Consorcio TRANSMETE , "Metodología de implantación de la Telemática en las PYMES"

[6] Consorcio TRANSMETE , "Documentación de los seminarios TRANSMETE"

[7] Consorcio TRANSMETE , "Evaluación de resultados del proyecto TRANSMETE en las PYMES"

Nuevos Sistemas para el Control de Tráfico Urbano: El sistema PETRI

Autor: Fernando de la Huerta Fernández
IBERDROLA, Ingeniería y Consultoría
División: UITESA, Sistemas de Telecomunicación y Telecontrol
Avda. Burgos, 8B. 28036 - Madrid
Tfno. : 34 - 1 - 3833180. Fax.: 34 - 1 - 3833311
E-mail: fhf@uitesa.es

ABSTRACT

The aim of this paper is showing the system PETRI. It is an example of the collaboration between Private Enterprise, Spanish University and Local Governments. The system has been developed in Spain completely. PETRI gives a new dimenssion of traffic control and allows data transmission of traffic information using paging networks.

1. Introducción

PETRI, es un sistema multiprocesador de memoria distribuida de muy altas prestaciones que permite el control de tráfico en tiempo real y el envío de la información obtenida en dicho sistema a múltiples usuarios. Este sistema ha sido desarrollado íntegramente en España mediante la cooperación de la empresa **IBERDROLA**, Ingeniería y Consultoría (División: **UITESA**), el **D**epartament **E**stadística i **I**nvestigació **O**perativa, perteneciente a la **Universitat Politècnica de Catalunya**, y el **Ajuntament de Barcelona**. Este proyecto pertenece al programa europeo **ESPRIT**, estando integrado dentro de la iniciativa **PA**rallel **CO**mputing for **S**pain (**PACOS**) , PCI Project (EP - 9602).

El objetivo último de **PETRI** es el desarrollo, implementación y pruebas limitadas de un prototipo plenamente operativo de un *Sistema para el Control de Tráfico Urbano y Envío de Información en Tiempo Real* consistente en:

1.- Un software paralelo que realiza las funciones de:
- Recogida de los datos en tiempo real suministrados por los sensores de tráfico.
- Identificación de las condiciones actuales del tráfico sobre la red analizada.
- Representación de la evolución del estado en cortos periodos de tiempo.
- Tomar las acciones oportunas y elaborar los mensajes de tráfico necesarios para advertir a los usuarios de las condiciones del tráfico y su evolución.

2.- Un módulo de salida capaz de diseminar la información de tráfico a través de los medios de comunicación disponibles: Paneles de Mensajes Variables, Radio, etc. Para ello se ha desarrollado un sistema de bajo coste que permite hacer llegar la información a múltiples usuarios a través de las redes de radiomensajería. Este sistema se denomina **ARDID.**

El sistema **PETRI** está pensado para proporcionar información de tráfico en tiempo real a los centros de control de tráfico de los Ayuntamientos de las grandes ciudades permitiendo una gestión eficaz del mismo y la difusión del estado actual y previsto a un gran número de usuarios de una forma sencilla y económica con vistas a mejorar la circulación en las grandes ciudades.

2. Características técnicas del sistema

Las características técnicas del sistema **PETRI** son las siguientes:

- Utilización de **sistemas multiprocesadores** de memoria distribuida.
- Utilización de **redes neuronales** para simulación de estados de tráfico, entrenadas a partir de datos reales.
- Utilización de algoritmos de cálculo basados en la **paralelización** de sus principales componentes.
- Primer sistema de control de tráfico que permite proporcionar el estado **futuro** del tráfico en función de las acciones tomadas por el operador de tráfico.
- Desarrollo del sistema **ARDID**, que permite la **radiodifusión informática de datos** a través de las redes de radiomensajería **a bajo coste y en tiempo real.**
- Posibilidad de desarrollo de aplicaciones para usuarios específicos:
  - Simuladores de tráfico.
  - Control de tráfico en tiempo real.
  - Gestión de tráfico y redes urbanas.
  - Otros servicios públicos urbanos.
  - Información de tráfico y urbana en tiempo real para múltiples usuarios a bajo coste.
- Desarrollo de aplicaciones industriales:
  - Simulador y predictor de tráfico.

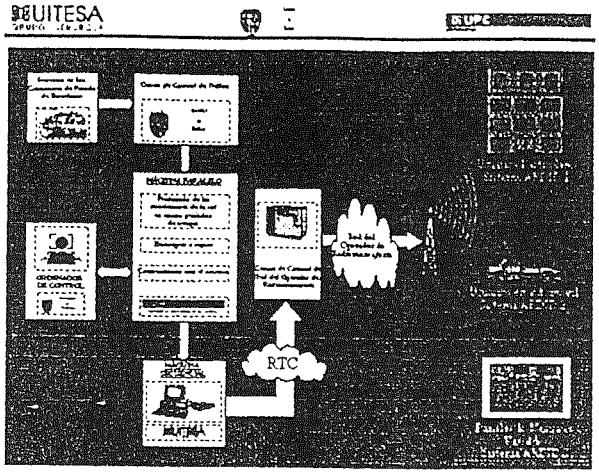


Figura 1.1.: Arquitectura del sistema PETRI

♦ Redes de distribución.

♦ Aplicaciones de control en tiempo real.

• Servicios de valor añadido proporcionados por el sistema **ARDID**.

La arquitectura del sistema **PETRI** queda reflejada en la figura 1.1.

3. Sistema de control de tráfico

Este sistema está formado por varios programas que funcionando de forma paralela se encargan de suministrar el estado actual y futuro de la red, permitiendo al operador ejecutar distintas funciones de control.

La estructura del **Sistema de Control de Tráfico**, como se puede ver en la figura 1.2. tiene tres componentes principales.

• **Módulo de Co-ordinación,** que se encarga de realizar las siguientes funciones: recibe información en tiempo real sobre las condiciones del tráfico a través del sistema de recogida de datos que toma los datos de los distintos sensores instalados en los cinturones de ronda.

Esta información es utilizada por los modelos de tráfico para describir las condiciones actuales de la red, y generar la previsión para un corto periodo de tiempo. Los datos actuales y futuros alimentan el Módulo de Reconocimiento Dibujado que suministra un análisis cualitativo del Sistema completo, e identifica y clasifica las áreas donde se están produciendo los problemas, o muy probablemente se van a producir en un futuro cercano o dentro en un horizonte de tiempo preestablecido, de acuerdo con los datos suministrados.

El **Módulo de Co-ordinación** esta formado por tres programas principales:

♦ Un programa que implementa un **algoritmo para la estimación dinámica** de los estados del tráfico en forma de matrices Origen-Destino dentro de la red.

♦ Un programa que implementa los algoritmos para estimación del estado de la red y la previsión en cortos periodos de tiempo.

♦ Un programa que identificará los problemas actuales y futuros dentro del horizonte de tiempo considerado.

• **Módulo de Diagnóstico,** es el segundo componente principal del prototipo que está formado por un conjunto de programas que generan un diagnóstico de la situación y proporcionan un conjunto de recomendaciones para evitar o prevenir los problemas identificados. Este módulo está formado por dos programas principales:

♦ El programa de diagnóstico y acciones recomendadas.

♦ El programa de paralelización de valoración y herramientas de evaluación de impacto.

• **Módulo de Diálogo,** es el tercer componente del prototipo, que reúne un conjunto de programas que forman un Sistema de Apoyo a la Decisión a medida para aplicaciones de control de tráfico. El objetivo de este módulo es completar los interfaces de usuario que harán de la aplicación un sistema amigable para el control de tráfico y potenciar su explotación comercial. Este módulo se encarga de presentar el estado de la red al operador donde las diferentes condiciones del estado del tráfico son mostradas una determinada escala de colores. El significado de la escala de colores se muestra en la esquina izquierda de la pantalla. El operador puede elegir entre distintas alternativas de representar el estado de la red, de acuerdo con la información producida por el simulador y seleccionar alguna de las recomendaciones presentadas por el sistema, visualizando el efecto que produce en la red la opción elegida.

4. Centro de control de comunicaciones

Está formado básicamente por un ordenador personal que de forma periódica conecta con el **Sistema de Control de Tráfico** para obtener la información sobre el estado actual y futuro de la

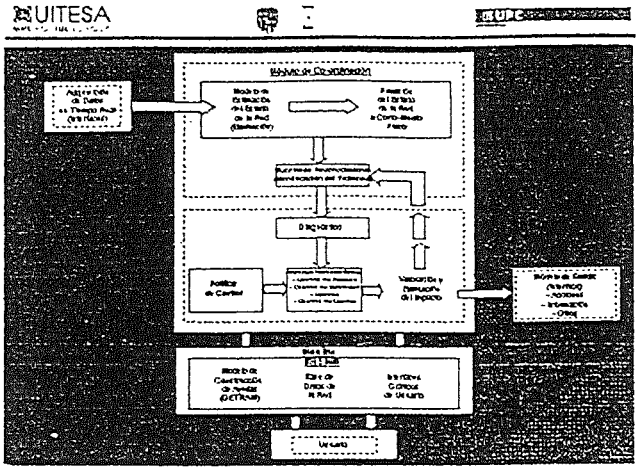


Figura 1.2 . PETRI, sistema de apoyo a la decisión para control de tráfico

red, y a continuación conectar con el **Operador de Radiomensajería** que se encarga de difundir la información vía radio haciéndola llegar a un gran número de usuarios.

El centro de control está formado por dos programas principalmente:

- **Programa de Comunicaciones,** que se encarga de establecer los distintos protocolos de comunicaciones entre las máquinas que pone en contacto. Así para comunicar el **Centro de Control de Tráfico** con el **Centro de Control de Comunicaciones** se utiliza el protocolo **Kermit** por su facilidad de uso con sistemas UNIX, mientras que para comunicar el **Centro de Control de Comunicaciones** con el **Centro de Control del Operador de Radiomensajería** se utiliza el protocolo TAP, estándar para conexión con los sistemas de radiomensajería.
- **Programa de Control,** que se encarga de establecer los horarios de conexión con el **Centro de Control de Tráfico.** Este programa permite establecer a voluntad del operador los distintos horarios de conexión con el **Centro de Control de Tráfico** con el fin de difundir la información suministrada por el sistema. El programa permite que las distintas conexiones se ejecuten de forma manual o automática en función de las necesidades del operador.

En la figura 1.7. podemos ver la arquitectura del Centro de Control de Comunicaciones.

5. Sistema ARDID

Este módulo se encarga de recibir la información de tráfico difundida vía radio a través de la red de radiomensajería y presentarla al usuario de una forma fácilmente comprensible para el usuario final. El sistema está formado por dos módulos fundamentales:

- **Receptor ARDID,** que se encarga de recibir los datos enviados a través del operador de radiomensajería. Su tamaño es muy reducido y se conecta directamente al puerto serie del ordenador personal. Se trata de un receptor de doble conversión en la banda de 170 MHz para FSK a una velocidad de 1.200 baudios. El receptor dispone de un amplificador de bajo nivel de ruido, que garantiza una alta sensibilidad (-120 dBm), y dos frecuencias intermedias de 10,7 MHz y 455 kHz para la obtención de una excelente selectividad y rechazo a la frecuencia imagen. El receptor se encuentra sintetizado mediante un PLL, de forma que se pueden utilizar los distintos operadores de radiomensajería existentes actualmente en España con el mismo receptor. El control del receptor se lleva a cabo mediante un microprocesador incorporado en el mismo.
- **Programas de Recepción y Presentación de Mensajes,** que son los encargados de presentar la información recibida vía radio en un formato comprensible por el usuario final. Este módulo por dos programas:
  - ✦ **Programa de Recepción,** que se encarga de estar permanentemente a la escucha y almacenar los datos recibidos vía radio a través de la puerta serie del ordenador personal.
  - ✦ **Programa de Presentación de Mensajes,** se encarga de presentar en pantalla los datos recibidos vía radio. El programa representa en pantalla la red objeto de estudio y sobre ella muestra la información suministrada por el **Centro de Control de Tráfico.** El modo de presentación puede verse en la figura 1.4.

En principio, el sistema **ARDID** ha sido desarrollado para presentar la información en un ordenador personal, pero el sistema es flexible pudiendo acomodarse a la necesidades del usuario final. Otras posibles opciones de presentación de la información son, paneles de mensajes variables móviles que llevarían incorporado el sistema y receptores situados a bordo de vehículos.

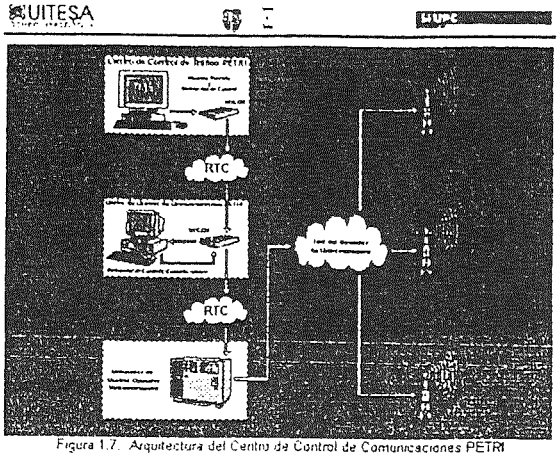


Figura 1.7. Arquitectura del Centro de Control de Comunicaciones PETRI

6. Prueba piloto del sistema

La prueba piloto del sistema completo, se ha desarrollado en varias fases. Una primera se hizo para comprobar el funcionamiento del sistema en modo secuencial, el cual suministra predicciones a 15 minutos vista. Una segunda prueba ha consistido en comprobar el funcionamiento del sistema en modo paralelo, el cual suministra predicciones a 5 minutos vista. Por último se va a realizar una última prueba con el sistema instalado en el Centro de Control de Tráfico del Ajuntament de Barcelona con el sistema completamente operativo y que servirá para verificar la bondad de la información suministrada por el sistema, tras lo cual el sistema quedará totalmente operativo.

7. Conclusiones

El principal objetivo de este proyecto es la migración de algoritmos de optimización y simulación probados para manejo numérico tomando los modelos de transporte necesarios, a un entorno de computación paralela con el fin de conseguir la velocidad necesaria para su utilización en aplicaciones en tiempo real, y en base a esta migración desarrollar un prototipo de sistema operativo en tiempo real para el control de tráfico y la difusión de la información obtenida.

Como resultado dos productos comerciales se han obtenido:

- Un sistema informático que constituye el corazón de una nueva generación de los Centros de Control de Tráfico cuyos usuarios serán los responsables del control de tráfico de las grandes ciudades.
- Un sistema de información de tráfico que puede ser utilizado por las corporaciones municipales y por cualquier operador que suministre servicios telemáticos al gran público, dando opción a incorporar otros servicios de valor añadido.

El sistema ha sido desarrollado mediante la cooperación entre la Empresa Privada, la Universidad y los Gobiernos Locales habiendo sido probado con éxito en un Proyecto Europeo.

8. Bibliografía

[1] DESPINA, DEmand Spreading through Pre-trip Information using ATT. Transport Telematics Project V2062, March 1995.

[2] PETRI, PArallel Environment for a real-time Traffic management and Information system. PArallel COmputing for Spain (PACOS) PCI Project (EP - 9602).

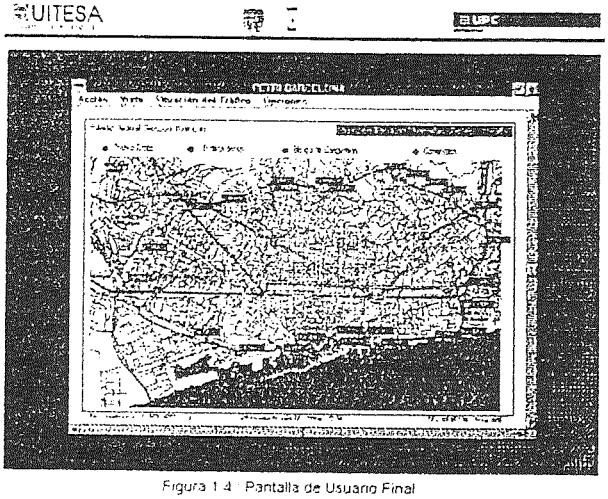


Figura 1.4 Pantalla de Usuario Final

Sistema de Información Telemático basado en Puntos Públicos de Acceso

DEPARTAMENTO DE ELECTRÓNICA Y SISTEMAS
FACULTAD DE INFORMÁTICA, UNIVERSIDAD DE A CORUÑA
CAMPUS DE ELVIÑA s/n, 15.071 A CORUÑA
Alberto Pan Bermúdez
Lucía Ardao Rodríguez
Fidel Cacheda Seijo
Angel Viña Castiñeiras
Correo electrónico: {alberto, lucia, fidel, avc}@gris.des.fi.udc.es

Abstract:

In this paper we describe the design of a information system based on telematic public access points. Its main features include: multimedia based information services, Internet connection, search capabilities and the use of smart cards as payment and access control means. A first implementation of the system is already in use. The conclusions obtained from its operation are also showed.

1. Introducción

Un Punto Público de Acceso, conocido por las siglas PPA, consiste típicamente en un ordenador ubicado en un recinto seguro y situado en un lugar público, que permite a los usuarios obtener un acceso electrónico instantáneo a información y servicios [1].

Es indudable la expansión que han sufrido este tipo de sistemas, apoyados en gran manera por compañías que desean ofrecer un nuevo tipo de servicios a todos sus clientes de una forma cómoda y sencilla. A través de los PPAs es posible hacer llegar un gran volumen de información a los distintos sectores de la población, haciendo uso de la tecnología informática para presentar diversos tipos de servicios de la forma más atractiva posible.

Del mismo modo, el estado actual de la tecnología en el campo del almacenamiento masivo facilita la incorporación de aplicaciones multimedia, lo que posibilita la presentación de los contenidos de una forma mucho más atractiva de cara al público y a través de una interfaz de usuario más amigable, obviando la necesidad de conocimientos informáticos por parte de los usuarios para utilizar de manera eficaz el sistema.

Desde la perspectiva empresarial, las ventajas que presentan los PPAs se basan principalmente en el bajo coste que implica la realización de un sistema de estas características y la mínima infraestructura requerida en contraposición, por ejemplo, al establecimiento de oficinas de información.

Todos estos factores han hecho que la expansión de los PPAs haya ido creciendo de forma paulatina hasta llegar a un punto en el cual se exigen contenidos más específicos, mayor dinamismo en la información mostrada y mayor variedad en los servicios ofrecidos, exigiendose gradualmente mayor calidad en los mismos.

Las nuevas exigencias llevan a la creación de Puntos Públicos de Acceso telemáticos, en los cuales se facilitan las actualizaciones remotas del contenido de cada punto de información y se posibilita la incorporación de innovadores servicios como pueden ser el audio o vídeo en vivo y el acceso al mundo de Internet [2].

Mediante la incorporación de la telemática a los PPAs se posibilita la actualización remota y automática de los contenidos de forma totalmente transparente para los usuarios. Así, el proceso de actualización es independiente de la ubicación geográfica concreta de cada uno de los múltiples PPAs, todos ellos conectados a algún servidor del sistema.

Sin embargo, la mayor ventaja se encuentra en la incorporación de Internet al mundo de los PPAs. Lo que ha dado en llamarse *Internet Kiosks*, ofrece a los usuarios una vía sencilla y cómoda de introducción en el mundo de Internet, permitiendo el acceso directo a todos los servicios que se ofrecen y puedan llegar a ofrecerse a través de ella. De esta forma, la información mostrada a través de un punto público de acceso no se encuentra limitada a la que sea posible almacenar localmente, sino que se permite acceder a una cantidad ilimitada de información: la ofrecida por medio del *World Wide Web*.

La incorporación de estos servicios presenta un incremento del coste debido al empleo de líneas de comunicación. En consecuencia, es necesario algún tipo de financiación para utilizar este tipo de servicios. La financiación puede derivarse de módulos publicísticos incorporados al contenido, o bien mediante el cobro a los usuarios por el uso de servicios utilizando algún tipo de medio de pago.

Este trabajo presenta una arquitectura para un sistema basado en Puntos Públicos de Acceso telemáticos. En la sección 2 se apuntan los principales requerimientos que debe de satisfacer un sistema de estas características. La sección 3 proporciona una descripción general del sistema. La sección 4 describe en profundidad los principales aspectos del diseño del mismo. La sección 5 muestra una implementación concreta y

ya en funcionamiento de la arquitectura propuesta. Finalmente, la sección 6 expone las principales conclusiones y valoraciones extraídas del estudio presentado.

2. Objetivos

En este punto se plantean los principales requerimientos que, a nuestro juicio, debe satisfacer un sistema telemático basado en PPAs. Estos son:

- Seguridad: El sistema debe ser protegido ante ataques originados por usuarios de los puntos de información (*ataques internos*) o tentativas de acceso no autorizadas desde Internet (*ataques externos*).
- Gestión remota: La arquitectura del sistema debe permitir recabar información sobre los puntos públicos de acceso mediante la monitorización de aspectos de rendimiento, utilización de servicios, líneas de comunicación, etc. Esta información incluye datos que pueden permitir tanto la realización automática de acciones remotas de ajuste, como la toma de decisiones de índole técnica y comercial.
- Fiabilidad: El sistema debe de garantizar unos tiempos de disponibilidad máximos en los puntos de acceso. Para ello, se debe de garantizar robustez ante fallos o caídas del sistema, incidencias de seguridad, etc. Además, en caso de producirse alguna contingencia se debe asegurar una actuación rápida que permita restaurar el correcto funcionamiento del punto de información. Esto pasa por que la arquitectura del sistema ofrezca las facilidades de gestión remota mencionadas en el punto anterior.
- Acceso condicional: Se debe poder limitar el acceso de los usuarios a determinados servicios en función de una serie de condiciones. Estas condiciones pueden referirse, por ejemplo, a la necesidad de realizar un pago electrónico o autentificar la identidad del usuario.
- Pago electrónico: El sistema de pago debe estar adaptado a las peculiares características de los puntos de acceso en aspectos como seguridad, soporte de pagos por tiempo o por evento, soporte de pagos de pequeño importe e incluso en ciertos casos debería proporcionar alguna forma de anonimato.
- Soporte para gran variedad de servicios: Los servicios ofertados por los puntos de acceso pueden ser de distintos tipos, en el presente trabajo se clasificarán los servicios según dos criterios. Un primer criterio se basa en el tipo de actualización requerido por los servicios, lo cual origina la siguiente clasificación:
  - *Servicios locales*: aquellos cuyo contenido se almacena localmente en los puntos de acceso y que, en general, no requieren ser actualizados.
  - *Servicios remotos*: aquellos servicios cuyo contenido se almacena remotamente en el servidor, lo que implica que las actualizaciones se realizan directamente en el mismo.
  - *Servicios locales actualizados remotamente*: aquellos que debido al gran volumen de información que tiene que ser transmitida durante su operación, no pueden tener sus contenidos almacenados en el servidor, ya que eso originaría unos tiempos de respuesta excesivamente altos, pero que, por otra parte, requieren actualizaciones periódicas.

  El segundo criterio divide a los servicios en base a sus características desde el punto de vista del acceso condicional, lo que da lugar a la siguiente clasificación:
  - Servicios de *acceso condicional:* se deben satisfacer ciertas condiciones para tener acceso al servicio. Pueden subdividirse en:
    - *Gratuitos*.
    - *Pagados*: Estos pueden requerir el pago por:
      - *Tiempo* de utilización.
      - *Evento* o acción realizada (e.g. pago por búsqueda realizada en un archivo documental).
- Escalabilidad: La arquitectura debe estar preparada para crecer sin degradar parámetros del rendimiento global del sistema. Estos parámetros incluyen aspectos como tiempos de respuesta, rapidez de las actualizaciones, etc.

3. Descripción general

Como se observa en la Fig. 1, la arquitectura presentada se basa en la existencia de uno o más servidores encargados de controlar y gestionar los múltiples puntos de información que componen el sistema, así como el sistema de seguridad necesario para evitar la realización de acciones no permitidas.

Como primer aspecto a destacar se encuentra la estructura de comunicaciones empleada por el sistema para mantener conectados a todos y cada uno de los puntos de información con su respectivo servidor. El medio empleado se basa en RDSI (Red Digital de Servicios Integrados), principalmente por los siguientes motivos: ancho de banda ofrecido, movilidad y coste.

Los servicios que típicamente se ofrecen en un punto de información telemático, si bien en su mayoría, no requieren una velocidad de conexión excesivamente elevada, sí requieren un mínimo que debe ser satisfecho con garantías. Por otra parte, debido a la forma de empleo de un punto de información, en donde existen variaciones en los tiempos de respuesta esperados dependiendo del servicio solicitado, e incluso de la presencia o ausencia de usuarios en determinadas situaciones, la línea de comunicación empleada debe ser capaz de soportar el tráfico variable generado, manteniendo los tiempos de respuesta esperados.

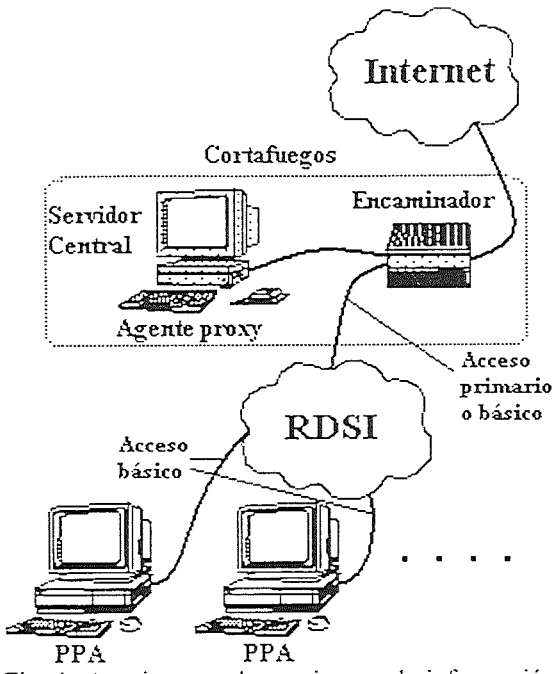


Fig. 1: Arquitectura de un sistema de información basado en PPAs.

Debido a la movilidad de los puntos de información, la línea de comunicación debe ser fácilmente instalable en cualquier punto de la zona de acción del sistema, que no necesariamente tiene porque ser local, hecho que no todas las líneas de comunicación soportan (e.g. líneas dedicadas).

Por último, el coste asociado a la línea contratada debe estar en función de la utilización de la misma, debido principalmente a la variación en el tráfico generado por los distintos puntos de información. En consecuencia, la utilización de la red RDSI para la conexión de los diferentes puntos públicos de información con un servidor central presenta las características expuestas, garantizando, por medio de un acceso básico, una velocidad de transferencia lo suficientemente elevada. Para evitar que el acceso al servidor se convierta en un cuello de botella es posible aumentar la capacidad en el mismo mediante un acceso primario.

En un sistema basado en redes abiertas, como es el caso de Internet, es necesario extremar el control ejercido tanto sobre los diferentes puntos de información, como sobre el servidor con el objetivo de evitar accesos indebidos.

Esto se realiza a través de un cortafuegos, en donde en sistemas de este tipo las precauciones a tener en cuenta se centran en evitar ataques tanto *externos* como *internos*.

Los ataques *externos* se producen por el empleo de una red abierta dentro de la estructura de comunicaciones, por lo que cualquier intruso puede intentar acceder al servidor central del sistema. Para prevenir este tipo de ataques están especialmente pensados los cortafuegos, controlando el acceso al sistema desde el exterior.

El caso de los ataques *internos* se deriva de aquellos usuarios que desde los puntos de acceso pertenecientes al sistema intenten corromper al mismo, con diferentes objetivos como, por ejemplo, el acceso a servicios evitando el pago. Contra este tipo de ataques además de los mecanismos particulares implementados directamente en cada punto público, se requiere la presencia de un cortafuegos que controle a través de un agente proxy el acceso al exterior del sistema, consiguiendo así mismo la monitorización de los accesos al exterior realizados desde todos y cada uno de los puntos públicos de acceso. En el apartado 4.1 se trata este aspecto en mayor profundidad.

Como se ha introducido anteriormente, la utilización de puntos de acceso telemáticos incorpora la existencia de servicios en los cuales es necesario condicionar el acceso, e incluso realizar pagos por la utilización de los mismos. Un ejemplo claro de esta situación es permitir el acceso a determinados servicios sólo a aquellos usuarios que cumplan determinada característica (poseer una tarjeta identificativa, conocer una contraseña, etc).

El hecho de tener que realizar un cobro al usuario de un punto de información hace que se incorporen a los mismos los medios de pago electrónicos a través de monederos electrónicos.

El uso de un punto de información puede generar gran cantidad de pagos de pequeño importe por la utilización de los servicios. Pensados especialmente para este tipo de situaciones se encuentran los monederos electrónicos utilizando tarjetas-chip pre-pagadas. En estos sistemas, el usuario realiza un desembolso inicial para la compra de la tarjeta, lo cual le permite posteriormente realizar diferentes gastos en los puntos de información que van decrementando el saldo de la tarjeta. Cuando el saldo se agota, la tarjeta-chip puede ser recargada.

La arquitectura permite el empleo de tarjetas-chip de clave compartida, con un coste asociado sensiblemente inferior a las tarjetas basadas en clave pública, obteniendo el mismo grado de seguridad en los puntos de acceso.

Mediante este medio de pago electrónico se ofrece soporte para los diferentes pagos de pequeños importes que se puedan generar durante la utilización de un punto de información, incorporando incluso ciertas formas de anonimato. Una explicación detallada del sistema empleado puede encontrarse en el apartado 4.2.

Uno de los principales aspectos que incorpora esta arquitectura se basa en la facilidad que ofrece para la actualización de los contenidos ofertados en el punto de información.

Las actualizaciones se pueden realizar de dos

formas principalmente. La primera de ellas consiste en la actualización directa en el servidor de los contenidos que serán accedidos a posteriori por los diferentes puntos de información. Esto determina que el servicio sea *remoto*. La segunda consiste en la actualización de los contenidos, a través del servidor, en cada uno de los puntos de información. Esto determina que el servicio sea *local actualizado remotamente*.

El primer tipo de actualización no requiere ningún tipo de arquitectura específica, simplemente hace uso de las posibilidades que ofrecen aquellos servicios alojados en el servidor y accedidos remotamente desde el cliente. Este tipo de actualización presenta la importante ventaja de su fácil implementación e inmediata repercusión en el contenido pero, en cambio, no es una opción válida para algunos servicios cuyo volumen de información es excesivamente elevado. En estos casos, si los documentos residiesen en el servidor, se obtendrían tiempos de respuesta inaceptables.

Para este último tipo de servicios, se ha diseñado una actualización remota y automática, que translada el control de la actualización a los puntos de información. Una vez que el servidor central ha determinado que se debe iniciar la actualización remota, el punto de información solicita al servidor de actualización únicamente aquellos datos necesarios para mantener el sistema actualizado desde la última actualización realizada.

En un sistema basado en puntos de información, la dispersión geográfica es un elemento importante e inherente al propio sistema. Por este motivo es de vital importancia poder determinar en todo momento la situación exacta de cada puesto de información.

La arquitectura propuesta permite la incorporación de técnicas de gestión de red al sistema, con el objetivo de poder detectar y corregir fallos del sistema, así como monitorizar la calidad de servicio ofrecida a los usuarios.

Mediante el empleo de técnicas de monitorización y muestreo exhaustivas, es posible determinar en todo momento el estado de cada punto de información. De esta forma, se permite determinar en todo momento aquellos puntos de información que se encuentran funcionando correctamente, y lo que es más importante, en algunos casos, es posible una recuperación automática y remota del equipo que presente problemas intentando de esta forma responder de la forma más rápida posible ante los fallos del sistema.

Así mismo, es posible visualizar en todo momento el funcionamiento de los componentes primordiales del punto de información, así como la carga de las líneas de comunicación empleadas. De esta forma se consigue controlar en todo momento el funcionamiento del sistema, pudiendo detectar posibles mejoras o reparaciones a realizar en base a datos obtenidos de las observaciones realizadas. El apartado 4.4 profundiza en este tema.

Por último, una de las mayores ventajas aportadas por esta tecnología se basa en su escalabilidad. Esta arquitectura no se encuentra limitada a un funcionamiento con un número fijo de puntos de información y un único servidor central.

Por el contrario, en caso de que el número de puntos públicos de acceso alcance un número elevado para ser gestionados a través de un único servidor central, existe la posibilidad de ampliar modularmente la arquitectura. En el apartado 4.5 se toca este aspecto en mayor detalle.

4. El sistema en detalle.

4.1 Seguridad

Las peculiares características de un sistema de información telemático basado en puntos públicos de acceso obligan a considerar la seguridad desde dos puntos de vista, derivados respectivamente de:

- La condición pública de los puntos de acceso al sistema.
- La conexión del sistema a una red abierta, típicamente Internet.

El primer aspecto consiste en intentar evitar acciones indebidas de los usuarios, tanto sobre el punto de acceso como sobre los servidores, que podrían causar fallos en el sistema o compromisos en la seguridad del mismo.

Esto conlleva, por una parte, dotar a los puntos públicos de acceso de medidas de seguridad física, tales como las provistas por equipos antivándalicos, una estructura externa resistente a ataques físicos, etc. Y por otra parte, es necesario crear una interfaz de usuario del punto de acceso que aune la facilidad de manejo con la limitación de aquellas operaciones que podrían resultar peligrosas (e.g. escribir en el dispositivo de almacenamiento, acceder a la interfaz del sistema operativo, etc.).

El segundo aspecto crucial para la seguridad del sistema consiste en el control de los accesos a los servidores desde máquinas externas conectadas a través de Internet, o cualquier otra red abierta. Para solucionar este problema se utiliza un cortafuegos, implementado mediante un encaminador y un agente proxy localizado en el servidor, actuando este último como "*bastion host*" [4].

El modo de funcionamiento del cortafuegos se basa en la configuración del encaminador y del agente proxy del servidor. El encaminador es configurado para no permitir la comunicación entre los puntos de acceso y la red abierta, del mismo

modo que tampoco se permiten conexiones desde el exterior a los puntos de acceso. Para permitir la comunicación entre los puntos de acceso y algún nodo de la red abierta se hace uso del agente proxy situado en el servidor que habilita o deniega el acceso en función de parámetros de autorización de los servicios.

En base a los requerimientos de seguridad de este sistema una solución estándar de cortafuegos como la aquí utilizada ofrece las garantías exigidas. Debe tenerse en cuenta que la seguridad ante ataques internos (ataques al servidor desde los puntos de acceso) es fuertemente incrementada por las ya comentadas restricciones de seguridad impuestas a los puntos públicos.

4.2 El Sistema de Pago y Control de Acceso

El Sistema de pago y control de acceso se basa en el uso de tarjetas-chip (también conocidas como tarjetas inteligentes) frente al empleo de tarjetas de crédito (debido a los altos costes por transacción inherentes a este método de pago). Esto permite dar solución a las necesidades de pago electrónico, al tiempo que proporciona una manera natural de implementar la estructura de autorizaciones necesaria para el acceso condicional a los servicios.

Desde el punto de vista del pago, las tarjetas-chip son pre-pagadas, lo que significa que las tarjetas son cargadas con un cierto importe y posteriormente el usuario puede utilizarlas para pagar los servicios del sistema. Este enfoque proporciona de forma automática varias ventajas: debido a que las tarjetas son pre-pagadas no es necesario realizar ninguna comprobación *on line* con una entidad bancaria o de otro tipo, lo que abarata considerablemente el coste de una transacción y hace al sistema adecuado para pagos de importes pequeños. Además, si las tarjetas no son distribuidas de manera personalizada, el sistema de pago es potencialmente anónimo lo que puede ser interesante en determinadas aplicaciones.

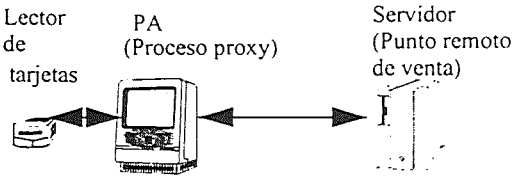
Las tarjetas-chip propuestas en la arquitectura se basan en la criptografía de clave compartida. Esto implica que la tarjeta-chip y el terminal de venta comparten una clave secreta que utilizan para autentificarse mútuamente mediante un protocolo basado en el uso de números aleatorios. La tecnología de tarjetas-chip de clave compartida es mucho más barata que su competidora, la tecnología de tarjetas-chip de clave pública. La ventaja de la criptografía de clave pública en los sistemas de monedero electrónico de este tipo, consiste en permitir a los terminales de venta contener únicamente claves públicas, lo que hace inútil vencer su seguridad con el propósito de cometer un fraude al sistema [5]. El problema de la seguridad de los terminales de venta es especialmente grave en el caso de los PPAs, donde dicho terminal está expuesto al público. Sin embargo, el uso de tarjetas-chip de clave pública podría encarecer en demasía el sistema de pago.

La arquitectura que presentamos soluciona el problema de seguridad de los puntos de venta con un incremento inapreciable del coste, haciendo uso de las líneas de comunicación ya existentes que conectan servidores y puntos de acceso. Básicamente, se transfiere la "inteligencia" del terminal de venta al servidor, de manera que el terminal del punto de acceso se limita a contener un proceso "*proxy*" que actúa como intermediario en las comunicaciones entre la tarjeta-chip y el servidor. De esta manera, la clave compartida reside físicamente en el servidor, con lo que no es vulnerable a ataques por parte de los usuarios. La clave no puede ser interceptada por intrusos en las comunicaciones entre tarjeta y servidor debido al uso del sencillo protocolo de autentificación basado en la transmisión de números aleatorios típico de clave compartida: un extremo que desee autentificar al otro genera un número aleatorio y se lo envía al otro extremo, quien debe de devolverlo encriptado con la clave compartida. El extremo emisor encripta a su vez el número aleatorio con la clave y comprueba que ambos resultados son idénticos.

En la Fig. 2 se muestra un sencillo esquema del sistema, en el que se muestra la función de intermediario que efectua el PA.

Con este sencillo esquema se consiguen, aparte del incremento de seguridad ya mencionado, otras características convenientes tales como: una mejor monitorización del sistema de pago (que ahora puede ser realizada desde los servidores), compartición de recursos del sistema de pago por parte de todos los puntos de acceso (e.g. si se usa un sistema de pago general, basta con contratar con la entidad propietaria del sistema de pago, un terminal de venta por servidor y no uno por punto de acceso), etc.

Por otra parte, el coste temporal extra asociado a las comunicaciones con el servidor es casi despreciable en nuestro caso debido fundamentalmente a los siguientes factores: la capacidad que deben tener las líneas de



ig. 2 : Estructura del sistema de pago

comunicaciones para dar soporte a los servicios telemáticos es ampliamente superior a la requerida por el sistema de pago (los datos intercambiados en cada transacción no sobrepasan los 50 bytes) y a que la velocidad de la comunicación entre el ordenador del PPA y el lector de tarjetas-chip (que obviamente también se tendría que realizar si la transacción fuese totalmente local) se produce a velocidades mucho más bajas (e.g. 9600 bps) que la comunicación con el servidor, con lo que es ahí donde se situará el cuello de botella del tiempo de respuesta de una transacción.

El uso de tarjetas-chip de uso exclusivo en el sistema puede resultar poco flexible para los usuarios. Hay dos maneras de solucionarlo: la primera de ellas consiste en utilizar un sistema de monedero electrónico de uso general (e.g VISA Cash, EURO-6000, etc). Esto requiere agregar a los servidores un módulo de punto de venta de estos sistemas. En el servidor actuaría otro proceso "*proxy*" entre el terminal de venta y la tarjeta-chip. La segunda manera se basa en utilizar tarjetas de crédito como medio para la carga de las tarjetas-chip. Este método consistiría en disponer de puntos automáticos de adquisición y recarga de tarjetas-chip específicas para los puntos públicos, donde la compra y recarga de las mismas se realizase de manera automática y con cargo a una tarjeta de crédito. El primer método mantiene todas las ventajas del sistema inicial, pero puede requerir algún acuerdo especial con el proveedor del método utilizado (e.g. VISA). El segundo impide que los usuarios puedan tener garantía de anonimato e incrementa el coste debido al establecimiento de puntos donde se aceptan tarjetas de crédito.

En cuanto al sistema de acceso condicional, este también se beneficia de las ventajas de la arquitectura debido a que toda la información referente a la seguridad y a las condiciones de acceso reside en el servidor, de manera que una intrusión en los puntos de acceso es inútil, del mismo modo que se facilita la actualización y monitorización de las reglas de acceso al sistema.

La complejidad del sistema de acceso puede variar desde la necesidad de poseer una tarjeta-chip autorizada para acceder a algún servicio, hasta construir una compleja estructura de autorizaciones y condiciones de acceso basándose en la información de autorización contenida en la tarjeta-chip, el sistema puede decidir si proporciona o no el acceso, así como mantener o modificar la información de autorización de la tarjeta.

4.3 Actualización de contenidos

La idea básica para la actualización de contenidos en el sistema consiste en el empleo de un tipo de procesos denominados "*servidores de actualización*", encargados de detectar la existencia de nuevos documentos en el servidor. Estos envían un aviso a otro tipo de procesos que se ejecutan en los puntos públicos de acceso, llamados "*clientes de actualización*". El principal aspecto se centra en que el peso de las decisiones recae en el proceso cliente, quedando los procesos servidores como meros "mensajeros", encargados únicamente de advertir a los clientes de los cambios producidos y proporcionar los datos que precisan para realizar la actualización. Las razones de este enfoque son varias y se irán mostrando en las líneas que siguen.

El hecho de realizar una actualización involucra diferentes actividades según el tipo de servicio a actualizar, *remoto* o *local actualizado remotamente*. En el primer caso, una vez realizada la actualización en el servidor, únicamente es necesario eliminar la información de la cache de la aplicación cliente, de manera que los contenidos del servicio tengan que obtenerse directamente del servidor. Este tipo de actualizaciones, normalmente, puede realizarse una vez es recibida la señal procedente del "servidor de actualizaciones", ya que no afectan de forma sensible a la calidad de servicio ofrecida al usuario. Sin embargo, la actualización de los servicios *locales actualizados remotamente*, debe realizarse de forma diferente debido a que se necesita un uso intensivo de las líneas de comunicación y de un sistema de almacenamiento local, lo que podría perjudicar el correcto funcionamiento del punto de acceso. Por este motivo, el "cliente de actualización" debe ser autónomo respecto al "servidor de actualización", lo que implica que debe de ser capaz de obtener del servidor los datos necesarios para la actualización (tales como el servicio afectado, el volumen de datos a actualizar, etc.), almacenarlos y realizar la actualización en el momento adecuado, que, por ejemplo, si la rapidez de la actualización no es un factor crítico, puede ser el horario nocturno, o en caso contrario, puede emplear los intervalos en que el punto de acceso no esté siendo utilizado por ningún usuario. En cualquier caso, el cliente debe de ser capaz de almacenar el estado de la actualización para proseguirla posteriormente. Por otra parte, poner la "inteligencia" del sistema en el cliente permite descargar al servidor del control de las actualizaciones de todos los puntos de acceso. Así, el papel del servidor se reduce a proporcionar un mensaje con información sobre la actualización a todos los clientes y, en caso de realizarse algún tipo de monitorización del proceso, debe recibir mensajes indicativos del estado de la actualización en el cliente. En este sentido, el sistema funciona de manera similar a los protocolos de "*Webcasting*" o tecnologías "push" actualmente en auge en Internet.

La idea general expuesta debe de ser ajustada a cada caso particular para obtener un protocolo plenamente adaptado a las necesidades de cada sistema. A continuación se esboza brevemente un protocolo concreto que, aun siendo sencillo, será adecuado para la mayoría de sistemas.

Para las actualizaciones en servicios *locales actualizados remotamente*, el proceso se inicia cuando el servidor de actualizaciones envía a cada PPA un mensaje en el que se indica: el servidor que envía el mensaje, el servicio al que se refiere la actualización, un identificador de la actualización, el número de documentos nuevos, la longitud total de los mismos, una fecha tope de terminación y una lista (opcional) de algunos documentos y sus longitudes. La lista es conveniente para servicios multimedia porque, frecuentemente, el tamaño de los documentos es muy variable. En caso de no utilizarse esta lista, el cliente podría comenzar la actualización de un documento de tamaño elevado y no disponer de suficiente tiempo para terminarla antes de que el siguiente usuario acceda al punto de acceso, causándole importantes restricciones en la calidad del servicio. De esta manera, listando los documentos de mayor volumen el cliente puede realizar una mejor planificación (e.g. puede dejar los documentos de mayor longitud para las horas nocturnas). Adicionalmente, la lista puede ser utilizada para asignar prioridades a los documentos.

El cliente puede enviar al servidor dos tipos de mensajes que permiten a este realizar una monitorización completa del proceso de actualización. El primero es un mensaje indicando el estado de la actualización, que puede ser enviado a intervalos predeterminados. En este mensaje, cada PPA especifica cuantos documentos han sido ya actualizados, su longitud, así como cuales de los documentos de la lista han sido tratados.

El segundo es un mensaje de error que indica al servidor que se ha surgido un problema intentando actualizar un documento.

El caso de actualizaciones de servicios *remotos* es mucho más sencillo ya que sólo se requiere un mensaje de aviso del servidor al cliente y otro de respuesta de la acción en sentido inverso.

4.4 Gestión remota

Es sencillo aplicar en la arquitectura técnicas de monitorización remota que permitan detectar de manera rápida anomalías de funcionamiento, caídas en la calidad del servicio, etc. La idea central consiste en que un servidor se gestiona a sí mismo y a los clientes que dependen de él.

La complejidad de las técnicas empleadas puede variar ampliamente según el nivel de gestión requerida. En el caso más sencillo bastaría con que un proceso en el servidor efectue un test de conectividad sobre sus clientes para detectar posibles caídas de los puntos de acceso o de las líneas de comunicaciones. Sin embargo, se requerirán, normalmente, controles e información más rigurosa que permitan recoger, por ejemplo, parámetros de comunicaciones (e.g % de paquetes perdidos en las comunicaciones con cada punto de acceso, tiempos de respuesta, etc.), parámetros de distribución de carga (e.g. carga de CPU, picos de utilización, etc.), etc. Este tipo de parámetros pueden ser recogidos y procesados utilizando técnicas estándar de Gestión de Red. En concreto, el uso de alguna herramienta de Gestión de Red basada en SNMP debería de ser suficiente para detectar a tiempo anomalías y caídas en la calidad de servicio de un punto de acceso determinado, a la vez que puede proporcionar información extremadamente útil para determinar la causa del problema. El enfoque a utilizar sería el clásico, con un agente SNMP en cada punto de acceso y el gestor localizado en el servidor. Ambos manejarían una MIB específica orientada a la gestión de los PPA. Los agentes recolectan la información especificada y lanzan alarmas ("*traps*") si se sobrepasan los niveles permitidos en ciertos parámetros. Otra posible opción sería utilizar agentes RMON ("*Remote MONitoring*") para la recolección de estadísticas de diversos parámetros del sistema [3].

Si bien, por razones de limitación en la extensión del presente artículo, no se puede tratar en detalle la estructura de una MIB para estos sistemas, sí se pueden apuntar algunos parámetros específicos a modo de ejemplo: distribución del tiempo de operación entre los servicios, utilización y picos de uso de recursos del PPA por servicio, número de usuarios para cada servicio y, en general, parámetros que no sólo permitan monitorizar el funcionamiento adecuado de los PPAs, sino que permitan adecuar los requerimientos concretos para cada servicio en cada PPA, proporcionando a la vez información para medir la aceptación de los servicios entre los usuarios.

Una de las principales ventajas de la gestión de red aplicado a sistemas de este tipo es la capacidad para realizar acciones de forma automática con un mínimo tiempo de respuesta. El modo de funcionamiento consiste en el empleo de acciones de recuperación asociados a determinados tipos de alarmas activadas por los clientes. Esto permite la detección de situaciones de compromiso de la seguridad o posibles errores de funcionamiento (e.g. es posible detectar mediante una alarma si una aplicación cliente necesaria en el PPA ha dejado de estar activa, tomándose la acción de reiniciar remotamente el punto público de acceso para restaurar las anomalías).

4.5 Protocolo de respaldo

La posibilidad de fallos o excesos de carga en servidores puede ser tratada de manera eficaz en sistemas a gran escala mediante una estructura basada en la agrupación por zonas de los PPAs, designando un servidor primario por zona y uno o varios servidores secundarios (que a su vez, pueden ser primarios para otras zonas) para propósitos de respaldo, siguiendo un esquema similar al planteado por DNS (Domain Name System). Ocurrirá normalmente que los contenidos ofertados en cada zona sean diferentes (e.g. servicios de información local), lo que implica que los servidores secundarios deben actualizar sus contenidos de manera automática, pudiendo utilizar para ello un protocolo muy similar al mostrado en el punto 4.3, utilizado para las actualizaciones en servicios *locales actualizados remotamente*. De esta manera, si un servidor primario no está accesible en un momento determinados, sus clientes pueden dirigirse a alguno de los servidores secundarios temporalmente mientras no se recupera la conectividad con el servidor primario. Se corre el riesgo de que el servidor secundario no haya tenido tiempo de actualizarse totalmente, pero se evita la ausencia de servicio al usuario. De la misma forma, un servidor primario sobrecargado puede ordenar a sus clientes dirigir sus peticiones relacionadas con un determinado servicio a un servidor secundario que le haya informado de la finalización de la actualización de dicho servicio.

Esta estructura puede generalizarse organizando los servidores del sistema de modo jerárquico, de tal forma que los servidores raíz se encargan de proporcionar los contenidos generales del sistema y los servidores de niveles inferiores introducen contenidos y servicios específicos para zonas o subzonas específicas. Esta estructura es, en muchos aspectos, similar a la conocida jerarquía DNS y, al igual que ella, proporciona fiabilidad basada en servidores secundarios, escalabilidad y extensibilidad automáticas, evita la replicación de contenidos (excepto para propósitos de apoyo) y permite una gestión independiente para cada zona.

Los protocolos que soportan esta arquitectura, así como otras consideraciones que deben de realizarse sobre ella, serán tratados en una futura comunicación.

5. Una implementación concreta: Proyecto KIM

En este apartado se muestra una implementación concreta y ya en funcionamiento de la arquitectura presentada: el proyecto KIM (Kiosco Interactivo Multimedia) realizado para el Grupo Voz de comunicaciones.

La primera versión fue presentada en noviembre de 1.996 y ha estado en funcionamiento desde entonces. Es esta versión la que se tomará como referencia para las líneas que siguen.

5.1 Los servicios del KIM

Con el KIM se pretende proporcionar al ciudadano medio un sistema de *puntos de información* de acceso público, que sea el soporte para una serie de servicios novedosos. La característica fundamental de este sistema frente a los sistemas de puntos de acceso habituales en su zona de operación, es su condición de *telemático*, es decir su capacidad de proporcionar servicios de naturaleza telemática como puede ser la consulta de bases de datos documentales o permitir el acceso a Internet.

A continuación se presentan los servicios que ofrece el sistema clasificándolos según los criterios que fueron presentados en el apartado 2.

Dentro de la clasificación de servicios *locales* se ofrece un servicio de información corporativa en formato multimedia (audio, video, rotativas en 3D) del GrupoVoz de comunicaciones. Atendiendo al otro criterio de clasificación, se trata de un servicio de *acceso libre*.

Dentro de los servicios *locales actualizados remotamente* se engloba el servicio de Hemeroteca on-line. Este servicio permite la consulta de los periódicos del último año del archivo hemerográfico del Grupo Voz, ofreciendo dos utilidades principales:

- Consulta de periódicos: Permite al usuario consultar el periódico de una fecha determinada.
- Búsquedas: Permite realizar búsquedas en el archivo hemerográfico. Las búsquedas pueden realizarse entre los periódicos de un período determinado (e.g el último trimestre o el último año) y se admiten búsquedas avanzadas (e.g buscar las palabras similares a una dada, buscar las palabras con la misma raiz léxica, etc.)

Los periódicos se visualizan en un formato que es una reproducción exacta del periódico impreso, utilizando para ello el formato *Adobe PDF (Portable Document Format)*. El servicio es actualizado diariamente con el fin de que el archivo hemerográfico contenga incluso el periódico del día. El servicio no es *remoto* (i.e. no se almacenan los documentos directamente en el servidor) debido a que el elevado tamaño de los documentos y los índices utilizados en las búsquedas, convertirían al servicio en extremadamente lento.

Desde el punto de vista del otro criterio de clasificación considerado, el servicio es de *acceso condicional gratuito*. Para el acceso a este servicio es necesario introducir en el lector una tarjeta chip autorizada, si bien actualmente no se sustrae dinero de ella durante la operación. Si el usuario retira la tarjeta mientras se está utilizando este servicio, la consulta se interrumpe automáticamente.

Dentro de los servicios *remotos* englobamos:

- *Conexión a Internet de pago*: Este es un servicio de *acceso condicional de pago por tiempo*. Desde el punto de acceso es posible navegar por Internet y conectarse a cualquier servidor WWW externo. Sin embargo, para ello el usuario tendrá que disponer de una tarjeta chip autorizada que posea saldo suficiente. Cuando el usuario esté conectado a Internet, se sustraerá dinero de esta tarjeta a intervalos definidos. Si el usuario retira la tarjeta o se agota el dinero de ésta, el acceso queda automáticamente interrumpido, permitiéndose exclusivamente el acceso a los servidores propios del sistema.

- *Radio en Internet*: Se ofrece un menú 'a la carta' con programas de radio así como radio 'en vivo' en Internet de la emisora del GrupoVoz. Junto con el nombre de los programas ofertados, se indica la fecha y hora de emisión así como el tipo de programa. Este servicio es de *acceso libre*.

- *Inserción de anuncios por palabras y suscripciones al periódico convencional:* El usuario del punto de acceso telemático tiene la posibilidad de contratar un anuncio por palabras para la edición impresa de los periódicos del GrupoVoz y de contratar una suscripción a cualquiera de las ediciones de los mismos. Ambos servicios son de *acceso condicional pagado por eventos*. En el caso de los anuncios por palabras, se utiliza la tarjeta chip autorizada para pagar el anuncio contratado. En el caso de las suscripciones, el pago se realiza utilizando una cuenta bancaria.

5.2 La implementación de la arquitectura en el KIM

En este apartado revisaremos los puntos presentados en el apartado 4 indicando que opciones de las ofrecidas por la arquitectura se han tomado para este sistema concreto, además de cual es la estructura de comunicaciones:

- Estructura de comunicaciones: El sistema lo componen un servidor central y un conjunto de clientes dispersos geográficamente que se comunican con este servidor central utilizando líneas RDSI de acceso básico. El uso de un sólo servidor es suficiente por el momento debido a que el número de clientes es reducido. Los clientes acceden al servidor a través de un router que está conectado con el servidor mediante una Ethernet a 10 Mbps. La salida a Internet se realiza a través del router mediante una línea Frame Relay a 64 Kbps.

- Seguridad: Se ha utilizado un cortafuegos combinado con técnicas de seguridad en los puntos públicos, tal y como se ha comentado en el apartado 4.1

- El sistema de pago y de control de acceso: El sistema de pago empleado es un sistema propietario de tarjetas chip pre-pagadas. Las tarjetas son anónimas ya que no están ligadas a ningún usuario concreto. No ha sido necesaria la integración con sistemas de pago de uso general, ya que el sistema fue planteado para su uso en superficies comerciales, por lo que es sencillo disponer puntos de venta y recarga de las tarjetas cerca de cada punto de acceso

- Actualización de contenidos: Se ha implementado la estructura de actualizaciones presentada en el apartado 4.3 para el servicio de *Consulta del archivo hemerográfico*. El protocolo concreto se ha simplificado ligeramente con respecto al del punto 4.3 para aprovechar que el servicio se actualiza a intervalos diarios.

- Monitorización, gestión remota y protocolos de backup: Debido a que este tipo de sistema requiere una gestión remota sencilla sólamente se han implementado los aspectos que chequean la conectividad de los puntos de acceso y que se encuentran lanzadas todas la aplicaciones necesarias para el correcto funcionamiento del punto de acceso. Adicionalmente, se ha implementado un mecanismo simple que muestra mensajes en el servidor describiendo que acción se está realizando en cada momento en cada punto de acceso (e.g. que servicio se está utilizando, como se accede a él, etc.). Actualmente se están realizando pruebas con protocolos más complejos basados en el uso de SNMP y agentes RMON.

5.3 Valoración del funcionamiento del KIM

En este apartado se indica como se ha comportado la arquitectura en la implementación particular que aquí se presenta.

- Seguridad: Se han producido ataques simples a la seguridad de los puntos públicos, tales como intentos de acceso al sistema operativo de los ordenadores clientes. Las medidas de seguridad tomadas en los puntos de acceso fueron suficientes

para evitarlos e informar de ellos.

- El sistema de pago y de control de acceso: Los costes por transacción se han mostrado como casi nulos, lo que ha confirmado que el sistema de pago es válido para pagos de pequeño importe. Si bien las restricciones de acceso condicional en el sistema han sido sencillas, y no permiten sacar conclusiones generales, el sistema de tarjetas-chip también se ha comportado bien en este aspecto, permitiendo realizar una buena monitorización de las reglas de acceso. También se ha comprobado que el uso de tarjetas-chip exclusivas del sistema no es un problema serio si pueden disponerse puntos de venta y recarga de las mismas cerca de cada punto de acceso.
- Actualización de contenidos: Las primeras pruebas con el servicio de *Consulta del archivo hemerográfico* mostraron la inviabilidad de plantear el servicio como *remoto*. Con el enfoque de actualizaciones, ha podido realizarse el servicio satisfaciendo adecuadamente los requisitos de velocidad. Pudo observarse al realizar las actualizaciones remotas que debido al elevado volumen de información a transmitir, las actualizaciones podían ser, en ocasiones, algo lentas. Sin embargo, esto no constituyó un problema en este caso concreto debido a que las actualizaciones podían realizarse en las horas nocturnas. En casos en que se requieran actualizaciones más rápidas, puede requerirse el uso de formatos que generen documentos de menor tamaño que el *Adobe PDF*.
- Monitorización, gestión remota y protocolos de backup: Los procesos implementados de monitorización y gestión remota, a pesar de su sencillez, se han mostrado como suficientes para este sistema, debido a que el número de puntos de acceso ha sido bajo. Entre las incidencias que ha permitido detectar han estado: caída de la línea entre un punto de aceso y el servidor y pérdida de corriente eléctrica en un punto de acceso. También se detectaron casos en que se produjeron fallos de alguna aplicación del punto de acceso (e.g el navegador de Internet). Esto provocó que el proceso de gestión remota reiniciase el punto de acceso donde se había producido el fallo, restaurando así su funcionamiento normal. A pesar del buen funcionamiento en líneas generales, para sistemas con un número de puntos de acceso elevado, parece conveniente generalizar las técnicas empleadas mediante el uso de herramientas más formales de gestión de red. Las primeras experiencias con SNMP y agentes RMON parecen confirmar esta apreciación.

6. Discusión de la arquitectura

Una vez expuestos los principios de la arquitectura y las conclusiones extraídas de la operación del primer sistema construido de acuerdo a ella, se procede a la valoración de la misma en base a los objetivos planteados inicialmente.

Como primera característica cabe destacar la capacidad para soportar un amplio espectro de tipos de servicios. El soporte para servicios *locales* es inmediato por la naturaleza del punto de información, mientras que el soporte para aquellos servicios *remotos* se posibilita por la conexión con los servidores. La parte más novedosa se centra en el soporte de servicios *locales pero actualizados remotamente* desde los servidores.

Desde el punto de vista del *acceso condicional* a los servicios, la incorporación al punto público de acceso de un mecanismo de monedero electrónico basado en tarjetas-chip ofrece soporte para prácticamente cualquier aplicación de control de acceso, al tiempo que proporciona una manera segura y eficiente de realizar el pago por los servicios. De todas maneras, si el *acceso condicional* se implementa en base a otro método no existe ningún inconveniente en la arquitectura que impida la incorporación del mismo al punto público de información.

También cabe destacar, respecto al pago electrónico, la posibilidad de incluir mecanismos de pago de uso general, frente al uso de un método propietario de uso exclusivo en el sistema. Esto permitiría incorporar una mayor flexibilidad en los pagos de los usuarios, al poder emplear los mecanismos de pago electrónico que utiliza para otros propósitos.

En cuanto a la seguridad incorporada por el sistema, el esquema basado en un cortafuegos estándar que se ha esbozado, unido a las precauciones tomadas en la implementación del punto de información para permitir al usuario realizar únicamente aquellas operaciones estrictamente necesarias para un empleo cómodo y eficiente de los servicios, parecen ser más que suficientes para sistemas de este tipo, donde normalmente el incentivo de fraude para posibles atacantes sea bajo.

Otro aspecto considerado en la arquitectura, que posibilita la detección y corrección de fallos se basa en la monitorización y gestión remota de los diferentes puntos de información realizada a través de técnicas de gestión de red. Esto enlaza directamente con las posibilidades de recuperación automática que ofrece el sistema ante algunos fallos en los puntos públicos.

También es importante la capacidad para detectar posibles zonas de mejoras en el sistema en función de los valores obtenidos de la monitorización y que abarcan tanto aspectos relacionados con la

utilización de las líneas de comunicación como con los diferentes elementos que componen el punto de información, así como con el grado de utilización de los diferentes servicios.

Por último destacar el alto grado de escalabilidad que presenta el sistema, de tal modo que su crecimiento no influencia sobre aspectos como los tiempos de respuesta ofrecidos a los usuarios en determinados servicios, tiempo invertido en las actualizaciones remotas, etc.

Esto es debido a la utilización del modelo jerárquico y organizado en zonas esbozado en el apartado 4.5, el cual permite el crecimiento gradual del sistema sin degradación en los parámetros de calidad de servicio observados por el usuario, al tiempo que permite una distribución natural entre los servidores de la carga asociada a las actualizaciones de los servicios.

De esta forma, se concluye que aún existiendo puntos mejorables, la arquitectura propuesta da soporte a los objetivos principales reseñados inicialmente.

Referencias

[1] Morris, G., Sanders, T., Gilman, A., Stephen, A. y Smith, S. "Kiosks: A technological Overview". LA-UR-95-1672. CIC-3, Los Alamos National Laboratory. Los Alamos, NM 87545. January, 1995.

[2] Shah, R. "Suggestions for Information Kiosk System using the World Wide Web". *The World Wide Web information Kiosks Special Interest Group*, April 1994.

[3] Rose, T. "The simple book. An introduction to networking management", 2nd ed. Prentice Hall 1997.

[4] Hare, C. and Siyan K. "Internet Firewalls and Network Security", 2nd ed. New Riders 1.996.

[5] Chaum, David. "Prepaid smart card techniques: A brief introduction and comparison". *Technical report DIGICASH* (1994).

El Controlador Domótico Maior-Domo

A. RUIZ DE OLANO
DPTO. ELECTRÓNICA Y COMUNICACIONES
IKERLAN
APDO. 146, E-20500 MONDRAGÓN
e-mail: arolano@ikerlan.es

Abstract

A new concept of home is being promoted by taking benefit of the fast evolving electronics and communication technologies, the Intelligent Home. Through this concept it is offered to the home dwellers an increased amount of security, comfort and savings. The Domotic Manager Maior-Domo, a product from Fagor Electrodomésticos, has been designed to make the mentioned benefits possible.

1. INTRODUCCIÓN

Desde 1990 Fagor Electrodomésticos está desarrollando, con la colaboración de Ikerlan, una serie de actividades encaminadas a la obtención de productos que posibiliten la entrada en los aparatos de automatización del hogar de varios de los avances tecnológicos existentes en el terreno de las Tecnologías de la Información y concretamente de las Comunicaciones.

Estas actividades de I+D abarcan varias líneas tecnológicas: el diseño de sistemas electrónicos analógicos y digitales, el desarrollo de los sistemas electrónicos de procesamiento que sean capaces de coordinar las nuevas funciones requeridas de los dispositivos físicos en los aparatos electrodomésticos y de ejecutar los algoritmos para la actuación y/o supervisión que les son propios en sus tareas de automatización, lo que significa el desarrollo de aplicaciones Software progresivamente más sofisticadas.

La automatización del hogar que se ofrece en el contexto del concepto 'Hogar Inteligente' (Domótica) requiere unas prestaciones adicionales y nuevas de los aparatos electrodomésticos, que en modo importante están relacionadas con la comunicación de datos de modo fiable entre los diversos aparatos, a los que llamaremos 'Domóticos' cuando las incorporan.

En este trabajo se presentan unas ideas generales sobre los aspectos más relevantes de la domótica desde un enfoque técnico y, a continuación, uno de los productos fundamentales para un Sistema Domótico, formado por un conjunto de electrodomésticos intercambiándose datos a través de la propia red de alimentación eléctrica (powerline). Este producto, diseñado y fabricado en este país, es el Controlador Domótico Maior-Domo Fagor.

2. DESCRIPCIÓN GENERAL DE LA DOMÓTICA

En los hogares convencionales el usuario de una instalación eléctrica conecta a ella sus aparatos que tienen, por lo general, un funcionamiento independiente y sin posibilidad de establecer supervisión o coordinación sobre éstos. La única interrelación entre ellos la realiza ocasionalmente el usuario sin poder delegar esta función en un sistema que lo lleve a cabo continuamente.

El abaratamiento progresivo de los componentes electrónicos que permiten la introducción de 'inteligencia' en ellos ha posibilitado la puesta en el mercado competitivo de productos que gestionan la energía y el confort. Los fabricantes de dispositivos para el hogar han desarrollado aparatos que permiten programar la calefacción, regular la temperatura ambiental, gestionar la energía en función de la ocupación de los locales, u otros criterios de racionalización teniendo en cuenta factores exteriores como el nivel de luminosidad, las diversas tarifas según la hora diaria, la temperatura exterior, la velocidad del viento... A todas estas instalaciones singulares se han sumado nuevas funciones de vigilancia y de comunicación: alarmas técnicas, alarma intrusión, televigilancia...

Allí donde se desea elevar el nivel de automatización, la instalación se hace más compleja, la cantidad de cableado adicional y los conocimientos requeridos para configurar, instalar, ampliar y mantener frenan el proceso de aceptación de estas mejoras en los conceptos mencionados.

Es conveniente que la tecnología a utilizar sea la adecuada para simplificar la implantación del sistema domótico permitiendo:

- Disminuir el cableado y en consecuencia los costes de instalación, mantenimiento y reconfiguración.
- Integrar servicios o instalaciones singulares, funcionando hasta ahora independientemente, para permitir la comunicación entre ellos.

El propósito de la Domótica, con los medios disponibles en la actualidad, se puede expresar del modo siguiente: "Aquello que permite una mayor calidad de vida a través de la tecnología, ofreciendo una reducción del trabajo doméstico, un aumento del bienestar y de la seguridad de sus habitantes y una racionalización de los consumos energéticos". En un futuro cercano otros aspectos importantes como son la salud, ecología, la cultura se irán incorporando a las posibilidades ofertadas por la Domótica (Fig. 1).

Los principales servicios disponibles actualmente por su intervención son los siguientes:

- Control y gestión de la energía.
- Seguridad.
- Automatización de sistemas e instalaciones domésticas.
- Comunicaciones.

En lo que respecta al medio de comunicación, en los diversas aplicaciones domóticas se utilizan hasta seis diferentes soportes (par trenzado, cable coaxial, fibra óptica, red eléctrica de alimentación, radiación infrarroja, radiofrecuencias).

3. LOS ESTÁNDARES PARA LOS SISTEMAS DOMÓTICOS

En el mundo de la Domótica existen varios modos mas o menos estándares para determinar los sistemas domóticos. Presentamos aquí algunos de los más significativos en el pasado reciente en Estados Unidos, Japón y Europa.

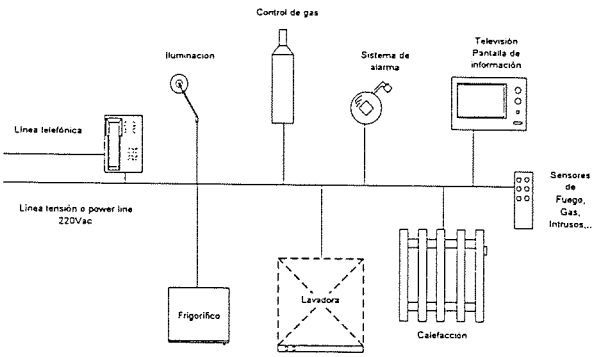


Fig. 1

3.1 Estados Unidos

Una característica de los diversas opciones es la menor sujeción a regulaciones con respecto a las existentes en Japón o Europa. Destacan principalmente los siguientes sistemas:

3.1.1 Smart house

Iniciativa orientada a investigar las posibilidades de la domótica en viviendas de nueva construcción.

Inicialmente se propuso una red domótica con cableado unificado, para reemplazar la multiplicidad de redes que coexisten en una vivienda convencional: electricidad, sonido, vídeo, teléfono, alarmas, etc.

3.1.1.1 Cebus

Destinado a la creación de una solución flexible para resolver el problema de la comunicación entre los diversos equipos domóticos sin que se produzcan interferencias.

La red está dividida en tres partes: la de alimentación eléctrica, la de control (formada por cable coaxial) y la de señal. Cada una de ellas cuenta con interfaces consistentes en receptores de radiofrecuencia e infrarrojos.

3.1.1.2 X-10

Sistema modular de bajo coste basado en comunicación por corrientes portadoras. Los módulos básicos consisten en unidades que se acoplan a las tomas de corriente normales, reciben señales digitales codificadas emitidas por un controlador a través de la red eléctrica, no existiendo comunicación de doble sentido entre el controlador y los módulos, por lo que no hay confirmación de que la orden se ha ejecutado correctamente.

Los módulos pueden realizar funciones de control de nivel, encendido y apagado de la iluminación y de puesta en marcha/paro de equipos audiovisuales. También se pueden utilizar para programar el funcionamiento de equipos eléctricos y algunos pueden realizar operaciones sencillas de control como regulación de temperaturas, alarmas acústicas, fugas, accionamiento a distancia por vía telefónica, etc.

3.1.1.3 Echelon

En esta solución se dispone de un circuito integrado específico que controla las redes de operación y permite convertir en inteligente cualquier dispositivo del sistema. Para ello integra los siguientes elementos:

- Protocolo de comunicaciones.
- Circuitos integrados con dispositivos de entrada/salida, microprocesadores y memoria.
- Dispositivos emisores/receptores que se encargan de conectar los chips con los medios de comunicación.

3.2 Japón

Los fabricantes de productos domóticos, a pesar de la diversidad de productos y la intensa competencia, acordaron un sistema estándar.

3.2.1 HBS (Home Bus Standard)

Este sistema no contempla qué es lo que debe soportar la red, sino cuál es su configuración básica:

- Una red de banda estrecha para controles.
- Una red de banda media para sonido.
- Una red de banda ancha para vídeo.

La configuración básica consta de 2 cables coaxiales y 4 pares trenzados, aunque también se puede soportar bajo un único cable coaxial.

Existe una variante del HBS destinada a apartamentos de viviendas colectivas denominado SHBS (Super Home Bus Standard).

3.3 Europa

Existen diversos trabajos de normalización impulsados por diferentes grupos empresariales, si bien ninguno de ellos ha conseguido establecer una solución que sea unánimemente aceptada por todos, se está en la actualidad en un proceso de convergencia entre las principales opciones que pudiera posibilitar una mayor interoperabilidad de sistemas con dispositivos de marcas distintas.

3.3.1 Batibus

La iniciativa de este sistema fue impulsada en Francia, con el propósito de establecer un bus de control doméstico estándar, de forma que su aplicación pudiera también extenderse al pequeño terciario. Por este motivo ha sido necesario el establecimiento de las especificaciones del bus en lo referente a:

- Medio de transmisión.
- Protocolo de comunicaciones.
- Lenguaje de comandos.

La característica más relevante de este sistema es la adopción de un único cable (par trenzado) de baja velocidad (4.8 Kbit/s) como medio de transmisión, al cual se conectan todos los equipos a controlar.

3.3.2 Eibus

Iniciativa alemana diseñada para realizar la gestión técnica de edificios que no permite la transmisión de señales de audio y vídeo ni telefónicas.

Al igual que Batibus, especifica los medios de transmisión, el protocolo de comunicaciones y el lenguaje de comandos.

Los medios de transmisión de este sistema son la red de distribución de energía eléctrica y un cable de 2 pares trenzados apantallados de baja velocidad (9.6 Kbit/s).

3.3.3 Global Home System

Desarrollado en Francia, está basado en la transmisión de señales por corrientes portadoras, y tiene como objeto hacer posible la compatibilidad de todos los productos domóticos de la empresa que lo diseñó.

Este sistema permite el mando local y a distancia de electrodomésticos, así como funciones relacionadas con la gestión energética.

3.3.4 Mediabus

Es un lenguaje domótico de comunicaciones desarrollado en Francia que permite la implantación de sistemas y aplicaciones domóticas en redes de televisión por cable.

Este sistema está desglosado en dos partes. Existe por un lado un sistema domótico en cada vivienda, y por otra parte un centro de gestión común, que recibe alarmas, lleva a cabo las tareas de telemantenimiento y gestión de equipos comunes, etc.

La comunicación bidireccional se realiza a través del cable coaxial de la red de televisión.

3.3.5 D2B

Inicialmente destinado a conseguir un estándar de comunicaciones para las viviendas que permitiese conectar todo tipo de equipos de diversas aplicaciones como seguridad, confort, control de audio y vídeo, etc.

Formado en la actualidad por un bus de tres cables que permite controlar todos los equipos de audio y vídeo conectados a dicho bus desde un menú situado en la pantalla del receptor de televisión y que posee una capacidad de comunicación bidireccional.

3.3.6 European Home System

Proceso de especificación independiente de fabricantes referente a medios físicos de transmisión, protocolos de comunicación, etc. para los sistemas electrónicos integrados en la vivienda. Este estándar incluye una serie de normas para las aplicaciones de control y comunicaciones,

incluyendo los principales medios de transmisión: pares trenzados, cables coaxiales, corrientes portadoras en red de alimentación, emisiones radio e infrarrojas.

4. EL SISTEMA DOMÓTICO FAGOR

Buscando la minimización de la complejidad para el usuario, en la solución base para el Bus Fagor se ha optado en primer lugar por la transmisión de la información a través de la red eléctrica. En consecuencia, no hay necesidad de instalar cableado adicional para el intercambio de información entre los diferentes componentes. Simplemente insertando el enchufe en la toma de corriente más próxima, se conecta a toda una red de información, espina dorsal de la automatización. Aparte del ahorro en gastos y esfuerzo de instalación, esta red de información permite hacer modificaciones cambiando la aplicación de una toma de red a otra, y las altas y bajas del sistema no requieren más que esa simple operación de enchufar/desenchufar en la red eléctrica.

Para conseguir esto las señales de datos digitales que procesan los microprocesadores que gobiernan los dispositivos domóticos han de ser convertidas, en señales senoidales que se puedan propagar por la red eléctrica, sin perder información y sin entrar en conflicto con la normativa Europea (EN50065) que establece unas firmes restricciones en las emisiones permitidas a la red que realice cualquier dispositivo electrodoméstico. Así pues, se requiere un interfase a la red eléctrica y el principal dispositivo de este interfase es un módem FSK (Frequency Shift Keying o codificación por cambio de frecuencia), el cual envía dos señales senoidales de diferentes frecuencias para cada señal digital. En concreto, utiliza una senoidal de 131.85 kHz para enviar un 1 lógico y la otra de 133.05 kHz para enviar un 0 lógico.

Un sistema domótico tal como el que FAGOR Electrodomésticos ha desarrollado en el marco del 'Hogar Inteligente' puede intercambiar datos por medio de la red eléctrica cubriendo toda la variedad de dispositivos típicos del hogar como son: Calderas eléctricas o a gas, Lavadoras, Lavavajillas, Hornos, etc. Para lograrlo se ha dotado a estos dispositivos de una electrónica adicional que permite estas comunicaciones y del software que implementa los protocolos de comunicación y añade algunas funcionalidades a las que les son típicas. Junto a estos aparatos, también se han desarrollado otros para las funciones siguientes:

- Detección de Gas, Agua, Incendio, Intrusos, Medición de Temperatura.
- Actuación: Cierre de Gas, Cierre de Agua, Alarma, Calefacción zonificada.
- Enchufes inteligentes, capaces de responder a requerimientos del sistema en red y en los que se pueden conectar electrodomésticos u otros dispositivos convencionales.

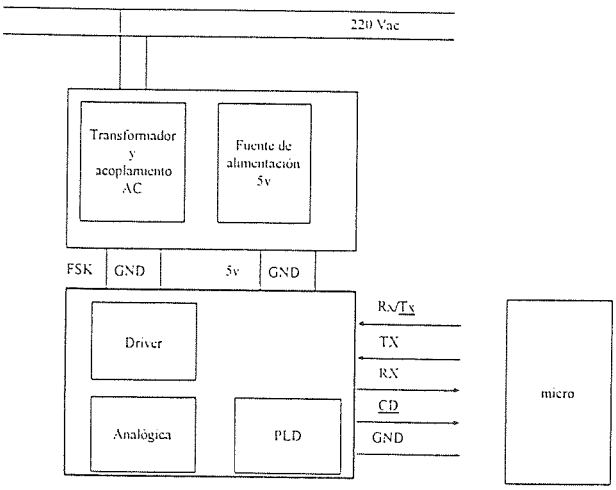


Fig. 2

Lo que se ofrece al usuario de estos sistemas es una mayor seguridad y confort con las funcionalidades de detección y actuación y además un ahorro económico por medio de una adecuada gestión de la energía utilizada por sus aparatos domóticos.

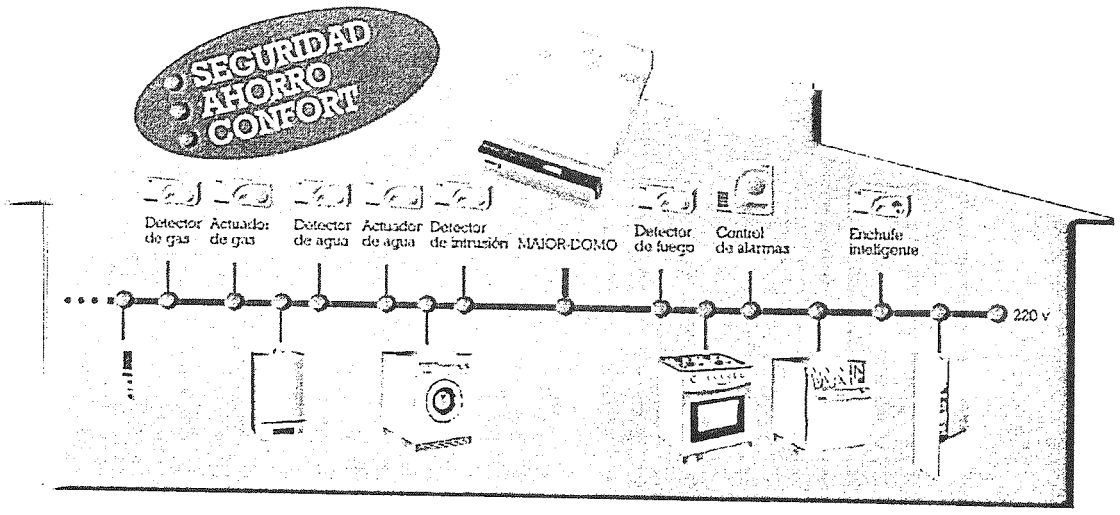


Fig. 3

5. EL CONTROLADOR DOMÓTICO

El elemento encargado de la coordinación de todas estas funciones es el Controlador Domótico.

Control

La coordinación del conjunto de funciones se ejerce mediante el control de todos los intercambios de mensajes que se realizan en el sistema dando permisos, órdenes y recibiendo informaciones en relación con el estado de operación de los elementos que integran el sistema. Este dispositivo además de su función de control, permite la configuración y monitorización de la red.

Contestador

Por otro lado además de estas funciones propias de Controlador Domótico este dispositivo es un completo contestador telefónico digital para el hogar y al estar conectado a la línea telefónica puede informar de cualquier evento ocurrido en el sistema del Hogar, si el usuario así lo desea, a los teléfonos prefijados por éste. Estos eventos mencionados son comunicados al Controlador por los dispositivos domóticos conectados a la red eléctrica.

Telecontrol

El Maior-Domo permite también, la activación/ desactivación telefónica-remota de los dispositivos e incluso la variación de las reglas de comportamiento del sistema.

El interface directo con el usuario se realiza a través de un conjunto de teclas y un visualizador de cuatro dígitos. También, el Maior-Domo permite la utilización de un terminal telefónico (tanto local como remoto) para recibir comandos sobre la Red Domótica y monitorizar su funcionamiento.

5.1 Funciones principales

5.1.1 Gestión de la Red Domótica Fagor

La función de gestión de la red ejercida por el Maior-Domo se refiere tanto a la gestión de la potencia contratada como a la gestión de la tarifación. Esto significa que el Maior-Domo es el encargado de conceder o negar el permiso de activación a un aparato integrado en la Red Domótica en función de la potencia disponible con respecto al límite establecido por el contrato del usuario con la entidad suministradora de energía eléctrica y la prioridad del dispositivo establecida por el usuario.

El Maior-Domo envía una señal indicando cambio de tarifa cuando la hora presente entra en una de las dos franjas de tarifa reducida establecidas previamente. Si alguno de los electrodomésticos integrados en la Red Domótica se hallara en espera de esta tarifa para la activación, será "despertado". Seguidamente pedirá permiso de activación al Maior-Domo cuando le llegue el mensaje de cambio de tarifa y éste, según las condiciones establecidas por la prestación de gestión de potencia, concederá o negará el permiso para activarse.

Asimismo, el Maior-Domo gestiona también las posibles situaciones de alarma o disfunción que se puedan producir y tiene la capacidad de llamar a números telefónicos predefinidos por el usuario e informar mediante una serie de mensajes hablados la situación de alarma que se ha detectado. Estos mensajes están precedidos por textos de cabecera de alarmas grabados previamente por el usuario.

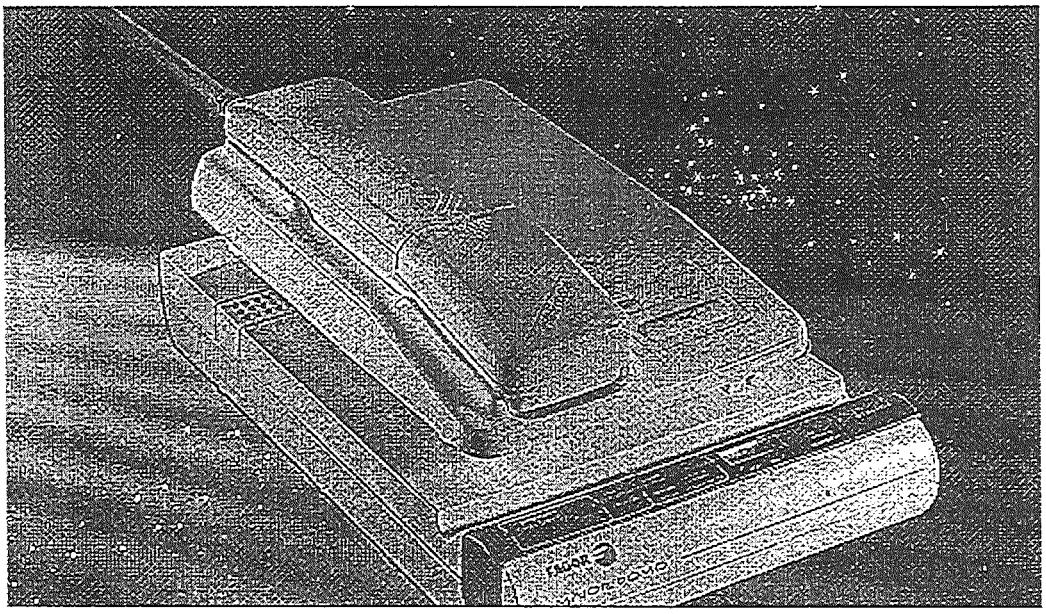


Fig. 4

5.1.2 Acceso Remoto a la Red Domótica

El Maior-Domo será el encargado de posibilitar el acceso remoto a través de la línea telefónica del usuario a los aparatos integrados en la Red Domótica. Para ello, el usuario introducirá un código personal mediante el cual se identificará y habilitará las funciones que el Maior-Domo dispone en modo de funcionamiento local, como pueden ser la activación y desactivación de los aparatos, o la petición de *status*.

5.1.3 Contestador Automático

Cuando la función de contestador automático está activada, el Maior-Domo Fagor procede a grabar las llamadas que no sean atendidas por el usuario tras haber reproducido un mensaje de bienvenida previamente grabado. Posteriormente, se podrán reproducir los mensajes almacenados, saltarse la reproducción de mensajes concretos, proceder al borrado selectivo o completo de los mensajes, y reproducir los mensajes a través del altavoz o utilizando el auricular del teléfono.

La memorización de los mensajes se realiza en memorias de silicio (no utiliza cintas magnéticas). Una de las principales características de este tipo de contestador digital es que permite un tiempo máximo de grabación de quince minutos o un máximo de 64 mensajes. Las funciones disponibles son las siguientes:

- Activación/desactivación de las funciones del contestador automático.
- Reproducción local (altavoz) de los mensajes recibidos, indicando el número de mensaje.
- Reproducción de mensajes desde un Terminal Telefónico remoto.
- Borrado de todos los mensajes (local y remoto).
- Grabación del mensaje de bienvenida.
- Reproducción del mensaje de bienvenida.
- Grabación del mensaje de cabecera de alarma.
- Reproducción del mensaje de cabecera de alarma.

5.2 Modos de Uso

El sistema domótico de Fagor está concebido teniendo como objetivo prioritario la facilidad de instalación tanto en nuevas viviendas como en las ya existentes, no necesitándose ninguna provisión además de los aparatos domóticos, el Maior-Domo y la red de 220v con sus tomas habituales.

Para que las preferencias y necesidades particulares del usuario sean introducidas en el sistema, el Controlador Domótico dispone de los medios para su programación específica por el usuario, quedando esta configuración almacenada indefinidamente en su memoria mantenida con batería.

5.2.1 Programación del Maior-Domo

Para la programación del Maior-Domo y su adecuación a las necesidades específicas del usuario, no se precisa un entrenamiento elaborado, ya que una serie de mensajes pregrabados van guiando al programador en las diversas opciones.

Para grabar el mensaje de bienvenida, el usuario debe pulsar la tecla "Grabar Mensaje Contestador". La duración de este mensaje está limitada a un máximo de un minuto.

Las funciones para atender llamadas se activan/desactivan mediante la tecla "Activar", y se confirma la operación mediante el correspondiente indicador.

Una vez que se haya atendido la llamada, el visualizador mostrará la leyenda "rec", se reproducirá el mensaje de bienvenida y se emitirá un pitido para indicar el comienzo de la grabación del mensaje.

Al concluir la grabación el visualizador volverá a mostrar la hora actual, mientras que el indicador luminoso de mensajes se activará si el mensaje era válido.

Para la reproducción de mensajes y una vez pulsada la tecla "Lectura/Pausa", el Maior-Domo inicia la reproducción de cada uno de los mensajes almacenados de forma secuencial, indicando en el visualizador el número de mensaje que está siendo reproducido en cada momento.

La totalidad de mensajes almacenados en el Maior-Domo pueden ser borrados manteniendo pulsada la tecla de "Borrar" durante dos segundos.

También es posible borrar un único mensaje (borrado selectivo), para ello se debe pulsar dos veces la tecla de borrado mientras se está reproduciendo el mensaje que queremos borrar.

El usuario tiene la posibilidad de reproducir o borrar remotamente los mensajes almacenados. Estas operaciones están contempladas en las funciones del Maior-Domo en Modo Local, por lo que el usuario puede llamar a su vivienda y, una vez accedido al Modo Local, proceder a reproducir o borrar los mensajes almacenados.

Las horas de comienzo y final de los períodos de tarifación económica son introducidas en el Maior-Domo por el usuario (o instalador) accediendo a la programación en Modo Local, guiándose por los menús hablados.

La configuración del sistema se realiza también desde el Modo Local.

Para posibilitar el acceso remoto del usuario, a través de la línea telefónica, a los aparatos integrados en la Red Domótica se introduce un código de identificación/autentificación de tres cifras que habilita diversas funciones que el Maior-Domo dispone en modo de funcionamiento local. Si el código introducido es correcto, el Maior-Domo responderá con el mensaje hablado **"activación y consulta remota de los electrodomésticos"**. En caso contrario, el Maior-Domo guardará silencio para no delatar su condición de gestor de un sistema domótico.

En una red tipo las operaciones sobre los aparatos se podrían codificar del modo indicado en la tabla 1:

Tabla 1

CÓDIGO APARATO	OPCIONES
1- LAVADORA	1- Activar / ON 2- Desactivar / OFF 3- Status
2- LAVAVAJILLAS	1- Activar / ON 2- Desactivar / OFF 3- Status
3- CALDERA	1- Activar / ON 2- Desactivar / OFF 3- Status 4- Programar
4- HORNO	1- Activar / ON 2- Desactivar / OFF 3- Status
5- MÓDEM	1- Reproducir Mensajes 2- Borrar Mensajes

Como se puede observar en este ejemplo, un comando completo está formado por tres campos: Código, Opción y Datos. Para verificar la configuración actual el usuario dispone de un comando mediante el cual el Maior-Domo muestra dicha configuración utilizando una combinación de mensajes hablados y visuales (a través del altavoz y visualizador). Los mensajes hablados indicarán la potencia asignada a la Red Domótica, e irán nombrando los electrodomésticos instalados en la red mientras que, simultáneamente, aparecerán en el visualizador la potencia y la prioridad de cada uno de ellos. Por ejemplo: mensaje "LAVADORA", visualizador "15P2". Indicando que la potencia de la lavadora es de 1.5 kW y la prioridad es 2. Al finalizar la verificación se volverá al menú principal.

El Maior-Domo dispone de un reloj cuya misión es suministrar la referencia horaria para los eventos y temporizaciones, permitiendo controlar los cambios de tarifa según las franjas horarias configuradas. El visualizador muestra normalmente la hora, y sólo deja de hacerlo cuando se entra en Modo Local, se están reproduciendo los mensajes grabados o se está grabando el mensaje de bienvenida.

El usuario puede cambiar la hora mediante las teclas "Hor." y "Min.".

Durante el modo de funcionamiento normal el Maior-Domo Fagor se alimenta a través de la red de tensión de 220v y todas las prestaciones descritas están a disposición del usuario. En caso de que falle la tensión de alimentación, el Maior-Domo dispone de una batería auxiliar con la que mantener únicamente algunas funciones básicas, como mantener la configuración actual de la Red Domótica y mantener los mensajes grabados en memoria.

Si transcurridas dos horas no hubiera vuelto la tensión en la red, el Maior-Domo se activará para emitir un mensaje de alarma a los números telefónicos predefinidos por el usuario indicando que se ha producido un fallo de alimentación. Este mensaje está pensado para que el usuario quede advertido de que en el congelador se está a punto de romper la barrera del frío, con lo cual se corre el riesgo de que los alimentos se descongelen, con la consiguiente problemática que esto supone.

6. CONCLUSIONES

El sistema domótico de Fagor y su controlador Maior-Domo que se ha descrito en este documento es una realidad que está siendo comercializada a través de canales especializados como es el destinado a constructores de viviendas. Si bien el mercado de la domótica es aún incipiente, la evolución hacia la inclusión de las prestaciones que ofrece en los dispositivos de consumo estándar es inequívoca.

7. REFERENCIAS

[1] The Intelligent Home.
(Documento Fagor, 1997)

[2] European Home System Specification Release (1.1 March 1992)

[3] Revista Domótica, (varios ej. 1996)

Estado del Arte sobre aplicaciones de la arquitectura modular de STREAMS en los sistemas de comunicaciones

Armando Ferro, Mikel Olabe e Iñaki Goiricelaia
{jtpfevaa|jtpolbam|jtpgoori}@bi.ehu.es
Telematic Networks Group
Escuela Técnica Superior de Ingenieros Industriales y de Telecomunicaciones de Bilbao
September 15th , JITEL'97

Abstract

The modern communication systems must support each time more protocols and several applications running together. The system's performance could be affected by this overhead and a new flexible manner to solve the problem is needed. It is often proposed the use of multiple identical general-purpose processors in parallel (a multiple-instruction, multiple-data, or MIMD approach). Such methods need a large numbers of proccesors and programmers can find difficulty in efficiently parallelizing software. In other way, a less-well-known form of parallel computation, based on STREAMS, is conceptually closer to the ways in which programmers think about algorithms and offer a more cost-effective, scalable, highly-integratable approach to flexible communications systems.

In this paper, we explain the concept of STREAM-based arquitecture and describe why it is a good match to solve problems in different enviroments. We are showing several applications of this technology that some companies are using around the world in different areas.

1. Introducción.

La arquitectura modular de STREAMS no es una innovación muy reciente. Originalmente fué propuesta y desarrollada por Dennis Ritchie[1] en los laboratorios Bell de AT&T sobre el año 1982. Sin embargo su implantación en los sistemas de comunicaciones no se realizó de forma generalizada a pesar de las ventajas aparentes que proporcionaba hasta algunos años más tarde.

Fué AT&T en el año 1987 el primero en hacer una implementación [2] de esta arquitectura a nivel de sistema operativo sobre la versión de Unix System V release 3. Posteriormente, en 1989 lo realizó sobre la release 4 [3]. Quizá fuese éste el producto que más ha difundido comercialmente la arquitectura propuesta por Dennis Ritchie. Sin embargo la implementación se ha realizado sólo a nivel de sistema operativo. En concreto como una nueva arquitectura para implementar drivers de dispositivos de una forma modular. Posteriormente AT&T coge propiedad del nombre comercial de STREAMS.

A partir de esto, han sido numerosos los fabricantes de diferentes productos los que han ido incorporando esta arquitectura en el diseño de sus productos. En el año 1993 ya estaba disponible como arquitectura de dispositivos sobre sistemas operativos en tiempo real. En la actualidad existen más de tres vendedores de S.O en tiempo real que ofrecen STREAMS. También han implementado esta arquitectura fabricantes de dispositivos de tarjetas de comunicaciones inteligentes que ofrecen mejoras en las prestaciones y rendimientos de los protocolos. En la actualidad, el concepto original de STREAMs se ha visto más ampliamente generalizado. Existen implementaciones de la arquitectura próximas al hardware. Estas soluciones proporcionan una capacidad de procesamiento en paralelo que permite aprovechar las posibilidades de la arquitectura en el multiproceso con rendimientos muy altos.

El concepto de STREAM como se ha mencionado anteriormente ha sufrido una generalización importante de como originalmente fue contemplado por su creador. De hecho se pueden encontrar términos ambiguos que pueden llevar a confusión. Generalmente al hablar de STREAMS se entiende comúnmente como la arquitectura software según la ha realizado AT&T en su S.O. Unix. Ésta básicamente implementa drivers de dispositivos de una forma modular. Sin embargo pronto se fueron viendo muchas más aplicaciones de esta arquitectura que caían fuera de los S.O. y de ahí que aparezcan sucedáneos de la tecnología de STREAMs conocidos como "*procesamiento modular*", "*Multisubsystem protocol arquitectures*", "*Stream-based computing*", etc. Todas ellas giran alrededor de la misma idea y , al menos en este artículo, se consideran variantes de la arquitectura original conocida con el nombre de STREAMS e implementada de diferente forma.

2. Definición de STREAM

Vamos a realizar una definición genérica de lo que se entiende por STREAM y posteriormente

mostraremos la arquitectura básica del mismo basándonos en el modelo definido por AT&T.

Un STREAM se puede considerar de forma genérica [4] como un objeto que acepta secuencias de caracteres. Es el destinatario de los datos mismos. Debe considerarse como algo más de lo que son las cadenas de caracteres utilizadas en lenguajes de programación para escribir o leer datos desde un dispositivo. Puede ser cualquier objeto que acepte datos, tal vez una clase tipo stream en lenguaje ANSI-C o un módulo software, o incluso un agujero negro que sea capaz de absorber datos sin devolverlos. Los STREAMS pueden ser usados para transportar datos internamente en una aplicación, a través de una red, utilizando ficheros locales, sobre dispositivos, etc. Los módulos STREAMs se pueden poner en cascada formando una cadena. Para ello se une la salida de un módulo a la entrada del siguiente. De esta forma los datos que atraviesen la cascada de los STREAMs se verán afectados por el procesamiento que se realice en cada uno de los módulos.

Los STREAMs pueden ser usados para transportar datos desde aplicaciones hacia las redes de comunicaciones o viceversa. Permiten la comunicación en ambos sentidos. El procesamiento en los módulos está orientado a la realización de eventos. Es decir, los datos son enviados a través de un módulo y éste realizará el procedimiento que le corresponde sobre el flujo de datos dejándolos disponibles a la entrada del siguiente módulo que a su vez procederá a realizar su función.

La clase genérica de un STREAM proporciona una interface uniforme a las demás clases específicas de STREAMs sin importar a la subclase que pertenezca. De esta forma se pueden intercomunicar entre ellas aún siendo de diferente naturaleza. Se puede definir una clase genérica que incorpora los métodos representados en el siguiente gráfico.

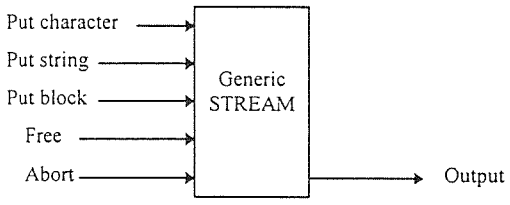


Fig.1 DIAGRAMA de STREAM Genérico

2.1 STREAMS según AT&T

Usaremos este modelo [5] para introducirnos en la arquitectura básica utilizada en los STREAMS. Aunque estos conceptos están más relacionados con la implementación software, se puede hacer una transposición hacia otras implementaciones.

El mecanismo de STREAMS utilizado por AT&T proporciona la infraestructura para el sistema operativo sobre la cual se pueden construir los servicios de comunicaciones. Estos servicios incluyen las comunicaciones entre los terminales y el ordenador principal, bien sea entre procesos de la misma máquina o entre procesos de diferentes máquinas.

El subsistema de STREAMS fue diseñado para unificar los mecanismos existentes en los sistemas UNIX para realizar las diferentes clases de procesamiento en dispositivos I/O de entrada y salida. En particular se pensó para reemplazar el mecanismo de *clist* soportado hasta el momento en versiones UNIX anteriores.

Los STREAMS proporcionan una nueva forma de trabajar con los dispositivos. Los usuarios pueden añadir (*"push"*) o eliminar (*"pop"*), según sea su deseo, elementos de procesamiento intermedio, denominados "*módulos*". Los módulos pueden ser guardados de tal forma que incluso uno puede ser usado al mismo tiempo por el mismo o varios STREAMs. Este cambio fundamental permite que módulos independientes que realicen tareas simples puedan ser combinados para realizar tareas más complejas.

La transferencia de datos en un STREAM software se realiza mediante el paso de mensajes entre elementos de procesamiento adyacentes. Solamente se pasarán los punteros a los mensajes, evitando la sobrecarga de la copia de datos. Los mensajes se pueden tipificar y se les puede asignar una prioridad para indicar como deben de ser procesados. El envío de mensajes para realizar las operaciones de I/O establece una diferencia fundamental de este mecanismo frente a los clásicos; las operaciones serán establecidas por el envío de datos (data-driven) y no por su espera (demand-driven). Es decir, antes un proceso que demandaba datos solicitaba los mismos invocando una rutina de lectura. Con los STREAMS las operaciones se realizarán cuando los datos estén disponibles.

2.1.1 Arquitectura de STREAMS

Un STREAM básico proporciona un camino de datos bidireccional entre un proceso de nivel de usuario y un dispositivo a nivel de kernel que se encargará de hacerlos viajar por la red. Los datos suministrados por el usuario viajarán hacia abajo (downstream) en dirección el controlador del dispositivo (driver) y los datos recibidos por el

controlador de dispositivo desde el hardware viajarán hacia arriba (upstream) para ser recibidos por el usuario. Los módulos tratarán al flujo de datos como una cadena de bytes vayan en un sentido o en el otro.

Un STREAM sencillo está compuesto al menos por dos elementos básicos de procesamiento: el STREAM cabecera (Head) y el driver.

El STREAM Head consiste en un conjunto de rutinas que proporcionan la interfaz entre las aplicaciones de los usuarios y el resto del STREAM en el núcleo del sistema. Es el único módulo que está en contacto con la aplicación de usuario. Este módulo no se puede sustituir, y el mismo proporciona acceso a diferentes aplicaciones cada vez que se monta un STREAM en el sistema. Sin embargo los módulos cabecera pueden ser preparados a medida para cada aplicación utilizando las opciones de procesamiento adecuadas.

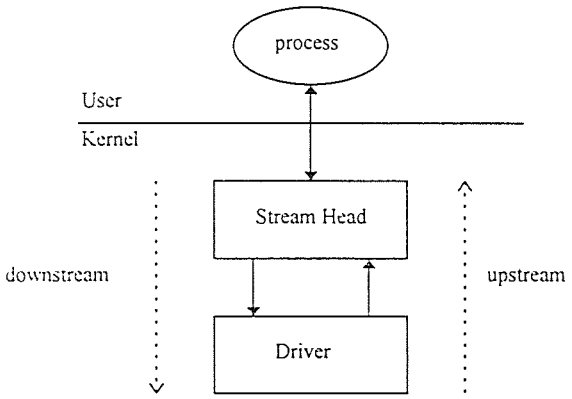


Fig.2 Un STREAM Simple

El otro elemento es el driver que se sitúa en el extremo inferior del STREAM en contacto con el dispositivo. Su trabajo es controlar el dispositivo y realizar la tranferencia de datos entre el STREAM y el dispositivo. Debido a que interactúa con el hardware se conoce al controlador *como hardware driver*. Este módulo puede sustituirse por un controlador virtual en software que no está en contacto con ningún hardware. Esto se realiza para determinadas aplicaciones y se conoce como *software driver* o *pseudo-driver*.

Las interacciones entre los módulos dentro del STREAM se realiza a través de mensajes intercambiados por las colas (queue). Estas son las que unen un componente al siguiente formando así el STREAM. Cada componente contiene al menos un par de colas, una para leer (read side) denominada *upstream* y otra para escribir (write side) denominada *downstream*. Estas colas sirven para almacenar mensajes, contienen información de estado y actúan como un registro de las rutinas que deben procesar los mensajes.

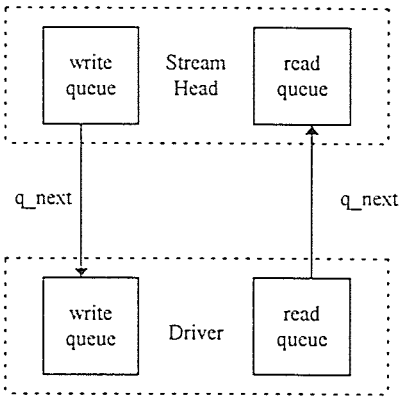


Fig.3 Colas en STREAMS

Un módulo es un elemento de procesamiento intermedio que puede añadirse o eliminarse de forma dinámica del STREAM. Los módulos son estructuralmente similares a los drivers salvo que realizan funciones de filtrado y procesamiento intermedio sobre los mensajes que pasan entre el elemento Head y el Driver. Por ejemplo, puede realizar funciones de encriptación o cambios de protocolos entre interfaces diferentes para adaptarlas.

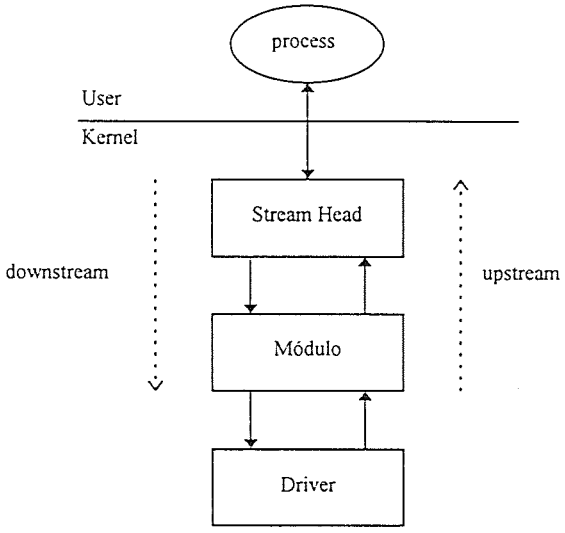


Fig.4 Un Módulo en un STREAM

Para poder configurar a medida la labor de un STREAM a parte de añadir y quitar módulos

intermedios se pueden realizar también otras operaciones como establecer o eliminar estructuras multiplexadas entre STREAMs diferentes. Varios STREAMs se pueden enlazar bajo un módulo controlador especial *denominado "driver multiplexador"* o "*multiplexor*".

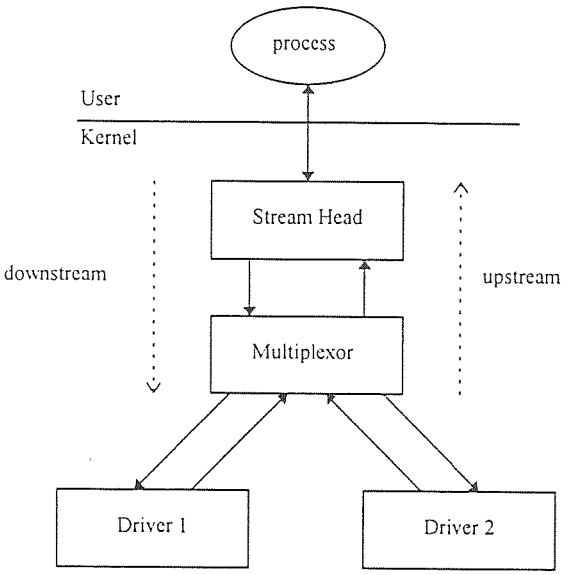


Fig.5 Un Modulo Multiplexor

El módulo multiplexor se encargará de rutar los mensajes entre la parte alta del STREAM abierta para acceder al driver y las cadenas de STREAMS enlazadas debajo del multiplexor. Esta estructura es muy apta para implementar sistemas de gestión de múltiples ventanas o varios protocolos sobre redes de comunicaciones de datos. Los sistemas de gestión de ventanas multiplexan varias ventanas sobre un mismo terminal. Las redes de comunicaciones multiplexan mensajes entre varios usuarios y permiten comunicaciones por diferentes medios físicos.

Una vez vistos los componentes principales de un STREAM nos podremos fijar en algunas características de interés a la hora de pasar información entre sus elementos. El paso de información en las implementaciones software se realiza a través de mensajes. Estos mensajes están predefinidos y su tipo indicará el propósito del mensaje y su prioridad. Se pueden asignar prioridades al flujo de información dentro de un STREAM, de tal forma que se subdividen *en bandas de prioridad* con el propósito de establecer el control de flujo y el almacenamiento de mensajes.

La arquitectura hasta ahora definida corresponde más bien a la implementación de los STREAMs sobre el kernel de un Sistema Operativo, sin embargo esta arquitectura ha ido evolucionando desde que se definió a nuevos entornos de aplicación y aunque conserva la esencia fundamental, su implementación ha sufrido modificaciones para adaptarse a cada plataforma y a los requerimientos de cada aplicación. Sirva pues lo visto como un acercamiento a la arquitectura genérica de STREAMs.

3. Ventajas de la utilización de STREAMS.

La arquitectura modular de STREAMs permite disponer de numerosas ventajas sobre otras arquitecturas debido en gran parte a su modularidad y flexibilidad en la aplicación. Son muchas las razones por las cuales es aconsejable su uso y seguidamente se expresan algunas de ellas:

- Proporciona una infraestructura estándar para soportar la implementación de servicios en redes de altas prestaciones.
- Permite intercambios de información full-duplex entre las aplicaciones y los dispositivos.
- Facilita la portabilidad de protocolos desarrollados en diferentes plataformas.
- Permite la reutilización de módulos estándares.
- Soporta flujos de información simultáneos de datos y control.
- Utiliza mecanismos comunes para el almacenamiento y el procesamiento de los datos.
- Soporta la multiplexación, una necesidad común en los protocolos de comunicaciones de las redes actuales.
- Permite compartir memoria en la gestión entre mensajes.
- Puede priorizar los flujos de información.

Estas son algunas de las ventajas que está proporcionando esta arquitectura sobre todo en la implementación como subsistema en muchos de los actuales sistemas operativos multitarea.

Sin embargo actualmente se está desarrollando esta tecnología con algunas variantes sobre entornos nuevos donde se pueden apreciar ventajas añadidas a las referenciadas con anterioridad.

4. Entornos de aplicación y desarrollos en curso

Dada la flexibilidad que proporciona esta arquitectura se ha ido introduciendo en diversos entornos. Existen en la actualidad varios grupos de trabajo en el mundo que están ayudando con su labor de investigación al desarrollo de esta tecnología. Debido a las ventajas claras que aporta la tecnología en determinados sectores, esta labor de investigación

no solamente se lleva a cabo en el ámbito científico y académico, sino que son numerosas las empresas privadas que están colaborando en su desarrollo implementando productos nuevos que incorporan las ventajas de esta arquitectura.

En cuanto a los entornos de aplicación de esta arquitectura hemos realizado una pequeña clasificación de los que se considera con mayor proyección.. Pueden existir entre estos entornos algunos solapes y es seguro que debido a la todavía reciente explosión de la tecnología no se incluyen muchos entornos nuevos que potencialmente pudieran aprovechar sus beneficios.

4.1 Arquitectura sobre Sistemas Operativos

Es sobre la plataforma de Sistemas Operativos donde inicialmente se ha ido desarrollando la arquitectura de STREAMS.

La implementación de STREAMS ha estado centrada en proporcionar un subsistema modular que permitiese soportar la implementación de controladores de dispositivos de una forma más inteligente y flexible.

Inicialmente, a partir del año 1987, han sido vendedores de sistemas operativos multitarea de entornos UNIX quienes han incorporado las primeras soluciones; UNIX de AT&T, SCO UNIX, etc.

Posteriormente, sobre el año 1993, los STREAMS han sido incorporados al entorno de Sistemas Operativos en tiempo real [6]. Actualmente se encuentran disponibles soluciones de al menos tres vendedores importantes de S.O en tiempo real.

La implementación sobre S.O. en tiempo real aporta los siguientes beneficios:

- Proporciona un API (aplication Program Interface) compatible con la interfaz estandar de I/O en UNIX.
- Se puede implementar tareas de los STREAMS como threads del propio kernel.
- El mecanismo de control de flujo de los STREAMS proporciona mejores prestaciones.
- Permite aislar el entorno de usuario del núcleo del S.O.

Estas implementaciones de STREAMS sobre S.O. son propietarias y no proporcionan código fuente de libre disposición que facilite la labor de investigación sobre dichas plataformas.

Existen productos comerciales que permiten implementar la arquitectura de STREAMs sobre un sistema operativo multitarea. Un ejemplo es el producto MPS (Mentat Portable Streams) de la empresa Mentat Inc. que ofrece no solamente el software para implementar la arquitectura, sino que además ofrece módulos basados en STREAMS que incorporan los protocolos más extendidos en las redes de comunicaciones (TCP/IP, XTP, X.25, FR, etc). Se pueden utilizar también módulos de otros fabricantes. Una característica muy importante de este producto es el soporte multiprocesador. Manipula la sincronización de actividades, la liberación de módulos y la gestión multithread.

Otra implementación comercial de STREAMs para S.O. en tiempo real es la proporcionada por la empresa Wind River Systems con su producto WindNet STREAMS. Proporciona soporte para implementar STREAMS sobre sistemas en tiempo real. Está basado en el producto de Mentat e incorpora capacidades especiales para asegurar el procesamiento en tiempo real.

Existe una implementación de STREAMs en código fuente para LINUX y QNX gracias al trabajo desarrollado por el grupo del proyecto LiS. Este proyecto tenía como objetivo disponer de una implementación de STREAMs que fuera de uso público. Sus trabajos están basados en la arquitectura de STREAMs de System V Release 4 e implementa multiplexación y un debuger que facilita el desarrollo de aplicaciones.

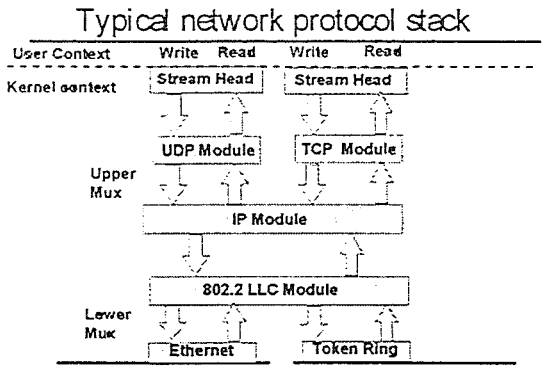


Fig.6 Ejemplo de aplicación de STREAMS en S.O.

4.2 Arquitectura sobre Periféricos Inteligentes

Una aplicación interesante de la arquitectura de STREAM y que ya muchos fabricantes están empleando es la posibilidad de implementar módulos del STREAM en un dispositivo periférico inteligente. Esto permitiría descargar al ordenador principal de las tareas relacionadas con los periféricos dejando más CPU disponible para sus funciones principales. Además esta arquitectura permitiría que el usuario pueda incorporar sobre el periférico nuevos módulos software a medida. Estos pueden realizar el

procesamiento específico que corresponda sin afectar los tiempos de disponibilidad de la CPU principal.

La descarga de los protocolos de las redes sobre un periférico externo inteligente permiten a los diseñadores de sistemas una mayor flexibilidad y mejores prestaciones. Muchas tarjetas inteligentes disponen hoy en día de procesarores que rivalizan en capacidad con los del ordenador principal. Desplazar el procesamiento de protocolos de la red al dispositivo externo permite disponer de más CPU para otras tareas. Además, implementar los protocolos de las redes sobre los periféricos permitirá disminuir el volumen de datos entre el dispositivo y la memoria del sistema. Los paquetes recibidos pueden ser unificados y las cabeceras de control eliminadas antes de transferirse al ordenador.

Se puede incluso mejorar las prestaciones en las redes de comunicaciones al ejecutar los protocolos sobre los periféricos externos, especialmente cuando el ordenador no es muy potente.

Si la capacidad de descargar el procesamiento hacia los periféricos es un beneficio tan importante, ¿ cómo es que no se ha tenido en cuenta hasta ahora ?.

Para disponer de protocolos de comunicaciones funcionando en los periféricos externos es necesario disponer de aplicaciones específicas no estándares para ese hardware, lo cual hace difícil su actualización y mantenimiento. Además si un desarrollador quiere implementar nuevos protocolos sobre la tarjeta precisa de un entorno de desarrollo específico para cada plataforma que además es propietario. Necesita conocer los entresijos de la arquitectura hardware del periférico, los servicios que proporciona el firmware y precisa de una documentación que es limitada. Resulta muy duro añadir nuevo software sin modificar el código existente. Luego la carencia de una arquitectura estándar impide implementar un código nuevo e independiente.

En vista de esto han surgido diferentes arquitecturas para periféricos inteligentes que permiten superar los inconvenientes anteriormente relacionados. El ejemplo más característico es el protagonizado por la empresa RNS [7] que ha diseñado un entorno de desarrollo para dispositivos inteligentes basándose en la arquitectura de STREAMs. El producto se conoce bajo el nombre de STRIDE (STREAMS Integrated Development Environment).

STRIDE proporciona dos componentes principales. El S.O. STRIDE que se ejecutará sobre la CPU del periférico inteligente y proporciona la arquitectura básica de STREAMS SVR4. Y el proceso de enlace de STREAMS (Remote Streams Construction Process) que permite unir la arquitectura de STREAMs del ordenador principal con el dispositivo inteligente. De esta forma los usuarios pueden decidir donde se instalarán los módulos que implementan los protocolos. Se podrán instalar en el S.O. del ordenador principal o para mejorar rendimientos sobre la arquitectura de STREAMs del periférico. La transparencia de este método es tan flexible que el usuario puede elegir que protocolos deben de ser ejecutados en la CPU del periférico.

Arquitecturas similares son proporcionadas por fabricantes de tarjetas de comunicaciones. Muchos de ellos a diferencia del anterior, venden soluciones propietarias con sus módulos específicos. Se basan en la arquitectura de STREAMs para obtener beneficios en la reutilización de sus desarrollos y en la facilidad para incorporar soluciones de terceros sobre su arquitectura. Proporcionan además la ventaja a los clientes de permitirles instalar sobre sus dispositivos módulos software desarrollados por ellos sin necesidad de conocer a fondo la arquitectura del producto.

Ejemplos representativos de esto puede ser la empresa Stallion que fue quizás una de las pioneras en la implementación de la arquitectura de STREAMs sobre su tarjeta IntelillyPort. Se trataba de una tarjeta multipuerto inteligente que permitía instalar módulos específicos sobre cada puerto. Esto aumentaba considerablemente los rendimientos de las comunicaciones al liberar a la CPU central del ordenador no solamente del protocolo sobre el puerto, sino también del tratamiento de las interrupciones en las comunicaciones serie.

Stallion sigue apostando por esta arquitectura con su producto Xtream que proporciona sobre su adaptador inteligente de comunicaciones de un rango de aplicaciones para redes WAN. Permite encapsular multiples protocolos sobre la tarjeta de comunicaciones liberando al sistema central de esa función.

Otro ejemplo a considerar es el del fabricante canadiense The Software Group que ha realizado un esfuerzo considerable migrando sus desarrollos sobre una tarjeta inteligente a una arquitectura de STREAMs. Esto le permite hoy en día mayor flexibilidad para integrar diferentes productos. Disponen de un producto denominado Netcom que ofrece conectividad a redes WAN. La arquitectura de STREAMs les permite encapsular diferentes protocolos en la misma tarjeta y ofrecer soluciones de conectividad a medida sin más que

incorporar los módulos adecuados sobre la arquitectura.

4.3 Procesamiento Paralelo

El procesamiento paralelo permite disponer de mayor capacidad computacional pero a consta de precisar de protocolos de sincronización de las actividades de los procesadores. Además es preciso disponer de un esquema de procesamiento que facilite la distribución de tareas entre procesadores. Esto obliga a los programadores a la hora de diseñar el código a tener en cuenta la estructura del sistema multiprocesador para generar algoritmos que puedan ser interpretados por el hardware.

La arquitectura de STREAMs ofrece una forma de diseñar aplicaciones más próxima a los algoritmos utilizados en procesamiento paralelo. Además independiza los flujos de información entre aplicaciones de la arquitectura hardware donde están soportadas. De esta forma una aplicación concebida para ser realizada en un sistema multiprocesador puede perfectamente ser desarrollada o implementada en un sistema monoprocesador y viceversa.

El aprovechamiento de varias unidades de procesamiento múltiple en un diseño que sea escalable precisa que las partes del algoritmo sean ejecutados en paralelo. Existen dos métodos para conseguir el paralelismo:

- Paralelismo de control. Las diferentes partes del algoritmo son ejecutadas al mismo tiempo por diferentes unidades de procesamiento.
- Paralelismo de datos. Cuando una parte del algoritmo que está siendo aplicada a una estructura de datos es ejecutada en paralelo por diferentes unidades de procesamiento sobre subdivisiones de la estructura de datos.

La arquitectura de STREAMs puede soportar ambos métodos de paralelismo.

Los STREAMs proporcionan un mecanismo de sincronización que apoya el control del paralelismo entre las partes del algoritmo que se realizan simultáneamente. Los STREAMs también imponen un estilo de programación que ayuda enormemente a la aproximación modular que precisa el algoritmo haciendo el paralelismo más evidente.

El número de operaciones que se deben planificar por unidad de procesamiento (*granularidad*) dependerá de la arquitectura de la máquina y del propio algoritmo a ejecutar. Una granularidad pequeña permite más oportunidades para el paralelismo pero puede incurrir en una excesiva sobrecarga en labores de planificación.

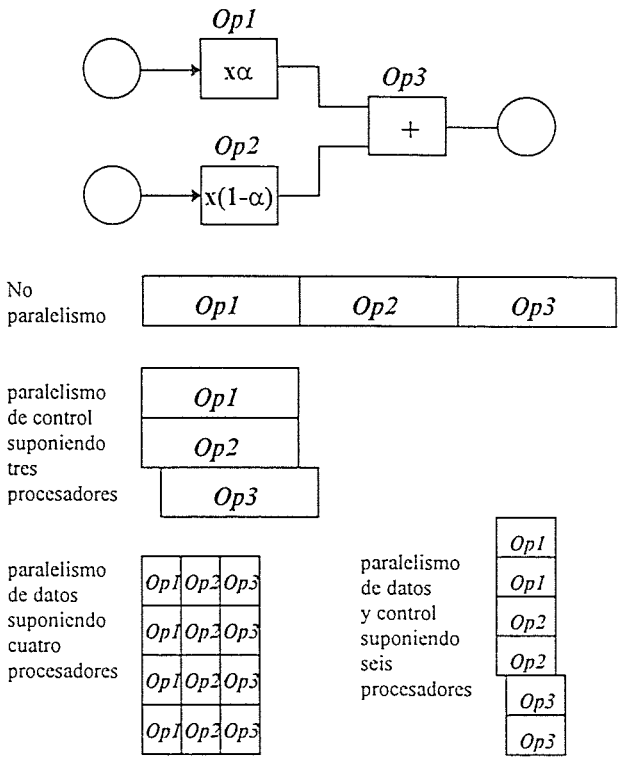


Fig.7 Paralelismo de control y datos

Debido pues a la necesidad de disponer de una granularidad pequeña para aprovechar mejor las capacidades del procesamiento en paralelo es por lo que en muchas aplicaciones la arquitectura de STREAMs se implementa sobre un hardware que dispone de unidades de procesamiento heterogéneas especializadas (stream processors) conectadas a través de un conmutador de altas prestaciones (crosspoint switch) para utilizar los recursos de memoria. Esta nueva arquitectura modifica en parte la estudiada hasta ahora para software pero en esencia mantiene el fundamento principal.

Un ejemplo de esta arquitectura lo constituye la arquitectura del sistema CHEOPS [8] desarrollada por el MIT Media Laboratory para procesamiento de secuencias de imágenes digitales en movimiento. Para ello los laboratorios han desarrollado un módulo de procesador propio denominado también Cheops que permite implementar hasta ocho procesadores de STREAMs (stream processors).

Es del todo conocido que los requerimientos de procesamiento para video digital son altamente exigentes sobre todo para aplicaciones de compresión/descompresión, procesamiento de la señal y efectos especiales. Incluso con la aparición de circuitos integrados y procesadores de propósito general más sofisticados que cubran las necesidades actuales, en el futuro se precisará de diez a mil veces

más operaciones por segundo y el correspondiente ancho de banda para transferencia con la memoria.

En busca de poder analizar las escenas de una imagen en movimiento de una forma más efectiva, el vídeo tenderá en los años venideros más hacia una representación segmentada no en base a bloques de pixeles, sino más bien a objetos o regiones determinadas por algoritmos interpretativos de la escena. El resultado será una colección de objetos y un escenario que describirá como formar la imagen. La arquitectura de STREAMs encaja perfectamente en este modelo de procesamiento de la información y proporciona para el tratamiento de la imagen y su diagnóstico un nuevo modelo que no solamente mejora los rendimientos del sistema facilitando el procesamiento en paralelo, sino que también permite realizar un análisis orientado a los objetos que facilita la interpretación de las imágenes en movimiento.

4.4 Sistemas Distribuidos

En el entorno de sistemas distribuidos la arquitectura de STREAMs tiene también su aplicación aunque sea aquí donde actualmente tenga menor impacto al quedar latente el principal inconveniente de su arquitectura. Las prestaciones que ofrecen los STREAMs son menos adecuadas que otras arquitecturas [9][10] más específicas para esos entornos que actualmente están implementando otros investigadores. También un inconveniente puede ser el tamaño que exige su implementación que hace más dificultosa su estructura distribuída.

Los STREAMs pueden facilitar la interoperatividad distribuida de sistemas e incluso sobre ellos se pueden diseñar sistemas distribuidos, pero hoy por hoy las prestaciones que proporciona no son comparables con las de otras arquitecturas más sencillas y adaptadas[9][10].

5. Tendencias

La arquitectura de STREAMs está ya muy implantada en el entorno de Sistemas Operativos tanto de tiempo compartido como en tiempo real. Las aplicaciones más comunes de este subsistema actualmente están enfocadas hacia el soporte de una arquitectura flexible y modular que permite la reutilización de código y la implementación de múltiples protocolos de comunicaciones encapsulados.

La tendencia actual en los S.O. es a soportar estructuras multiprocesador y es aquí donde los STREAMs pueden jugar un papel destacado. Ejemplos de ello son el S.O. Unix de SCO y Windows NT.

Es en los periféricos inteligentes donde quizás tenga mayor auge esta arquitectura a corto plazo. Proporciona numerosas ventajas tanto a los desarrolladores como a los usuarios. Permite la reutilización del código del desarrollador o incluso de terceros. Facilita la integración con los S.O.. Permite la incorporaciones de módulos del cliente. Mejora las prestaciones de los sistemas.

Hoy en día en el mundo de las telecomunicaciones aparecen numerosos protocolos que además viajan encapsulados unos dentro de otros. Esto exige a los sistemas que soporten todos los protocolos con la consecuente pérdida de rendimientos. Esto puede ser fácilmente simplificado y mejorado utilizando dispositivos inteligentes que implementen los protocolos de las redes. Un ejemplo puede ser el mercado de tarjetas de RDSI donde los fabricantes deben proporcionar soluciones de acceso a diferentes redes entre ellas TCP/IP.

En cuanto al procesamiento paralelo la tendencia clara en cuanto a STREAMs es su implementación en arquitecturas hardware específicas. Habrá que esperar las soluciones de los fabricantes.

6. Conclusiones

La arquitectura de STREAMS ha demostrado ya su consistencia en el entorno de los sistemas operativos y está demostrando su eficacia en la implementación de arquitecturas de dispositivos inteligentes mejorando sustancialmente los rendimientos de los sistemas y manteniendo la flexibilidad. Actualmente se está introduciendo en el entorno de procesamiento paralelo sobretodo en la implantación de S.O. multiprocesador.

Referencias

[1] Ritchie, D.M. , "A Stream Input-Output System". AT&T Bell Lab. Technical Journal,63,8,10-84 (1982)

[2] AT&T , "UNIX System V Release 3.2 STREAMS Programmer's Guide". *Prentice Hall 1989.*

[3] UNIX System Laboratories "STREAMS Modules and Drivers, UNIX SVR4.2". UNIX Press. Prentice Hall.

[4] Henrik Frystyk Nielsen "Data Flow using STREAMS"

[5] Stephen Rago "UNIX System V Network Programming" Addison Wesley.

[6] Thomas Herbert "STREAMS protocols and drivers in real-time embedded systems" ROCSLUG (1997).

[7] RNS "STREAMS on a Front-End Adaptor".

[8] M.Bove,J.Watlington"Cheops:A Reconfigurable Data-Flow System for Video Processing".MIT Media Lab.

[9] J.González, J.Ballesteros "The Inherently Distributed Adaptable Off microkernel" Univ. Carlos III (1997).

[10]J.Ballesteros, L.Fernández "The Network Hardware is the Operating System" Universidad Carlos III (1997).

Modelado del servicio de intermediación electrónica (brokerage) según el modelo de referencia de ODP: perspectiva de negocio

JUAN I. ASENSIO [†], JOSÉ I. MORENO [††] y VÍCTOR A. VILLAGRÁ [††]

[†] Dpto. de Teoría de la Señal, Comunicaciones e Ingeniería Telemática
E.T.S.I. de Telecomunicación, Universidad de Valladolid
C/Real de Burgos s/n, 47011 VALLADOLID
Correo electrónico: juaase@tel.uva.es

[††] Dpto. de Ingeniería de Sistemas Telemáticos
E.T.S.I. de Telecomunicación, Universidad Politécnica de Madrid
Ciudad Universitaria s/n, 28040 MADRID
Correo electrónico: {jmoreno,villagra}@dit.upm.es

Abstract:

The concept of Electronic Brokerage is based on the use of new Information Technologies so as to provide a service capable of facilitating and organising the relationship between the Supply and the Demand in an "Electronic Marketplace" environment. This new service will allow service providers to publish and advertise their offers and, at the same time, will help end-users to access the offered services and information in an easy and efficient way.

1. Introducción

En el contexto del futuro "*Mercado Global*", los servicios de intermediación (soportados por los comúnmente denominados "*brokers*" o intermediarios), jugarán un papel de vital importancia: poner en contacto Oferta y Demanda de una manera rápida, flexible y eficaz.

Un ejemplo de servicio de intermediación es el que se pretende ofrecer mediante la arquitectura que se está desarrollando en el ámbito del denominado proyecto ABS[1] ("*Architecture for Information Brokerage Service*" o "*Arquitectura para Servicios de Intermediación de Información*") perteneciente al programa ACTS ("*Advanced Communications Technologies and Services*" o "*Tecnologías y Servicios de Comunicaciones Avanzadas*") de la Comisión Europea.

El proyecto ABS tiene un triple objetivo:

1. **Definir y especificar una arquitectura de servicios de intermediación de información** basándose en conceptos de RM-ODP [1] ("Reference Model of Open Distributed Processing" o "Modelo de Referencia para Procesamiento Distribuido Abierto") y TINA-C [2] ("*Telecommunications Information Networking Architecture Consortium*" o "*Consorcio para la Arquitectura de Red de Información de Telecomunicaciones*").
2. **Diseñar e implementar varios prototipos** de sistemas de intermediación electrónica utilizando, entre otras tecnologías, CORBA [3] ("*Common Object Request Broker Architecture*" o "*Arquitectura Común de Intermediación de Peticiones de Objetos*") y JAVA.
3. Validar la arquitectura llevando a cabo pruebas de campo nacionales e internacionales en entornos reales de *Comercio Electrónico*.

El objetivo de esta comunicación es describir en profundidad el servicio de intermediación que se está desarrollando en el ámbito del proyecto ABS, tratando de definir con exactitud su papel dentro del entorno de *Comercio Electrónico*. Al mismo tiempo se intentará mostrar cómo la metodología de desarrollo utilizada en ABS es adecuada para el tipo de servicios que tendrán que coexistir en dicho entorno. Para todo ello, en la sección 2 se describirá con más detalles el servicio de intermediación electrónica en el contexto general de los servicios de soporte para *Comercio Electrónico*, en la sección 3 se introducirá la aproximación metodológica adoptada por ABS para especificar, diseñar e implementar su servicio de intermediación electrónica, en la sección 4 se profundizará en los resultados de una de las fases intermedias de desarrollo: el *modelo de negocio*, en el que se analizarán los agentes que integran el servicio y sus características particulares para, en la sección 5, introducir algunas de las características más importantes de la primera versión de la arquitectura ABS que ya ha sido completada. Por último, en la sección 6 se presentarán las principales conclusiones de este trabajo indicando, al mismo tiempo, cuáles son las fundamentales líneas de trabajo que aún tiene que abordar el consorcio ABS.

[1] Este trabajo está financiado parcialmente por la Comisión de la Unión Europea bajo el proyecto ABS, proyecto número 206 del programa ACTS. El consorcio ABS está compuesto por France Telecom-CNET (Francia), DeTeBerkom (Alemania), Telecom Finland (Finlandia), Portugal Telecom-CET (Portugal), Telis (Francia), VTT (Finlandia), NTUA (Grecia), DIT-UPM (España), Degriftour (Francia) y PoP (Alemania). Las opiniones presentadas aquí no representan necesariamente las del consorcio ABS.

2. El servicio de intermediación en el contexto de las plataformas para soporte del Comercio Electrónico

El Comercio se basa en el intercambio de bienes y servicios. Cuando alguien necesita algo, se crea una demanda en el Mercado. Por otra parte, si hay proveedores que ofrecen bienes y servicios, aparece una Oferta. Los agentes de la Demanda y la Oferta (los clientes y los proveedores, respectivamente) deben ponerse en contacto para llevar a cabo una determinada transacción comercial. Se necesita algún tipo de plataforma que soporte ese encuentro entre Oferta y Demanda y su interacción subsiguiente. Hace tiempo, esa plataforma consistía en un mercado físico pero, hoy en día, gracias a las Tecnologías de la Sociedad de la Información, ha surgido el denominado "*Mercado Electrónico*" que permite el correspondiente "*Comercio Electrónico*" entre las partes involucradas en una transacción, sin necesidad de desplazamiento físico.

Para que el "*Mercado Electrónico*" pueda soportar transacciones comerciales de forma adecuada, son necesarios los denominados "*Servicios de Mediación*" ("*Mediation Services*"). Un "*Servicio de Mediación*" se puede definir como aquel servicio que permite a los individuos y organizaciones obtener lo que necesitan mediante el intercambio de información y productos. Estos servicios actúan de soporte durante la fase de negociación de las intercambios comerciales y, posteriormente, proporcionan servicios logísticos y de transacción a dichas interacciones. La Fig. 1 muestra el concepto de "*Servicio de Mediación*" y su relación con la Oferta y la Demanda

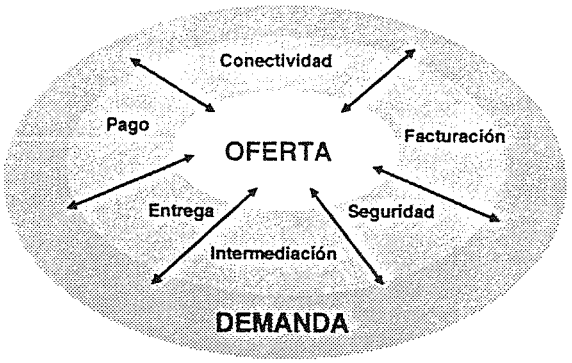


Fig. 1 Los "*Servicios de Mediación*" facilitan la interacción entre la Oferta y la Demanda en el ámbito del "Mercado Electrónico".

El servicio de intermediación electrónica ("*Electronic Brokerage Service*") es uno de esos "*Servicios de Mediación*". Su utilidad se pone de manifiesto en la primera fase en las relaciones entre Oferta y Demanda puesto que informa a los clientes (la Demanda) de un conjunto de proveedores (Oferta) que podrían satisfacer sus necesidades y los ayuda en la decisión de cuál de ellos es el más adecuado en base a una serie de criterios. Al mismo tiempo, ayuda a los proveedores en su afán de acercar las ofertas a los potenciales clientes. En otras palabras, pone a la Oferta y a la Demanda en contacto para que las transacciones comerciales puedan llevarse a cabo.

El objetivo principal de ABS es desarrollar una arquitectura de servicios de intermediación electrónica que ofrezcan un conjunto muy preciso de facilidades de valor añadido:

1. *Costes bajos para clientes y proveedores*: mediante el servicio de intermediación electrónica, un cliente no tendrá que visitar diferentes lugares para saber cuál es el proveedor que más le conviene y, además, los proveedores tendrán la oportunidad de acceder a nuevos tipos de clientes. Los intermediarios electrónicos pueden, al mismo tiempo, mantener bases de datos de clientes en las cuales almacenar información sobre sus preferencias para, de ese modo, reducir los costes de las búsquedas.
2. *Estructuración de las fuentes de información*: la eficiencia en el emparejamiento entre ofertas y demandas estará basada en la forma en que éstas están estructuradas en el intermediario electrónico. Además, la estructura que se obtenga permitirá a los clientes "navegar" a través del conjunto de ofertas disponibles, participando más activamente en el proceso de emparejamiento.
3. *Combinación de las fuentes de información*: el intermediario electrónico combinará diferentes fuentes de información para ofrecer al cliente una respuesta a su petición que le produzca un mayor grado de satisfacción. Es más, el intermediario electrónico podría obtener información acerca de los productos ofertados de otras fuentes diferentes a los propios proveedores de productos como, por ejemplo, clientes y evaluadores independientes.
4. *Presentación de la información*: cuando hay más de un proveedor que puede satisfacer la demanda de un cliente, el intermediario electrónico debe presentarle la información de la mejor manera posible para facilitar su toma de decisiones.
5. *Acceso homogéneo a fuentes de información heterogéneas*: el cliente debe percibir una visión homogénea de todos los proveedores que existen en el "*Mercado Electrónico*".
6. *Fiabilidad del contenido*: el intermediario electrónico mantendrá la información sobre ofertas consistente y actualizada. También supervisará los aspectos legales de la información y los productos que se intercambian en el "*Mercado Electrónico*".

7. *Confidencialidad*: cuando sea acordado, el intermediario electrónico no revelará las identidades de los participantes de una determinada transacción comercial.

La provisión de un servicio con las anteriores características implica el procesamiento de información dispersa geográficamente: información mantenida por diferentes proveedores, información de los diferentes usuarios del servicio, etc. El servicio de intermediación electrónica es claramente un *servicio distribuido*.

Para hacer frente a las especiales características de este tipo de servicios, el consorcio ABS adoptó una metodología de desarrollo cuya descripción es el objetivo del próximo apartado.

3. Metodología empleada

El consorcio ABS decidió utilizar los conceptos de modelado y desarrollo contenidos en la denominada *Arquitectura de computación de TINA-C* [1] (*"Telecommunications Information Networking Architecture Consortium"* o *"Consorcio para la Arquitectura de Red de Information de Telecomunicaciones"*). *TINA-C* es un consorcio de operadores de redes, proveedores de equipos de telecomunicación y fabricantes de ordenadores que, desde principios del año 1993, viene trabajando en la definición de una arquitectura software que, con la mayor independencia posible con respecto a la infraestructura de telecomunicaciones que la soporte, permita la introducción rápida y flexible de nuevos servicios avanzados de telecomunicación. Dicha arquitectura software deberá permitir igualmente, la gestión de todos esos servicios y de las redes que los sustentan [4,5].

La arquitectura de TINA-C ha aunado estándares y tecnologías existentes del mundo de las telecomunicaciones y de la informática creando un conjunto de conceptos y principios básicos para el desarrollo, operación y mantenimiento de software para aplicaciones de telecomunicación, aplicaciones que pueden ser centralizadas o pueden estar distribuidas en un entorno de procesamiento heterogéneo.

Debido a la elevada cantidad y complejidad de los conceptos contenidos en la arquitectura de TINA-C, ésta se descompone en cuatro subconjuntos que constituyen otras tantas arquitecturas:

- *La arquitectura de servicios*: define un conjunto de conceptos y principios necesarios para el desarrollo y gestión de servicios de telecomunicación.
- *La arquitectura de red*: define un conjunto de conceptos y principios útiles a la hora de desarrollar y gestionar los recursos de transmisión y conmutación que servirán de soporte a los servicios de TINA.
- *La arquitectura de gestión*: indica cómo se han de desarrollar las aplicaciones software que permitan una gestión eficiente de todos los elementos de la arquitectura de TINA: servicios (incluidos los propios servicios de gestión), componentes software, equipos de telecomunicación, etc.
- *La arquitectura de computación*: contiene los conceptos y principios básicos necesarios para especificar, diseñar e implementar aplicaciones software distribuidas indicando, al mismo tiempo, cuál es el entorno de soporte que dichas aplicaciones requieren.

Estas cuatro arquitecturas están completamente interrelacionadas. Así, por ejemplo, la arquitectura de servicios necesita un determinado grado de abstracción de los recursos de red, abstracción proporcionada por la arquitectura de red. La arquitectura de gestión se especializa para gestionar componentes de la arquitectura de servicios, de red y de computación. Y, lo que es más importante desde el punto de vista de la metodología de desarrollo empleada por el consorcio ABS, la arquitectura de computación define conceptos que han de ser utilizados para desarrollar aplicaciones dentro de la arquitectura de gestión, de red y de servicios. Todo software que se utilice dentro de TINA, ha de satisfacer los requisitos impuestos por los conceptos y principios contenidos dentro de la arquitectura de computación.

Esos conceptos y principios son los que ha adoptado el consorcio ABS para desarrollar su servicio de intermediación electrónica. No obstante, como se verá más adelante, esta no es la única influencia que el consorcio ABS ha recibido de la arquitectura de TINA-C.

Los conceptos de desarrollo software contenidos en la arquitectura de computación de TINA son una herencia, prácticamente directa, de los contenidos en el estándar *Modelo de Referencia para Procesamiento Distribuido Abierto* [2] (*RM-ODP* o *Reference Model for Open Distributed Processing*) de ISO/IEC e ITU-T.

RM-ODP proporciona un *marco de trabajo* para la descripción y estandarización de sistemas ODP o *Sistemas de Procesamiento Distribuido Abierto*: sistemas de procesamiento distribuido (sistemas de procesamiento de información cuyos componentes pueden estar

situados en diferentes lugares y cuya comunicación está sujeta a posibles fallos o retrasos) que cumplen los preceptos de las normas de RM-ODP. Se puede decir que RM-ODP es una evolución del Modelo de Referencia OSI para abarcar todas aquellas particularidades de los sistemas de procesamiento distribuido [6].

RM-ODP es un modelo de referencia de alto nivel que no trata en absoluto cuestiones relativas a implementación y que no prescribe la utilización de ninguna tecnología particular. El objetivo último es utilizar el marco proporcionado por RM-ODP para desarrollar nuevos estándares, de más bajo nivel, que pertenezcan a alguna de estas dos áreas [7]:

- Estándares de Modelos de referencia específicos que particularicen los conceptos y principios de RM-ODP a algún campo de aplicación determinado.
- Estándares de funciones de soporte de sistemas de procesamiento distribuido, previamente identificados en RM-ODP, pero incluyendo todos aquellos detalles y particularidades específicas de un determinado campo de aplicación.

La arquitectura de TINA se puede considerar englobada dentro del primer grupo de estándares: los conceptos de modelado agrupados en su arquitectura de computación no son más que una particularización de RM-ODP al campo de las redes y servicios de telecomunicación.

El marco de trabajo definido en RM-ODP está compuesto de:

- Cinco *perspectivas o puntos de vista* desde los que especificar un sistema ODP
- Unos *"lenguaje de perspectiva"* que definen conceptos necesarios para especificar un sistema ODP desde la perspectiva correspondiente.
- Especificación de unas *funciones* de soporte para los sistemas ODP.
- Unas *"prescripciones de transparencia"* que indican cómo utilizar las funciones de soporte de ODP para conseguir transparencia de distribución.

Los dos primeros aspectos indican cómo especificar los sistemas ODP mientras los dos últimos puntos están más relacionados con las características que deben poseer los entornos en los se ejecuten las aplicaciones de procesamiento distribuido: los denominados *Entornos de Procesamiento Distribuido* o *DPE* (*Distributed Processing Environment*).

La metodología de desarrollo adoptada por el consorcio ABS se basa en la utilización de las cinco perspectivas de ODP (particularizadas por la arquitectura de computación de TINA) y sus correspondientes lenguajes para especificar completamente la arquitectura del servicio de intermediación electrónica. Esas cinco perspectivas son las siguientes:

- *Perspectiva de Negocio* (*Enterprise Viewpoint*): se centra en el propósito y ámbito del sistema al efecto y las políticas por las que se rige.
- *Perspectiva de Información* (*Information Viewpoint*): se centra en los aspectos semánticos de la información en el sistema y cómo ésta es procesada.
- *Perspectiva Computacional* (*Computational Viewpoint*): se utiliza para especificar la funcionalidad del sistema. Dicha funcionalidad estará soportada por un conjunto de objetos que interactúan entre sí y que son susceptibles de estar distribuidos.
- *Perspectiva de Ingeniería* (*Engineering Viewpoint*): se centra en la infraestructura que se necesita para soportar la distribución de los objetos computacionales.
- *Perspectiva de Tecnología* (*Technology Viewpoint*): se centra en la elección de la(s) tecnología(s) necesarias para soportar todo el sistema.

La especificación de un sistema de procesamiento distribuido bajo estas cinco perspectivas se puede equiparar a la aplicación de las fases de desarrollo típicas de la Ingeniería Software de la forma que indica la Fig. 2 [8].

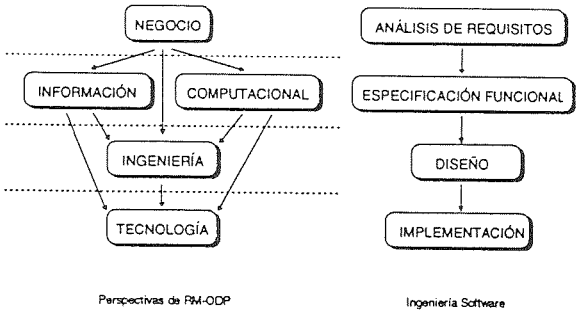


Fig. 2 Relación entre las perspectivas de RM-ODP y las fases de desarrollo software.

El objetivo de esta comunicación es, como se dijo en el apartado 1, describir en detalle la especificación que, desde la perspectiva de negocio, se ha hecho del servicio de intermediación electrónica.

4. Intermediación electrónica: perspectiva de negocio

Como ya se ha dicho, el análisis de un sistema ODP desde la perspectiva de negocio tiene como objetivo primordial conocer el sistema dentro del entorno en el que va a operar, su

propósito y cuáles son las políticas que gobiernan su funcionamiento.

Aplicando la perspectiva de negocio se obtendrá un modelo (el modelo de negocio) en el que el propio sistema ODP, los usuarios en sus diferentes papeles, el entorno y las políticas comerciales y de gestión aparecen como objetos o entidades interrelacionadas. Mediante este modelo, se separan los objetivos del sistema de la forma en que estos son llevados a cabo [7].

Conceptos fundamentales en el análisis de un sistema ODP y su entorno desde la perspectiva de negocio son:

- *Actor (stakeholder o actor):* es la abstracción de un individuo o un grupo de individuos que desempeñan uno o varios *papeles* en el entorno del sistema.
- *Papel (role)*: cada uno de los posibles tipos de relación que un actor puede tener con el sistema que se está describiendo, su entorno y el resto de actores. Un actor puede desempeñar más de un papel.
- *Dominio*: agrupación de actores que interactúan entre sí para alcanzar un determinado objetivo.
- *Política:* conjunto de reglas que describen los derechos y obligaciones de los actores, de acuerdo a sus papeles, en el entorno de actuación del sistema descrito. En un modelo de negocio, múltiples políticas pueden ser descritas haciendo referencia a aspectos tales como utilización de recursos, relación entre actores, estructuración de estos, etc.

El consorcio ABS ha aplicado todos estos conceptos, para obtener el correspondiente modelo de negocio del servicio de intermediación electrónico. En los siguientes subapartados se describirán los principales componentes de este modelo [9] que se pueden resumir en los actores y dominios mostrados en la Fig. 3.

4.1 Actores y sus correspondientes papeles en el ámbito del servicio de intermediación electrónica

El consorcio ABS ha identificado seis clases principales de actores en el ámbito del servicio de intermediación electrónica. Al mismo tiempo, y desde esa misma perspectiva, se han estudiado los principales requisitos demandados por cada tipo de actor (requisitos referentes a su propio comportamiento y al de los actores de las otras clases).

4.1.1 Usuario del Servicio de Intermediación

El *Usuario del Servicio de Intermediación* es una entidad (ser humano, máquina o aplicación) que utiliza el servicio de intermediación para

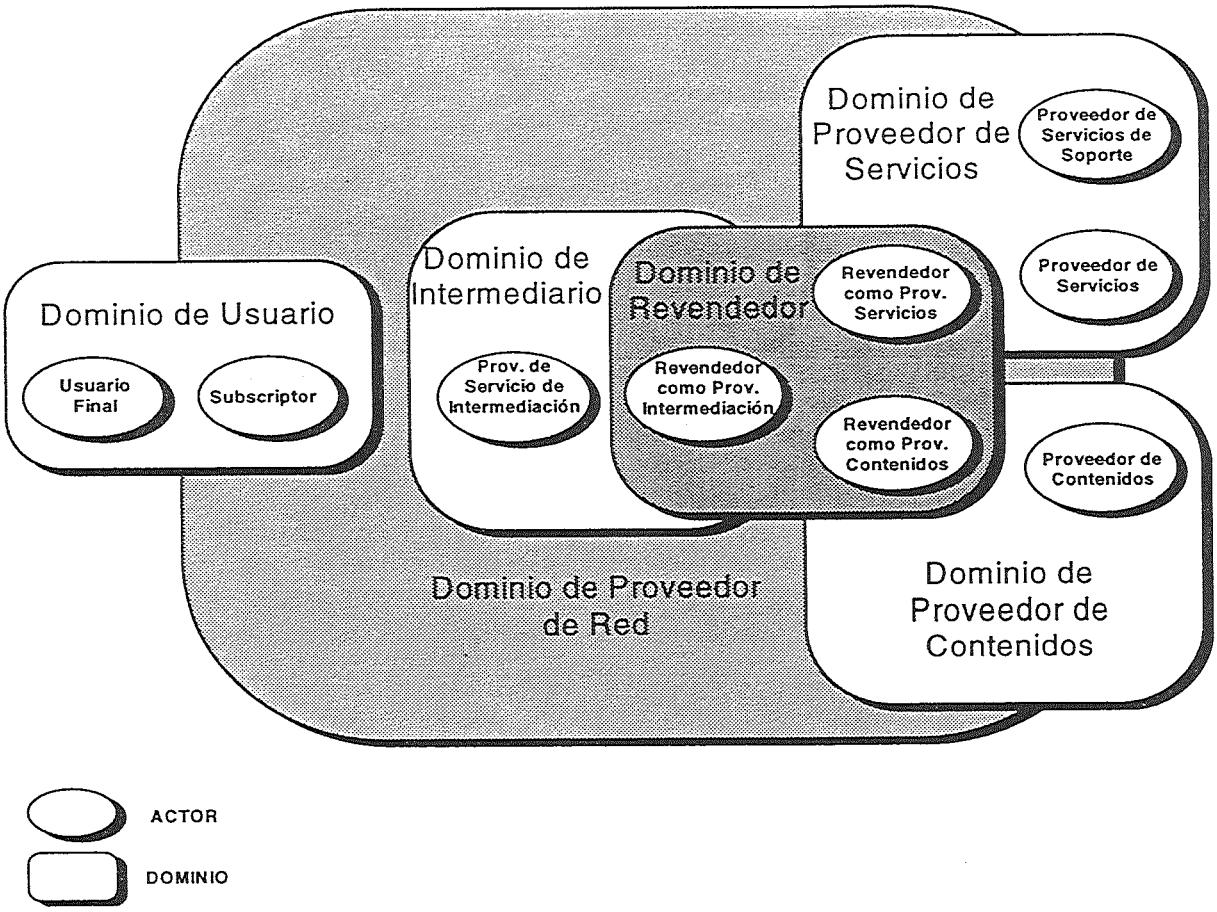


Fig. 3 Actores y Dominios Administrativos en el Modelo de Negocio del Servicio de Intermediación Electrónica.

satisfacer una serie de requisitos. Los *Usuarios del Servicio de Intermediación* se pueden dividir en dos grandes grupos:

- *Subscriptores*: son *Usuarios del Servicio de Intermediación* cuyas características de acceso y utilización del servicio han sido previamente reflejadas en un contrato. Un *subscriptor* puede ser un individuo o una organización.
- *Usuarios finales:* son *Usuarios del Servicio de Intermediación* que no están sujetos a las cláusulas de un contrato.

El *Usuario del Servicio de Intermediación* puede desempeñar los papeles de *"Cliente"* y *"Usuario"* dependiendo de si puede o no, respectivamente, negociar las características del servicio ofrecido por el proveedor (el papel desempeñado está íntimamente relacionado con el hecho de que estemos hablando de *Subscriptores* o *Usuarios finales*).

Los principales requisitos demandados por los *Usuarios del Servicio de Intermediación* son los siguientes:

- *Calidad de Servicio*: los *Usuarios del Servicio de Intermediación* necesitan conocer la fiabilidad de la información que el servicio les ofrece. Como mínimo, será imprescindible que se pueda conocer cuál es la fuente de información (este conocimiento puede ser suficiente para conocer con bastante precisión la fiabilidad y calidad de dicha información).
- *Seguridad:* los Usuarios del Servicio de Intermediación demandan los mismos requisitos de seguridad a nivel de aplicación que los demás actores involucrados en el "*Comercio Electrónico*": confidencialidad, autenticación, integridad y no repudiación por parte de origen/receptor.
- *Valor añadido del servicio:* obviamente, los *Usuarios del Servicio de Intermediación* únicamente estarán dispuestos a pagar por el servicio si la información que se les ofrece es imposible o extremadamente difícil de conseguir por otros medios.
- *Simplicidad de uso:* los *Usuarios del Servicio de Intermediación* querrán que el acceso al servicio esté integrado con otras aplicaciones del ámbito del *"Comercio Electrónico"* y que dicho acceso sea sencillo e intuitivo.

4.1.2 Proveedor del Servicio de Intermediación

El *Proveedor del Servicio de Intermediación* es una entidad que ofrece información a los *Usuarios del Servicio de Intermediación* acerca de servicios o contenidos mantenidos por terceras entidades.

El Proveedor del Servicio de Intermediación es una extensión del *"Comerciante de Objetos"* (*"Trader"* en inglés) que en el campo del Procesamiento Distribuido permite establecer enlaces dinámicos entre diferentes objetos [10]. En el servicio de intermediación electrónica, se traslada el concepto del *"Comerciante de Objetos"* a la intermediación de bienes y servicios estableciendo "enlaces" dinámicos entre clientes (demanda) y proveedores (oferta).

El *Proveedor del Servicio de Intermediación* puede desempeñar tres papeles diferentes:

- *"Cliente"* de los *Proveedores de Servicios* y de los *Proveedores de Contenidos* (que se describirán posteriormente) con los que establece una relación contractual para obtener información sobre los bienes y servicios que son objeto de la intermediación propiamente dicha.
- *"Operador"*: el *Proveedor del Servicio de Intermediación* opera y gestiona su propio servicio.
- *"Proveedor"*: el *Proveedor del Servicio de Intermediación* proporciona el servicio de intermediación a los *Usuarios del Servicio de Intermediación.*

En cuanto a los requisitos del *Proveedor del Servicio de Intermediación*, podemos destacar:

- *Seguridad*: los requisitos de seguridad de los *Usuarios del Servicio de Intermediación* son válidos también para este actor.
- *Información de los Usuarios*: El Proveedor del Servicio de Intermediación necesita información acerca de las características y las preferencias de los *Usuarios del Servicio de Intermediación* para poder estructurar de manera adecuada los resultados que les ofrece.

4.1.3 Proveedor de Servicios

El *Proveedor de Servicios* es una entidad que proporciona servicios de telecomunicación, servicios de información, servicios de aplicación, etc. que, en todos los casos, son servicios disponibles en la red de comunicaciones que soporta el servicio de intermediación. Podríamos dividir todos estos servicios en dos grandes grupos:

- Servicios que pueden ser motivo de intermediación (ofertas que se tratan de emparejar con ciertas demandas) como, por ejemplo: vídeo bajo demanda, video-conferencias, etc.
- Servicios de soporte: como, por ejemplo: seguridad, pago, certificación, comunicación, etc.

El *Proveedor de Servicios* puede desempeñar todos estos papeles en el ámbito del servicio de intermediación:

- *"Cliente"*: puede tener relaciones contractuales con los *Proveedores de Contenidos* (analizados

posteriormente) para obtener información necesaria para su servicio.
- "Operador": el *Proveedor de Servicios* opera y gestiona su propio servicio
- *"Proveedor"*: el *Proveedor de Servicios* ofrece su servicio a los *Usuarios del Servicio de Intermediación* y a los *Proveedores del Servicio de Intermediación* y *Revendedores* (que serán analizados en una subsección posterior).

En cuanto a los principales requisitos de los *Proveedores de Servicios*:
- *Seguridad*: los requisitos de este actor son los mismos que los descritos para los *Usuarios del Servicio de Intermediación.*
- *Calidad de Servicio*: los *Proveedores de Servicios* exigirán a los *Proveedores del Servicio de Intermediación* una cierta calidad en el servicio ofrecido puesto que de ésta dependerá el número de clientes con los que podrá contactar y, por tanto, el volumen de negocio.
- *Valor añadido del servicio*: lo que un Proveedor de Servicios estará dispuesto a pagar por el servicio de intermediación dependerá de la demanda de mercado de sus servicios y de su interés en promocionar determinados servicios.
- *Simplicidad de uso*: el Proveedor de Servicios no estará dispuesto a alterar sus métodos de trabajo internos para adaptarse al servicio de intermediación. Todos los cambios que, a ese respecto, haya que llevar a cabo deben estar plenamente justificados y recompensados.

4.1.4 Proveedor de Contenidos

El Proveedor de Contenidos es la entidad que ofrece bienes o información, susceptible de formar parte de transacciones comerciales, a usuarios o clientes, mediante la utilización del servicio de intermediación.

El único papel que desempeña este actor es el de *"Proveedor"* puesto que ofrece sus contenidos al resto de los actores que participan en un escenario de intermediación electrónica.

En cuanto a los requisitos, estos son básicamente los mismos que tenía el *Proveedor de Servicios.*

4.1.5 Revendedor (Retailer)

El *Revendedor* (*Retailer*, en inglés), es una entidad que proporciona servicios y contenidos a un conjunto de usuarios y clientes pero con la particularidad de que son servicios y contenidos comprados a terceros. La actividad del *Revendedor* se lleva a cabo previo acuerdo contractual con los *Proveedores de Servicios* y los *Proveedores de Contenidos.* Es posible igualmente que el *Revendedor* no llegue a comprar los servicios y contenidos que oferta pero que tenga un acuerdo con los *Proveedores de Servicios* y los *Proveedores de Contenidos* correspondientes para actuar en su nombre a todos los efectos.

El *Revendedor* incorpora funciones de intermediación para ayudar a los clientes y usuarios a seleccionar la oferta que más se adecua a sus necesidades. La diferencia entre *Revendedores* y *Proveedores del Servicio de Intermediación* estriba en que estos últimos no intervienen en las transacciones comerciales que, tras el proceso de intermediación, se establecen entre los *Usuarios del Servicio de Intermediación* y los *Proveedores de Servicios y Proveedores de Contenidos.*

El Revendedor puede desempeñar los siguientes papeles:
- *"Cliente"*: el *Revendedor* puede tener contratos firmados con los *Proveedores de Servicios* y *Proveedores de Contenidos* para obtener bienes y servicios que posteriormente venderá.
- *"Operador"*: el *Revendedor* opera y gestiona su propio servicio.
- *"Proveedor"*: proporciona bienes y servicios a los *Usuarios del Servicio de Intermediación.*

En cuanto a los requisitos del *Revendedor*, estos se pueden considerar como una combinación de los de los actores anteriormente descritos puesto que un *Revendedor*, dependiendo de las circunstancias, desempeña los papeles propios de los otros tipos de actores. Podemos hablar de los siguientes requisitos:
- *Flexibilidad:* el Revendedor no sólo ofrece información sobre una gran diversidad de servicios y contenidos sino que, además, él mismo debe estar preparado para llevar a cabo transacciones comerciales con todos esos tipos de bienes. Este hecho requiere que los métodos de trabajo internos del *Revendedor* sean fácilmente adaptables a cualquier área de negocio.
- *Calidad de Servicio*: si el *Revendedor* va a utilizar las funciones de intermediación, los requisitos impuestos por los *Usuarios del Servicio de Intermediación al Proveedor del Servicio de Intermediación* son aplicables exactamente igual en este caso.
- *Seguridad*: además de los requisitos de seguridad demandados por los demás tipos de actores, debido al papel activo que juega el *Revendedor* en las transacciones comerciales tras finalizar la fase de intermediación, los aspectos relacionados con la *Redundancia* y *Copias de Seguridad* de la información procesada por el Revendedor son de una importancia primordial.
- *Valor añadido del servicio:* todo lo dicho sobre este tipo de requisitos para otras clases de

actores es igualmente aplicable a los *Revendedores*.

- *Simplicidad de uso*: la incorporación de funciones de intermediación a los Revendedores debe evitar, en la medida de lo posible, incrementos en la carga de los recursos de procesamiento y comunicación disponibles. Del mismo modo, los métodos y la organización de trabajo interna deben ser afectados lo menos posible.

4.1.6 Proveedor de Red

El *Proveedor de Red* es una entidad que proporciona todos las funcionalidades de interconexión necesarias para que los otros actores puedan alcanzar sus objetivos. Aunque un *Proveedor de Red* se podría considerar como un *Proveedor de Servicios* se ha preferido considerarlo como un tipo de actor diferente ya que los servicios que ofrece no son específicos de la intermediación.

Un Proveedor de Red pues desempeñar los siguiente papeles:

- *"Operador"*: el *Proveedor de Red* opera y gestiona su propia red.
- *"Proveedor"*: el *Proveedor de Red* ofrece servicios de interconexión a los otros actores.

4.2 Dominios en el modelo de negocio

Como ya se ha comentado, un Dominio, desde la perspectiva de negocio, es la agrupación de varios actores que interactúan entre sí para alcanzar un determinado objetivo. Los dominios pueden ser *Administrativos* (formados por criterios de organización, políticos, geográficos, etc.) o *Funcionales* (atendiendo a las funcionalidades desempeñadas por los actores involucrados).

Desde la perspectiva de negocio tienen más importancia los dominios administrativos puesto que en ellos se pueden plasmar mejor los objetivos empresariales que se pretenden alcanzar mediante la utilización del sistema bajo estudio.

En ABS, se han definido los siguientes dominios administrativos para el servicio de intermediación (dicho dominios aparecen en la Fig. 3):

- **Dominio de Usuario**: en este dominio se agrupan los *Usuarios del Servicio de Intermediación* en todos sus papeles.
- **Dominio de Intermediario**: es el dominio administrativo asociado al *Proveedor del Servicio de Intermediación.*
- **Dominio de los Proveedores de Servicio**: en este dominio se reúnen los *Proveedores de Servicio*. En la figura aparecen dos tipos de *Proveedores de Servicio*: aquellos que proporcionan servicios de soporte y aquellos que proporcionan servicios sujetos a intermediación.
- **Dominio de Proveedor de Contenidos:** es el dominio administrativo asociado con los *Proveedores de Contenido.*
- **Dominio de Revendedor:** es el dominio administrativo asociado a los *Revendedores*. En la Fig. 3, la superposición de este domino con otros diferentes trata de expresar cómo el *Revendedor* puede comportarse de diferentes maneras dependiendo de las circunstancias.

Con la identificación de los tipos de actores (y sus requisitos) y los correspondientes dominios administrativos, el modelo de negocio del servicio de intermediación se puede considerar suficientemente detallado como para proceder con el análisis de dicho servicio bajo otras perspectivas.

4.3 Relación con el modelo de negocio de TINA

Recientemente, el consorcio TINA ha publicado la última versión de su modelo de negocio [11]. Es importante resaltar la similitud de dicho modelo con el desarrollado dentro del consorcio ABS para la arquitectura del servicio de intermediación electrónica. De hecho hay una correspondencia prácticamente biunívoca entre los tipos de actores definidos para el servicio de intermediación de ABS y los papeles (*roles*) definidos en el modelo de negocio de TINA:

- Los *Usuarios del Servicio de Intermediación* de ABS son similares a los *"Consumidores"* (*consumers*) definidos en TINA.
- Los actores en el *Dominio del Intermediario* se pueden emparejar con el papel de intermediario definido en TINA.
- En TINA también se contempla el papel de los *Revendedores*.
- Los *Proveedores de Servicios* y *los Proveedores de Contenidos* de ABS se puede hacer corresponder con el papel de *Proveedores de Servicio* de TINA.
- El *Proveedor de Red* de ABS sería idéntico al *Proveedor de Conectividad* que aparece en el modelo de negocio de TINA.

Estas similitudes no hacen sino validar el modelo de negocio desarrollado en ABS indicando, además, que el servicio de intermediación electrónica es un servicio que se podría diseñar e implementar de acuerdo a los principios contenidos en las diferentes arquitecturas de TINA-C (en el consorcio ABS únicamente se aplican los principios de la arquitectura de computación y algunos conceptos aislados de otras arquitecturas como se verá posteriormente).

5. Arquitectura del Servicio de Intermediación

Aunque la presente comunicación trata de centrarse en los aspectos referentes al modelo de negocio del servicio de intermediación, en esta sección se va a intentar dar algunas pinceladas de los resultados de subsiguientes etapas de modelado (siempre según las perspectivas del Modelo de Referencia de ODP). En concreto, el consorcio ABS ya ha finalizado los modelos de información y computacional de una primera versión de la arquitectura del servicio de intermediación.

Dichos modelos son objeto de las siguientes subsecciones.

5.1 Intermediación electrónica: perspectiva de información

El modelo de información trata de describir qué información está contenida en el intermediario electrónico y cómo ésta se procesa. Básicamente, el intermediario electrónico mantiene información acerca de los *Usuarios del Servicio de Intermediación* (el denominado *"Modelo de Información de Negocio"*), información que trata de modelar las diferentes áreas de negocio que se soportan (*"Dominio Conceptual"*), información sobre los contenidos de los proveedores (*"Dominio de Recursos"*) e información acerca de *Proveedores de Servicio* que proporcionan servicios auxiliares al servicio de intermediación (*"Modelos Auxiliares"*). Los *"Dominios de Recursos"* y los *"Dominios Conceptuales"* constituyen el denominado *"Modelo Núcleo"* del servicio de intermediación. Todos estos submodelos se pueden apreciar en la Fig. 4.

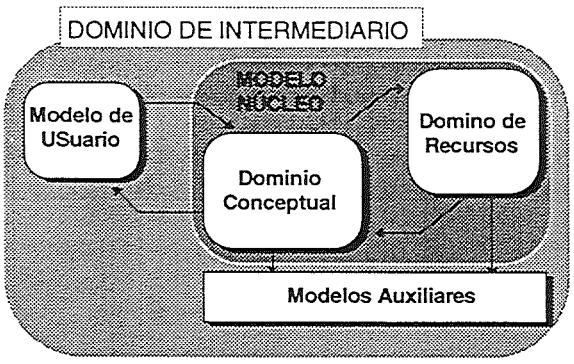


Fig. 4 Relación entre los submodelos de información del servicio de intermediación electrónica.

Un aspecto clave del modelo de información es la denominada *"Red Conceptual"*. Es parte del *"Dominio Conceptual"* y se puede definir como una estructura multi-relacional de nodos y relaciones entre nodos. Cada uno de esos nodos representa un concepto de la vida real. Un conjunto de estos conceptos, junto con las correspondientes relaciones, se pueden proyectar en las denominadas *"Perspectivas"* que no son más que una abstracción de una particular área de negocio.

La Fig. 5 muestra un esquema de en qué consiste la *"Red Conceptual"*. En dicho esquema se puede apreciar la *"Red Conceptual"* (RC) en la parte superior. Los conceptos de la RC se agrupan de acuerdo a diferentes áreas de negocio. Cada grupo de conceptos (cada área de trabajo) constituye una *"Perspectiva"* formada por *"Objetos Proyectados"* (OP, denominados *"Shadow Objects"* en inglés). Un OP se podría definir como la proyección de un concepto sobre una *"Perspectiva"*. Un mismo concepto de la RC se podría proyectar sobre diferentes *"Perspectivas"* dando lugar a diferentes OP. De esa forma, por ejemplo, Valladolid se podría proyectar sobre la *"Perspectiva"* de *"Vuelos en Europa"* como un OP que hace referencia a posibles destinos. Del mismo modo, se podría proyectar sobre la *"Perspectiva"* de *"Turismo Cultural en España"* como un posible sitio a visitar. Los OPs de diferentes *"Perspectivas"* están relacionados por los denominados *"Puentes"* que permiten relacionar diferentes áreas de negocio (en el ejemplo, *"Vuelos en Europa"* con *"Turismo Cultural en España"*).

Asociados a los OPs se encuentran los *"Recursos"* (componentes del *"Domino de Recursos"*) que describen los contenidos y las características de los proveedores involucrados en el servicio de intermediación. Dicho de otro modo, los *"Recursos"* contienen información sobre cómo acceder a la información que puede satisfacer la demanda de los usuarios.

Este modelo de información cuenta con dos ventajas sobresaliente:

- Es un modelo general: las características de la *"Red Conceptual"* permiten añadir nuevas áreas de negocio de una manera muy sencilla.
- Es un modelo flexible: permite "navegar" por una determinada área de negocio, pasar de un área a otra mediante los *"Puentes"*, y, además de navegar, se permite realizar búsquedas a partir de las cuales el intermediario electrónico combinará diferentes fuentes de información para encontrar la respuesta que más satisfaga al usuario. El usuario podrá pasar del modo de navegación al de búsqueda o viceversa cuando así lo desee.

5.2 Intermediación electrónica: perspectiva computacional

La Fig. 6 muestra el modelo computacional del servicio de intermediación electrónica (el "núcleo" de la arquitectura de ABS) en el que se trata de especificar los objetos que soportan las diferentes funcionalidades del intermediario electrónico. Dicho modelo incorpora diferentes elementos de:

- La *Arquitectura de Servicios* de TINA-C [12]: se pretende utilizar los componentes de la

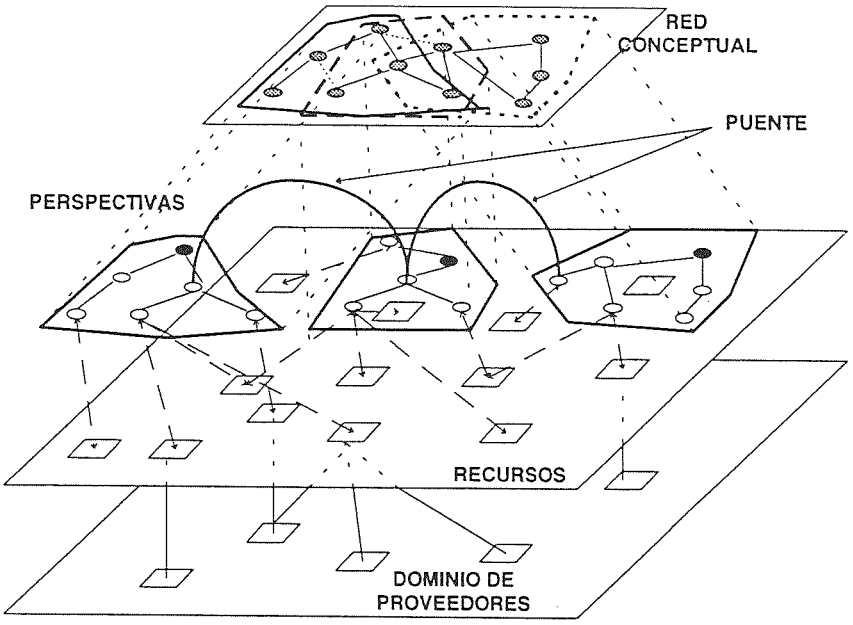


Fig. 5 La "Red Conceptual" del Modelo de Información del Servicio de Intermediación Electrónica.

denominada "Sesión de acceso" de la arquitectura de servicios de TINA-C.

- El *Modelo de Referencia de OMG EC-DTF* [13] (el *Object Management Group Electronic Commerce Task Force* o Grupo de Trabajo sobre Comercio Electrónico del OMG u Object Management Group) : algunos de los componentes del modelo reúnen las funcionalidades de los bloques de navegación y gestión de servicio de las denominadas *"Facilidades EC-DTF"*. También algunos componentes de las denominadas *"Facilidades Comunes"* del modelo de referencia de EC-DTF aparecen reflejadas en los componentes funcionales de la arquitectura de ABS.

Daremos a continuación una breve descripción de los bloques funcionales que constituyen el modelo de computación de la arquitectura del servicio de intermediación de ABS:

- CNM (*Conceptual Network Manager o Gestor de la Red Conceptual*): es responsable de la gestión de la Red Conceptual y de las

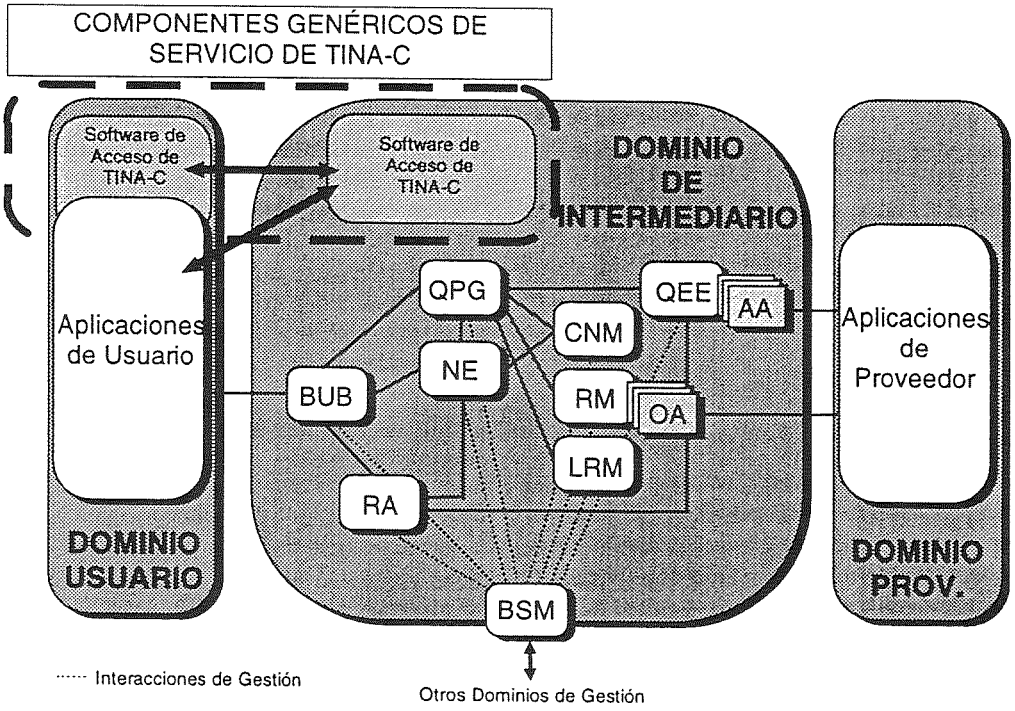


Fig. 6 Bloques Funcionales de la Arquitectura del Servicio de Intermediación Electrónica de ABS.

"Perspectivas". Podría decirse que éste es el bloque fundamental de toda la arquitectura.

- **RM** (*Resource Manager* o *Gestor de Recursos*) y **OA** (*Offer Agents* o *Agentes de Ofertas*): el **RM** gestiona las descripciones de las ofertas (información sobre los proveedores y datos almacenados localmente). Los **OAs** proporcionan capacidad de registro de ofertas, por parte de los proveedores, en el **RM**.
- **LRM** (*Local Resource Manager* o *Gestor de Recursos Locales*): es el responsable de la gestión de todas las bases de datos locales en las que se encuentran almacenada información sobre la información y los bienes ofertados por los proveedores.
- **BUB** (*Broker User Block* o *Bloque Intermediario-Usuario*): representa al usuario en el dominio del intermediario. Es el extremo, en el dominio del intermediario, de una sesión de acceso por parte de un usuario. También es responsable de discernir si la consulta de un usuario corresponde a una búsqueda o a una operación de navegación. Otra funcionalidad de este bloque es el mantenimiento de información sobre los usuarios (los denominados *"Perfiles de Usuarios"*).
- **QPG** (*Query Plan Generator o Generador de Planes de Peticiones*): su misión es identificar el concepto y la *"Perspectiva"* de la *Red Conceptual* que mejor se ajusta a las peticiones de un usuario y, tras comprobar si los datos almacenados por el **LRM** son o no de utilidad, generar un plan en el que se indica qué proveedores podrían satisfacer dicha petición. Ese plan se pasará al **QEE** para que éste acceda al dominio de los proveedores correspondientes y, de este modo, obtener información adicional.
- **QEE** (*Query Execution Engine o Sistema de Ejecución de Peticiones*) y **AA** (*Access Agents* o *Agentes de Acceso*): el **QEE** tiene como responsabilidad ejecutar el plan generado por el **QPG**. Para ello, y utilizando información disponible acerca de los Proveedores, configurará los *Agentes de Acceso* para que estos accedan al dominio de dichos *Proveedores*. La misión de los **AA** será recoger aquella información que pueda ser utilizada para responder a la petición del usuario. En este punto se puede ver cómo la información que maneja el intermediario no es estática sino que está completamente actualizada ya que se obtiene, en la mayor parte de las ocasiones, directamente del proveedor de servicios o contenidos donde se genera.
- **RA** (*Result Assistant* o *Asistente de Resultados*): su misión es procesar los resultados para mostrarlos de la manera más conveniente al usuario que los ha pedido. Dicho procesamiento estará basado en la información contenida en los *"Perfiles de Usuario"* gestionados por el **BUB**.
- **NE** (*Navigation Engine* o *Sistema de Navegación*): es el módulo responsable de la gestión de las peticiones de navegación por el *"Dominio Conceptual"* por parte del usuario.
- **BSM** (*Broker Service Manager*): es el módulo responsable de la gestión del servicio de intermediación electrónica, de la aplicación que implementa el intermediario electrónico y del sistema(s) sobre los que se ejecuta dicha aplicación. Permite, además, llevar a cabo dicha gestión desde otros dominios (*SNMP, gestión OSI*, etc.).

6. Conclusiones

En esta comunicación se ha descrito un nuevo servicio de información avanzado que jugará un papel fundamental en el futuro *"Comercio Electrónico"*: el *Servicio de Intermediación Electrónica* que se está desarrollando en el contexto del proyecto ABS.

La utilización de las perspectivas definidas en el modelo de referencia de ODP, junto con los matices incorporados por la arquitectura de computación de TINA-C, se está mostrando como una metodología de desarrollo válida en el caso del servicio de intermediación electrónica de ABS. Hasta el momento, el consorcio ABS ha finalizado los modelos correspondientes a las perspectivas de negocio, información y computacional de la primera versión de un prototipo de servicio de intermediación electrónica que se prevé esté finalizado en octubre de 1997. Con dicho prototipo, cuyas labores de implementación comenzaron en abril de 1997, y que está basado en tecnologías como JAVA y CORBA, se pretende llevar a cabo una serie de pruebas de campo en Portugal, Francia y Alemania con usuarios y proveedores reales en diferentes áreas de negocio. Los resultados de dichas pruebas servirán para refinar la primera versión de la arquitectura de tal forma que se incorporen todos los servicios de valor añadido que son deseables en el servicio de intermediación electrónica y que también se han detallado en esta comunicación.

Otro aspecto a considerar en el futuro es la utilización del nuevo *"Idioma de Negocio"* (*"Enterprise language"*) cuya estandarización se está llevando a cabo en el seno de ITU e ISO y que pretende ofrecer una serie de conceptos y principios mucho más completos que los contenidos en el Modelo de Referencia de ODP. Dichos conceptos y principios permitirán modelar sistemas ODP, desde la perspectiva de negocio, con un mayor nivel de detalle y riqueza semántica.

Referencias

[1] TINA Consortium, "Overall Concepts and Principles of TINA". Versión 1.0. Febrero (1995).

[2] ITU-T X.903 / ISO/IEC 10746-3, "Information technology - Open distributed processing - Reference model: Architecture" (1995).

[3] Object Management Group, "The Common Object Request Broker: Architecture and Specification". Revisión 2.0. Julio 1995.

[4] W. J. Barr, T. Boyd y Y. Inoue, "The TINA Initiative". *IEEE Communications Magazine*, **31**, 3, 71-76 (1993).

[5] L. A. de la Fuente, M. Kawanishi, M. Wakano, T. Walles y C. Aurrecoechea, "Application of the TINA-C Management Architecture". En A. S. Sethi, Y. Raynaud y F. Faure-Vincent (editores), "Integrated Network Management IV". *Proceedings of the 4th International Symposium on Integrated Network Management.* Chapman & Hall (1995).

[6] Jong-Hwa Yi, "Contribución al modelado de funciones de comunicación para aplicaciones de procesamiento distribuido abierto". *Tesis Doctoral.* Dpto. de Ingeniería de Sistemas Telemáticos, Universidad Politécnica de Madrid (1996).

[7] Draft International Standard ITU-T X.901 / ISO/IEC 10746-1, "Information technology - Open distributed processing - Reference model: Overview" (1995).

[8] Kerry Raymond, "Reference Model of Open Distributed Processing (RM-ODP): Introduction". Tutorial presentado en *International Conference on Open Distributed Processing (ICODP '95)* (1995). Disponible electrónicamente en *http://www. dstc.edu.au/AU/research_news/ /odp/ref_model/papers/icodp95.ps.gz*

[9] Proyecto ABS, "Broker Business Model". Documento Público D23, (1996).

[10] Draft International Standard ISO/IEC 13234-1, "Information technology - Open Distributed Processing - ODP trading function - Part 1: Specification", (1995).

[11] TINA Consortium, "TINA Business Model and Reference Points". Versión 4.0. Febrero (1997).

[12] TINA Consortium, "Service Architecture". Versión 4.0. Diciembre (1996).

[13] Object Management Grouop, "Electronic Commerce DTF Reference Model". OMG document *ec/96-09-02*. Septiembre (1996).

Experiencias en el uso de Redes Intranet en colectivos escolares con aplicaciones Multimedia

DVIKASNET

IÑAKI MOKOROA SEGUES
GERENTE UNIDAD DE TELECOMUNICACIONES
IBERMÁTICA S.A.
P. MIKELETEGI Nº 5 20009 DONOSTIA
Correo electrónico:imokoroa@ibermatica.es

Abstract:

Telematic and Multimedia Systems open a new technological horizon for school educational systems. DVIKASNET, an experience carried out by a consortium of companies located in the Basque Country, in which the objetive was to run a private net for both enjoyment and education, directed at children between 12 and 16 years of age, has produced, due to the research made by its promotors at the implementation stage and the following statistical study about the uses and tendencies of the target user, a set of basic principles about the advantages and disadvantages of this type of system, and the consecuences for a general implementation.

1. Introducción

Las tecnologías Multimedia se han revelado como una de las herramientas más importante en el desarrollo de los nuevos sistemas de formación, tanto en los aspectos de autoformación, como en los aspectos de apoyo a la formación tradicional.

La posibilidad de ofrecer al estudiante textos unidos a imágenes, sonidos y videos, nos permiten crear cursos mucho más didácticos, más fáciles de comprender y que mezclen la diversión con la formación, provocando una mejor actitud del alumno en cuanto al estudio, sobre todo en la juventud.

Por otra parte, la involucración de la informática en este tipo de desarrollos, permite crear cursos que se puedan ir adaptando a las características del alumno, guiándole a través de los contenidos a aprender, repitiendo aquellos puntos que se detecten más conflictivos y pasando por encima de aquellos que se conozcan.

En los últimos años, los avances conseguidos en las tecnologías de Telecomunicación, así como la implantación de nuevas infraestructuras más potentes, han abierto nuevas e importantes posibilidades, tanto en los sistemas tradicionales de formación, como en los nuevos sistemas Multimedia.

Los actuales sistemas basados en Internet acercan la formación a los hogares de los propios estudiantes independizándoles de un sistema rígido de horarios. Por otra parte, estos nuevos sistemas ofrecen la posibilidad de intercomunicarse con los profesores y el resto de alumnos de forma fácil y rápida, reduciendo notablemente los tiempos de consulta/respuesta tradicionales y mejorando la comunicación global.

Aunque todos estos sistemas antes mencionados teóricamente son perfectos para el desarrollo de la formación, en la práctica presentan una serie de dudas, aceptación tanto por el alumnado como por el profesorado, equipamiento de los hogares, facilidad para perderse entre los contenidos, preparación técnica necesaria, etc., de los que sólo podemos obtener conclusiones realistas a través de experiencias realizadas en condiciones lo más cercanas a la realidad.

Vamos a comentar en esta ponencia los diversos aspectos que han tenido lugar en el desarrollo de una experiencia promovida por tres empresas del País Vasco, Ardatz, El Diario Vasco e Ibermática, basada en una Red Lúdico-Educativa privada, en la que han participado ikastolas de la provincia de Gipuzkoa y del País Vasco-Frances.

2. Objetivos.

Sin duda alguna, la aplicación de las nuevas tecnologías en la sociedad va a cambiar radicalmente las formas de vida en los próximos años. La nueva televisión digital, la Multimedia y la Red Internet van a penetrar en los hogares produciendo unos cambios socioeconómicos de difícil calibración actualmente.

Este profundo cambio, va a exigir a las personas un trabajo de adaptación en el uso de todas estas tecnologías que van a estar implicadas en las diferentes áreas de la vida. Sin duda alguna, las personas que no hayan conseguido adaptarse tendrán problemas para poder desenvolverse en este mundo.

Por otra parte, una de las mayores preocupaciones de los Gobiernos en los últimos años ha sido la implantación de infraestructuras suficientes para poder hacer frente a las necesidades que las nuevas tecnologías de comunicación van a exigir, considerando que las infraestructuras de comunicaciones van a estar íntimamente relacionadas con la capacidad competitiva de los países.

Por último, las empresas que van a proveer a la sociedad de estos nuevos servicios tienen un amplio abanico de posibilidades y un gran mercado, pero, de igual manera, tienen grandes riesgos debido al desconocimiento que existe de la aceptación de dichos servicios. Ello hace que las empresas tengan un interés creciente en el desarrollo de pruebas piloto que permitan identificar las líneas del negocio.

Teniendo en cuenta todo lo anteriormente expuesto, DVIKASNET intenta conseguir dos objetivos principalmente:

- Familiarizar a los educadores y a jóvenes de la Comunidad Autónoma Vasca en el uso y aplicación de las nuevas tecnologías Telemáticas y Multimedia.

- Obtener una serie de datos que permitan definir los planes para la incorporación de las Tecnologías, y diseñar los servicios con garantías de aceptación.

Para ello se ha desarrollado una experiencia de servicios Internet, con un grupo reducido de usuarios y un seguimiento continuado de los resultados.

Este servicio se ha complementado con un estudio estadístico en cuanto a las capacidades tecnológicas en los hogares con objeto de analizar los volúmenes de mercado y los plazos de posible implantación.

3. Situación de partida.

Los primeros pasos estaban direccionados a conocer la situación de partida en cuanto a infraestructuras en las ikastolas, nivel de preparación de los profesores, características de los alumnos y, por último, unos primeros análisis del mercado, potenciados con nuevos estudios realizados durante la vida del proyecto.

3.1 Infraestructura escolar.

Una primera revisión de la situación en cuanto a equipamiento nos permitió descubrir que había grandes diferencias entre los diferentes centros.

En primer lugar, ningún centro escolar tenía acceso a Internet y en prácticamente todos los casos era necesaria la instalación de una nueva línea telefónica para poder ofrecer el servicio.

En cuanto a ordenadores las diferencias entre unos centros y otros eran muy importantes. Algunos centros tenía un número importante de Pcs de última generación y con capacidades Multimedia, casi todos tenía Pcs en red o independientes, pero en muchos casos el nivel del Pc no era el suficiente para los servicios que se pretendían ofrecer.

3.2 Profesorado.

Sin duda alguna, el éxito del proyecto estaba totalmente involucrado con el interés y capacidad de los profesores para poder dirigir y potenciar el uso del sistema en cada uno de los colegios.

Sin un convencimiento profundo de los profesores en la utilidad del sistema y un apoyo fuerte para que no se encontrasen problemas en el desarrollo de las clases el proyecto no podía ser llevado a cabo.

La situación se presentaba bastante difícil, ya que el nivel de conocimientos en muchos casos era muy bajo, generalmente aprendido por autoformación, prácticamente nulo en cuanto aplicaciones de comunicaciones y acceso a Internet y básicamente centrado en unas cuantas aplicaciones de uso generalizado (tratamiento de textos, hoja de cálculo, dibujo y bases de datos).

3.3 Estudiantes.

Aunque el nivel de número de Pcs en los hogares ha ido creciendo durante los últimos años de manera mas que proporcional, y en este aspecto nos podemos considerar privilegiados en el País Vasco con respecto a otras comunidades, el uso real en las edades en las que se posicionaba el proyecto era muy escaso.

Por otra parte, unos cuantos chavales, que tenían ordenador en casa y que lo dominaban perfectamente, presentaban un claro problema de diferenciación con el resto de compañeros e incluso podían suponer un problema para el profesorado, llegando en ocasiones a dominar los sistemas mejor que los propios profesores y dejándoles en una situación difícil ante el resto de la clase.

Por otra parte, los estudios realizados sobre como atraer la atención de los jóvenes a los

servicios, indicaban claramente que la Multimedia, la interactividad y la competitividad eran elementos fundamentales.

3.4 El Mercado Internet.

El análisis del estudio realizado por la empresa DBK.S.A. sobre el mercado Español en Internet, compaginados con datos extraídos de los Web de AIMC nos han permitido comprobar que los niveles de posibles usuarios domésticos es en la actualidad muy bajo y, aunque está experimentando un crecimiento rápido, tardará varios años en llegar a una masa crítica suficiente en la Comunidad Autónoma Vasca.

Se debe tener en cuenta que el número de estudiantes entre 12 y 16 años va a decrecer fuertemente en los próximos años, y que la cantidad de nuevas ofertas que se van a producir en los próximos años en las áreas de educación y ocio Telemático y Multimedia van a provocar confusión en los demandantes de los servicios.

4. Primeras decisiones.

Todo lo anteriormente expuesto nos llevo a definir una serie de puntos a cumplir para asegurar la consecución de los objetivos del proyecto.

- Definir un periodo de prueba suficientemente amplio para asegurar la obtención de datos reales sobre la experiencia que permitiesen tomar una decisión de ampliación o de suspensión. El periodo planteado ha sido de Enero a Septiembre de 1997.

- Trabajar durante la prueba piloto con un número de Centros Escolares reducido, que permitiese el retorno de la experiencia con un control total. Se seleccionaron 26 Centros.

- Realizar la experiencia en Centros distribuidos geográficamente de forma que se pudiese analizar diferentes realidades. La selección se realizó entre Centros de las diferentes comarcas de Guipúzcoa, de enseñanza pública y privada y se incluyeron dos Centros Escolares del País Vasco Francés.

- Conseguir equipamiento para estos 26 Centros tanto en la parte de Pcs Multimedia como en la de líneas. Se consiguió la colaboración de IBM y Telefónica.

- Evitar problemas de funcionamiento del sistema. En cada uno de los Centros se realizó la instalación directa y las pruebas iniciales que asegurasen el perfecto funcionamiento.

- Apoyar a los profesores en la utilización. Se creó un equipo de soporte telefónico en horario de oficina.

- Realizar un seguimiento continuado del sistema. Se desarrollaron un conjunto de estadísticas de uso y se realizaban consultas periódicas a los Centros para comprobar los problemas, detectar las ventajas y recoger información sobre nuevas demandas.

- Mantener el interés continuado en la utilización del sistema. Se realizo a través de concursos en el propio sistema.

5. Descripción del proyecto.

5.1 Arquitectura técnica.

Los puestos de usuario seleccionados son Pcs Multimedia (tarjeta sonido y Cdrom) con procesador Pentium, 16 Mb de memoria y módem corriendo bajo sistema operativo Windows 95, y utilizando explorer 3.0 como navegador, al que se le incorporaban un conjunto de plug-in, necesarios para algunos de los servicios ofrecidos.

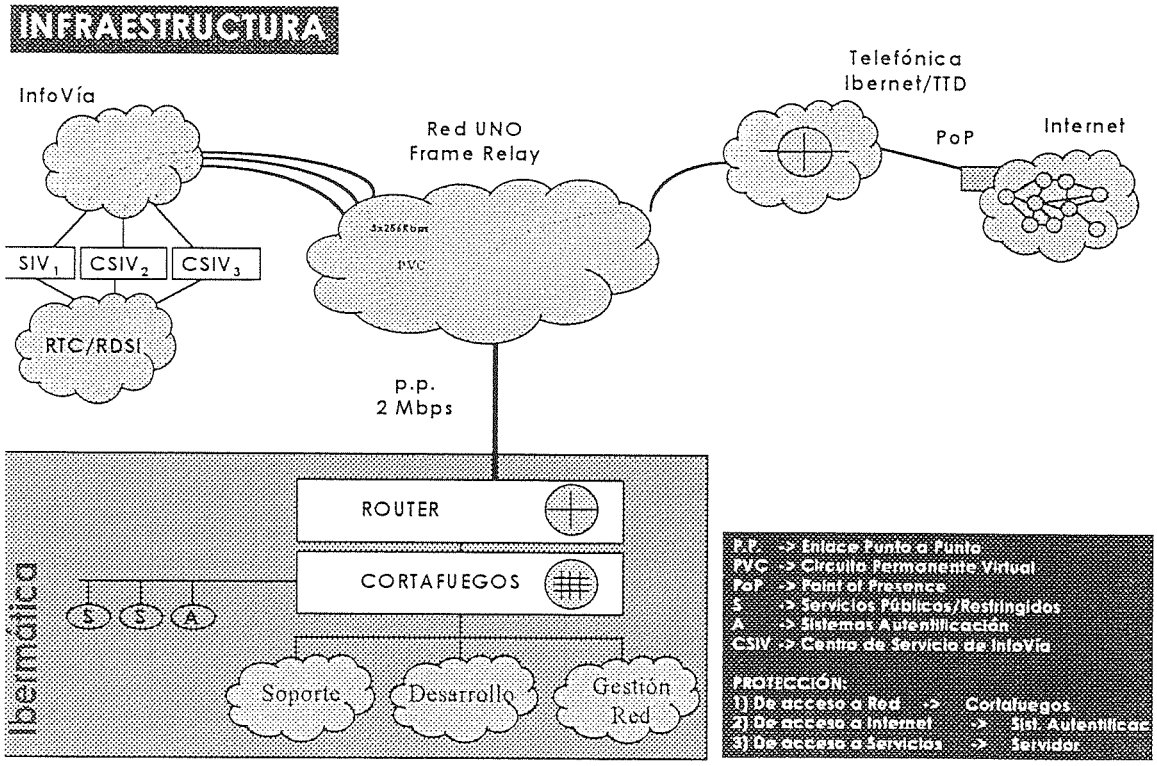
La conexión con el Centro de Servicio se realiza a través de Infovía, utilizando en algunos casos la Red Telefónica conmutada y en otros la Red Digital de Servicios Integrados. En este último caso se incluía en el Pc la placa RDSI.

El centro Servidor basado en una servidor Sun, con software Netscape, está conectado al exterior con una línea de 2Mb con caudales mínimos de 256Kb para Infovía y 128Kb en cada dirección para Internet.

Así mismo, se tiene instalado un equipo dedicado como cortafuegos, encargado de controlar los accesos de los diferentes usuarios.

Con objeto de dar mayor vistosidad al proyecto y ofrecer un campo más atractivo a los estudiantes, se implantó un Chat totalmente gráfico (PALACE) que permitía a los contertulios navegar por distintas aplicaciones, crear habitaciones privadas, jugar con otros participantes a juegos de mesa, disfrazarse y recoger diferentes objetos.

Se definieron dentro del proyecto tres tipos de usuarios diferentes dependiendo de las



posibilidades de cada uno de ellos.

- Profesores. Acceso libre a Internet y a los diferentes servicios del sistema.

- Alumnos. Acceso libre a todos los servicios del sistema.

- Visitantes. Acceso restringido a algunos de los servicios del sistema con objeto de que no causaran interferencias en votaciones y concursos.

Todo el sistema esta basado en Web, con gestión a través de una Base de Datos de los contenidos, actualización automatizada partiendo de ficheros de texto, iconos en movimiento y desarrollos en Java.

5.2 Contenidos.

Uno de los elementos más importantes para conseguir el éxito del proyecto esta relacionado íntimamente con los contenidos que tenían que cubrir los siguientes objetivos.

- Que el profesorado viese en ellos auténticas ventajas para su trabajo diario.

- Que incitasen a los alumnos a realizar nuevas conexiones, investigar por otros servicios e interactuar con el sistema.

- Bilingüismo, potenciando la participación tanto de los hispano parlantes como de los vasco parlantes.

Teniendo en cuenta los objetivos antes marcados se crearon un conjunto de secciones que describimos a continuación.

- **Ikasnet**

Es la página principal. Presenta un conjunto de iconos temáticos con los que se puede acceder a todo el resto de contenidos con una única pulsación. Se han incluido cajas de texto continuado, dibujos en movimiento y música.

- **Agenda**

El objetivo de esta sección es poner a disposición de los centros un lugar donde puedan dar información sobre actividades que realizan y de direcciones útiles. Está dividida en dos subsecciones, actividades y direcciones.

En el nivel de actividades se ofrece una lista de eventos que van a tener lugar en las

próximas fechas y que pueden ser de interés para los chavales. Esta actividades son aportadas por los

propios centros y se borran automáticamente cuando expira la fecha de realización.

En el nivel de direcciones se ofrece una lista de diferentes organizaciones culturales, recreativas y de ayuda clasificadas por tipos y que puede ser accedida desde un menú de la parte superior.

- **Ikastop 10**

El objetivo de este apartado es ofrecer un lugar interactivo, que permita participar a los alumnos y descubrir las preferencias de los demás. Esta repartido en cuatro secciones, música, cine, libros y vídeo.

Las sección de música es la más completa y la que más juego da. A parte de una lista de los títulos más vendidos, que cambia mes a mes, se permite escuchar treinta segundos de cada una de las canciones.

La participación de los alumnos se consigue a través de un sistema de votación que permite al alumno tras identificarse dar su voto a su canción preferida.

Inmediatamente la lista es actualizada y presentada de nuevo, de forma que el alumno puede comprobar al instante las repercusiones de su voto.

En caso de no encontrarse la canción preferido se puede enviar una petición y voto por correo electrónico.

El resto de secciones presentan únicamente una lista de los más vendidos que va cambiando mes a mes.

- **Kuxki**

Esta sección constituye la parte humorística y está basada en la implantación de las cuatrocientas tiras humorísticas de Kuxki publicadas en el Dv Gaztea.

Se permite la navegación por la diferentes viñetas a través de un conjunto de botones. Uno para selección por temas de la historieta, otro para la selección por fecha de publicación y un tercero para moverse adelante y atrás en el grupo seleccionado.

- **Ikasposta**

El objetivo de esta sección es crear una zona de comunicación entre los diferentes usuarios. Esta dividida en tres secciones, conferencias, correo electrónico, debates.

La sección de conferencias pone en marcha un Chat gráfico en el que los alumnos pueden navegar por diferentes habitaciones, conversar con los otros usuarios e interactuar con diversos objetos.

La sección de correo electrónico permite el intercambio de mensajes entre los diferentes usuarios e incluso con el exterior.

La sección de noticias presenta un foro de discusión sobre cuatro temas concretos que son actualizados mensualmente.

- **Azken ordua**

Esta sección permite la presentación de noticias a nivel mundial relacionadas con tres temas, informática e Internet, viajes y aventura y naturaleza y ecología.

- **Kuxkigrama**

Otra de las zonas interactivas basada en la participación en diferentes juegos educacionales. Esta dividido en dos secciones.

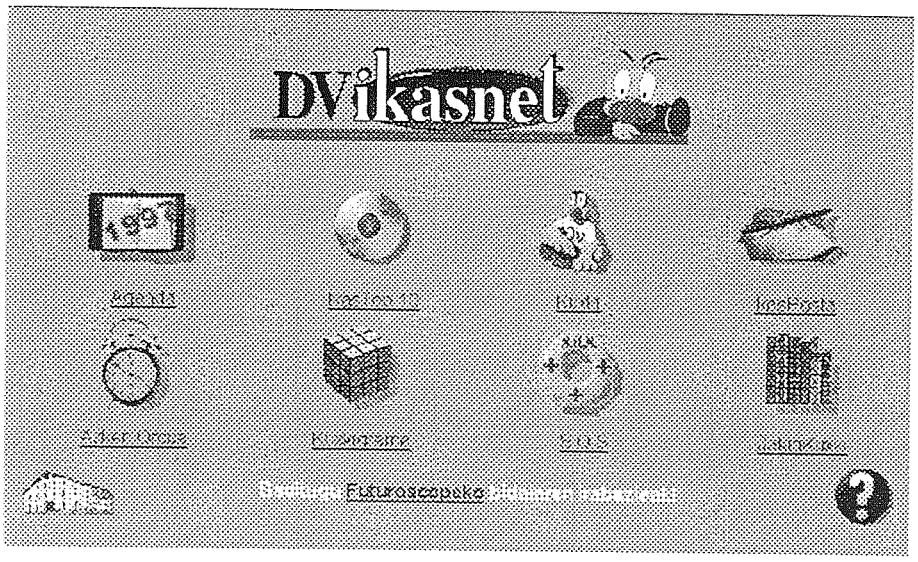
La primera presenta tres tipos de juegos, dos crucigramas, un juego de las diferencias, y un juego de palabras similar al ahorcado en el que la mascota Kuxki, recibe un tartazo o se come la tarta dependiendo de si acierta o no.

En la segunda se presenta un concurso de preguntas y respuestas, basado principalmente en el conocimiento de la cultura Vasca, desarrollado para la participación por equipos, que va cambiando mes a mes y en el que las puntuaciones se van acumulando.

Con el juego se persiguen dos objetivos. Por una parte, fomentar la colaboración entre los alumnos de una clase para la obtención de las respuestas. Por la otra provocar actividad de investigación y búsqueda de información en el conjunto.

- **S.O.S**

El objetivo de esta sección es ofrecer al usuario la información necesaria y suficiente para poder manejarse sin problemas con los diferentes servicios.



Contiene una lista de usuarios para ayudar a localizar las direcciones de correo de otros participantes, una guía de usuario que muestra información sobre los diferentes programas que dan servicio, y una ayuda en la que podemos encontrar la lista de problemas más frecuentes junto a su posible solución.

- **Jakinkuxki**

En esta sección se pretenden incluir todos aquellos temas relacionados con la cultura y la educación.

Inicialmente se ha pensado implantar un diccionario Euskara-Castellano que ayude a trabajar con el bilingüismo, y una enciclopedia. En un futuro, en esta sección se incluirán todo el conjunto de cursos Multimedia, así como la academia virtual.

5.3 Desarrollo

No hay duda que las características del proyecto nos obligaban a llevar una perfecta planificación con objeto de ir involucrando a profesores y alumnos de forma progresiva y evitar los lanzamientos y retiradas masivas.

En primer lugar era necesario conseguir que el profesor se sintiera dominador del sistema antes de enfrentarse con los alumnos. Para ello se realizo una formación especial a los profesores, dejando un plazo de un mes para que pudiesen investigar, probar y conocer todos los servicios.

La implantación del sistema se hizo progresivamente durante los meses de Enero y Febrero, manteniendo los concursos y los apartados de votaciones en pruebas hasta la puesta en marcha general.

Durante todo el periodo piloto se ha realizado un seguimiento continuado de la utilización del sistema por los usuarios, potenciando con cambios aquellos elementos de mayor interés y retirando los que tenían pocas consultas.

Por otra parte, y a través de consultas con los profesores se han puesto en marcha algunas secciones y servicios que en un principio fueron rechazadas.

6. Resultados obtenidos.

El proyecto piloto ha permitido obtener una serie de líneas básicas de actuación para la implantación de este tipo de sistemas. El siguiente

resumen pretende presentar lo que al entender del grupo promotor del proyecto ha sido de mayor relevancia.

- Analizada la utilización del sistema, el interés demostrado es ampliamente satisfactorio, ya que salvo en dos centros concretos y por causas ajenas al proyecto, la utilización ha sido máxima y continua.

- Las encuestas realizadas en los propios centros han demostrado que el interés entre los alumnos ha sido superior al interés entre los profesores en general, muy probablemente porque no consideraban los contenidos suficientemente educativos.

- No se han detectado problemas de utilización por parte de los chavales en ninguna de las secciones, demostrando muy buenas dotes para salir por si solos de los diferentes puntos conflictivos en los que se podían encontrar.

- No se han producido prácticamente problemas que no pudieran ser resueltos telefónicamente.

- Las áreas de mayor interés han sido las participativas, principalmente la música, el Chat y los concursos, quedando las áreas informativas con un nivel de uso muy bajo, demostrando que solo es posible mantener el interés con participación e interactividad.

- Otro de los puntos positivos del sistema ha sido el fuerte uso realizado en los sistemas de comunicación, correo, Chat y grupos de discusión, organizando los propios centros horas de conferencia compartidas entre varios grupos.

7. Conclusiones

La experiencia demuestra que las nuevas tecnologías pueden ser un elemento de apoyo para la mejora de la educación, prestando especial atención en la creación de materiales de calidad y utilizada como herramienta de apoyo de la enseñanza tradicional.

Para ello, exige un esfuerzo presupuestario importante en la dotación de equipamiento a los Centros Educativos, así como en la creación de materiales y servicios.

Es necesario realizar una preparación específica al profesorado en la utilización de estas nuevas tecnologías, en que materias son aplicables y como deben ser aplicadas.

La introducción debe ser paulatina, comenzando en la actualidad y provocando el uso continuado del sistema a través de sesiones periódicas, evitando uno de los riesgos más importante que es la adición.

Los estudios de mercado indican que la introducción de este tipo de sistemas en el hogar de forma generalizada va a tardar unos cuantos años más. Por ello se deberá tener especial cuidado para la no potenciación de diferencias producidas por las distintas capacidades de acceso a los nuevos sistemas desde el hogar.

Por último una duda. ¿ Somos conscientes de que las posibilidades comerciales de estos sistemas pueden cegarnos ante verdaderos peligros como la adición, la perdida de la comunicación directa, y la libertad dirigida?.

Referencias

[1] Barbara Means and Kerry Olson. SRI Consulting. Bussiness Intelligence Program. Technology,s Role in Education Reform (D96-1989).

[2] DBK, s.a. El mercado Español de Internet. Mayo de 1997.

[3] Comunity Networks. Lessons from Blacksburg, Virginia. Andrew Michael Cohill, Andrea Lee Kavanaugh. Artech House Publishers. (1997).

Entorno JAVA para el diseño y evaluación automática de cuestionarios

Ana Obregón Cuesta, Jesús Cid Sueiro [†]
DEPARTAMENTO DE TEORIA DE LA SE—AL Y COMUNICACIONES E INGENIERIA TELEMATICA [††]
ETSI DE TELECOMUNICACIÓN, UNIVERSIDAD DE VALLADOLID
C/ REAL DE BURGOS, S/N 47011 VALLADOLID
Correo electrónico: anaobr@casamaro.tel.uva.es, jesus@tel.uva.es

Abstract:

This paper presents a new Java environment to perform, evaluate and design tests in a computer network. The system provides an interface to the instructor that makes easy the test design; moreover, it generates, automatically, a web page with the test, that can be used by any authorised student. After a student completes the test, it can access to evaluation information. Also, the system uses the data collected from tha answers of different students to carry out an on-line statistical analysis and a neural learning based on self-organizing feature maps and the ART algorithm. The system provides information to the instructor about the knowledge state of a particular learning group, and it can be applied to student modelling in any Computer Aided Instruction System.

1. Introducción

La aparición de los ordenadores personales en el mercado ha provocado un gran avance en el desarrollo e implantación de los sistemas de instrucción asistida por computador (IAC). La forma en que los ordenadores pueden contribuir a mejorar la calidad de la enseñanza en todos los niveles educativos es objeto de investigación permanente por parte de psicólogos y pedagogos; son incontables los trabajos de investigación dedicados a averiguar cómo puede contribuir la informática a facilitar de la forma más eficiente posible el aprendizaje de un dominio de conocimiento, y cuál debe ser el papel del docente en un sistema de instrucción basado en ordenador ([1], [2], [3], [4] o [5] son sólo algunos ejemplos). El debate sobre las posibilidades educativas del ordenador no había sino comenzado cuando el concepto mismo de ordenador como herramienta de uso individual empieza a sustituirse por el de ordenador como sistema conectado a una red, local o dispersa geográficamente, que comparte información y recursos con otros ordenadores; el estudio sobre las posibilidades de las redes de ordenadores como herramientas que promuevan estrategias de aprendizaje cooperativo o competitivo, o que posibiliten una enseñanza eficaz a distancia centran la atención de numerosos especialistas.

Los primeros programas de ordenador educativos estaban diseñados para instruir determinadas habilidades prácticas; más tarde, los tutores informáticos mostraban al alumno una lección sobre cualquier materia; los sistemas de instrucción modernos son cada vez más flexibles, permitiendo adaptar la presentación de información a las características individuales del estudiante, y permiten al estudiante decidir su propia ruta de aprendizaje, o incluso aprender por descubrimiento [6]. Estos sistemas requieren un flujo de información bidireccional estudiante-ordenador y ordenador-estudiante. Para que el ordenador pueda adaptar la instrucción a los progresos del estudiante, es imprescindible disponer de un mecanismo que permita determinar el estado en el que se encuentra el proceso de aprendizaje. Para representar dicho estado, los programas de IAC utilizan sistemas específicos que recogen información acerca de lo que el estudiante conoce y desconoce, su "estilo cognitivo", grado de memorización, receptividad a consejos, etc. Estos esquemas formales se denominan Modelos del Estudiante.

Numerosos sistemas IAC extraen la información que necesitan para elaborar el modelo de un estudiante a partir de las respuestas a cuestionarios que se presentan al mismo durante diferentes etapas del proceso de aprendizaje. Recientemente, Harp [7] ha propuesto un modelo del estudiante basado en un tipo de redes neuronales denominado Mapas Autoorganizados [8]. A diferencia de la mayoria de los esquemas tradicionales, el propuesto por Harp es un modelo numérico que estima la situación del estudiante comparando sus respuestas a los cuestionarios con información estadística obtenida de respuestas generadas por un conjunto de estudiantes de prueba en diferentes estados de conocimiento. Por tanto, para parametrizar el modelo de Harp es necesario obtener respuestas de un conjunto de prueba.

El presente trabajo describe un entorno informático destinado a facilitar el diseño, realización y procesado de cuestionarios a través de una red informática. Se ha diseñado un interfaz que permite a un instructor humano diseñar un cuestionario tipo test escribiendo los enunciados de las preguntas, las posibles respuestas y un indicador de la respuesta correcta; el sistema genera automáticamente una página web a la que cualquier

estudiante provisto de una clave puede acceder. Las respuestas de todos los estudiantes son almacenadas procesadas dinámicamente por medio de herramientas estadísticas convencionales (cálculo de medias, varianzas, histogramas) y técnicas basadas en redes neuronales (mapas autoorganizados y modelos ART).

Aunque en principio el objetivo del sistema es que permita elaborar modelos de estudiante para sistemas IAC, puede ser utilizado por cualquier instructor humano para recabar información acerca de la marcha general del grupo de aprendizaje durante el tiempo de instrucción, o incluso por el estudiante para evaluar su grado de comprensión de un tema del programa.

En la sección siguiente se describen los elementos del entorno desarrollado; su funcionamiento se ilustra en la sección 3 con un ejemplo basado en una experiencia realizada en la ETSI de Telecomunicación; algunas conclusiones y las líneas futuras de trabajo finalizan esta contribución.

2. Estructura y funcionamiento del entorno JAVA

2.1. ¿Por qué Java?

Los creadores del lenguaje de programación "Java", lo definen como:

Un lenguaje de programación simple, orientado a objetos, distribuido, interpretado, robusto, seguro, de arquitectura neutral, portable, de altas prestaciones, multithread y dinámico."

De todas estas características, las que nos motivaron para escoger Java como lenguaje en el que programar el sistema son: que es distribuido, robusto, seguro y portable.

El hecho de que sea un lenguaje distribuido implica que puede actuar directamente con protocolos TCP/IP, como HTTP y FTP. Esto hace que las aplicaciones Java puedan abrir y acceder a objetos a través de la red vía URL con la misma facilidad con que un programador accede a su sistema local de ficheros. De esta forma, nuestro sistema de IAC podría incorporar fácilmente elementos ampliamente utilizados en Internet, como imágenes GIF, animación, Vídeo o Sonido, permitiendo al fin tener un sistema IAC dinámico y con grandes posibilidades de diseño.

Tanto la robustez como la seguridad son dos aspectos muy importantes a la hora de diseñar una herramienta que se va a utilizar en un entorno de red al que puede accceder un gran número de personas, o que incluso podría tener una apertura a Internet. Java es lo suficientemente robusto como para afirmar que en sus aplicaciones no existe la posibilidad de acceder a memoria y corromper los datos. De esta forma, se evitaría que cualquier usuario pudiera acceder a un cuestionario previamente diseñado y cambiara los enunciados de las preguntas, sus posibles respuestas o incluso la puntuación asignada a cada una de ellas. La seguridad inherente a Java evita también cualquier posible intrusión de virus.

Por ultimo, se necesitaba un lenguaje portable, de forma que el sistema sea independiente de la plataforma. Usuarios desde distintos entornos, como un PC o una estación de trabajo, e independientemente del browser utilizado, podrían utilizar por igual la herramienta. El sistema que se presenta se ha utilizado tanto en una estación de trabajo Sun Sparc 5 con sistema operativo Solaris, como en un PC Pentium con Windows'95, y usando como browser Netscape o bien Internet Explorer, y en todos los casos se obtuvieron practicamente los mismos resultados, excluyendo algunos detalles de diseño, como los colores. De esta forma, los sistemas IAC podrían ser ampliamente utilizables en redes locales en Universidades o centros privados, y, al ser independientes del entorno y de la plataforma, tendrían una dimensión comercial mucho mayor.

2.2 Elementos del sistema

Desde el punto de vista del usuario, la aplicación que se presenta en este artículo presenta dos funcionalidades bien diferenciadas.

En primer lugar, el usuario dispone de una interfaz Web para el diseño de tests o cuestionarios. Esta funcionalidad estará en principio reservada a las personas acreditadas para incorporar cuestionarios a la base de datos; típicamente el profesor.

Por otra parte, el usuario cuenta con un entorno de trabajo, también en interfaz Web, que le permitirá ejecutar y acceder al análisis de los datos relativos a cuestionarios previamente introducidos en la base de datos. El apartado de ejecución será el utilizado típicamente por alumnos que, de alguna forma, deben superar ese cuestionario. Por último, el apartado de Análisis podrá ser compartido tanto por profesores como por alumnos.

A continuación se presenta un diagrama de bloques que muestra los diferentes elementos de que consta el sistema; así como la relación entre ellos. (Fig1).

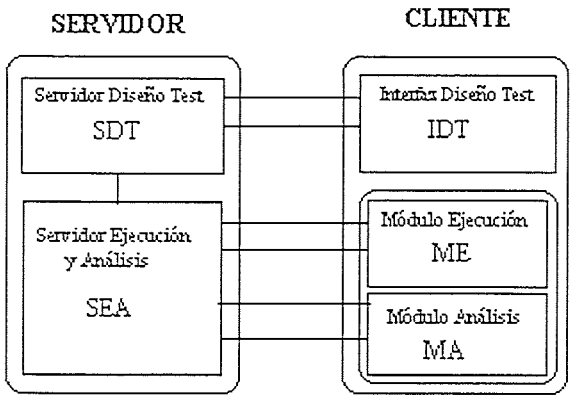


Fig 1: "Diagrama de Bloques "

En la Fig.1, se puede observar que hay dos grandes bloques: Cliente y Servidor. Previamente se ha explicado en que partes consiste el sistema desde el punto de vista del usuario, lo que se corresponde al bloque Cliente.

Debido a la arquitectura Cliente-Servidor utilizada, cada uno de los elementos del bloque Cliente tendrá asociado un elemento (programa) Servidor.

En el siguiente apartado se explicará el funcionamiento de la aplicación, la arquitectura Cliente-Servidor, y se detallarán cada uno de los elementos del diagrama de bloques, así como las relaciones entre ellos.

2.2 Funcionamiento.

Para tener una idea global del funcionamiento de la aplicación que se está presentando, lo primero que hay que tener en cuenta es que está basada en una arquitectura Cliente-Servidor.

Para realizar la interfaz Web que utiliza el usuario, se utilizan "applets" Java, programas escritos en Java que no se ejecutan directamente, sino desde un browser adecuado, y que van embebidas en páginas HTML. De esta forma, al cargar una página Web mediante la introducción de su URL, se puede ejecutar un applet Java embebido en ella. Los applets, por razones de seguridad, sólo se pueden comunicar a través de la red con el host del que provienen. Por lo tanto, para que el interfaz gráfico de usuario (IDC, ME y MA) pueda intercambiar información con la base de datos de los cuestionarios, que lógicamente residirá en el host en el que esté instalada la aplicación, es necesario que se comunique de alguna forma con un programa servidor que resida en dicha máquina, y que se encargue de actualizar los datos que el interfaz le proporciona, y enviarle la información que le pide.

En las aplicaciones cliente-servidor, el servidor proporciona algún servicio, como actualizar una base de datos o enviar algún tipo de información al cliente. El cliente, por su parte, utilizará estos servicios que le proporciona el servidor con algún fin, como recoger los datos del cliente, o mostrarle la información que había pedido. La comunicación que tiene lugar entre el cliente y el servidor debe ser robusta; no se deben perder datos, y estos deben llegar al cliente en el mismo orden en que fueron enviados por el servidor.

Java proporciona un canal de comunicación independiente del sistema que usa el protocolo TCP y que las aplicaciones cliente-servidor suelen usar para comunicarse en Internet. Este canal de comunicación está soportado por sockets, y es el que hemos utilizado para desarrollar nuestro sistema.

El servidor estará continuamente esperando peticiones de conexión por parte del cliente. Por lo tanto, para que el usuario pueda interaccionar con la interfaz gráfica desde cualquier máquina, es imprescindible que el programa servidor correspondiente se esté ejecutando en el host servidor. El servidor (o los servidores: SDT y SDA) de nuestra aplicación, tienen la ventaja de ser multicliente; es decir, que cuando un cliente solicita una petición al servidor, este lanza un proceso para atenderle, y sigue escuchando otras posibles peticiones por parte de nuevos clientes. Debe notarse la importancia que esto tiene para un sistema distribuido de IAC, que permite que múltiples usuarios puedan ejecutar el mismo o distintos cuestionarios simultáneamente desde distintas máquinas, o distintos profesores puedan estar a la vez diseñando sus respectivos tests.

Una vez explicada brevemente la arquitectura cliente-servidor, y los motivos por lo que se ha utilizado, pasamos a explicar el funcionamiento del sistema en sí.

Siguiendo con la terminología utilizada al hablar de la arquitectura cliente-servidor, la parte Cliente de nuestro sistema, con la que interactuará directamente el usuario, cuenta con una interfaz Web, lo que la hace muy intuitiva y amigable, y, por lo tanto fácil de utilizar. Esta compuesta por applets

Java integrados en paginas HTML a las que se podrá acceder desde cualquier máquina autorizada para utilizar la herramienta, a través de un Browser como Netscape o Internet Explorer sin mas que rellenar la dirección URL correspondiente.

Internamente, cada uno de estos applets "Cliente", se comunica a través de sockets con su respectivo programa servidor, también escrito en Java y que reside en la máquina en la que esta instalada la aplicación. De esta forma, cada vez que el usuario quiera almacenar los resultados de una ejecución del test, o bien consultar información a cerca de ese test, interaccionará directamente con la interfaz Web del cliente; pero este a su vez enviará los datos o hará peticiones al servidor a través de los sockets. El servidor será el que realmente actualize los datos o ejecute las operaciones pertinentes y, una vez obtenidos los resultados, se los enviara al Cliente, que los mostrará a través de su interfaz de forma que sean visibles para el usuario.

Así, la aplicación consta de dos grandes bloques con sus respectivos programas Cliente y Servidor: un interfaz para el diseño de cuestionarios (IDT) que trabaja con el Servidor de Diseño de tests (SDT), y otro interfaz para la ejecución y análisis de un cuestionario concreto (ME y MA) que se comunican con el Servidor de Ejecución y Análisis (SEA).

El IDT, cuyo aspecto ilustra la figura en la página siguiente, es el que utilizaría el encargado de diseñar o componer el test (tipicamente, el profesor); que, para elaborarlo deberá indicar su título, los enunciados de las preguntas, sus posibles respuestas y a cada una de ellas, deberá asignarle una puntuación. Para un entorno de IAC, el interfaz podría flexibilizarse en gran medida, de forma que permitiese al diseñador introducir el número de preguntas que él desease, con un número de respuestas independiente en cada caso, y un con un sistema de puntuación libre.

En este punto, cabe destacar que la herramienta está diseñada para trabajar con dos tipos bien diferenciados de cuestionarios. En primer lugar, los que llamaremos "test", para los cuales sólo una de las respuestas de cada pregunta se considera válida o correcta, y tendrá una puntuación de 1; el resto de las respuestas serán incorrectas y tendrán una puntuación de 0. Este es típicamente un formato de examen. Por otra parte, están los que llamaremos "encuestas"; para este prototipo, no hay una única respuesta válida por pregunta, sino que cada una puede tener una puntuación cualquiera dentro de un margen de valores.

Ambos prototipos de cuestionarios podrán ser implementados por la herramienta, que en algunos casos, como al realizar los histogramas, proporciona distintas utilidades según se esté ejecutando un tipo de cuestionario u otro. Los resultados de cada uno tendrán distintos significados, que tendrá que evaluar en cada caso el usuario, y más concretamente el diseñador de los cuestionarios o el propio sistema ICA.

Una vez que el diseñador haya rellenado correctamente toda la información acerca del cuestionario, deberá pulsar un botón indicando que ha finalizado el diseño. Así, los datos introducidos se almacenarán correctamente en una base de datos en la máquina servidor para que el Cliente pueda acceder a ellos cada vez que un usuario quiera ejecutar el test o encuesta. El encargado de actualizar de esta forma la base de datos será el programa servidor SDC, que a su vez, al recibir el nuevo cuestionario, creará para él una nueva zona donde almacenar sus datos, independiente de la del resto de los cuestionarios..

El otro gran bloque es el que proporcionara un cuestionario, previamente introducido mediante el IDT, para poder ejecutarlo o consultar información relacionada con él. A su vez, podemos dividir este bloque en dos módulos: el cuestionario en sí, listo para ser ejecutado ME, y un modulo de análisis que permitirá obtener información acerca del mismo, MA.

Al acceder a la pagina Web correspondiente a este bloque, aparecerá en pantalla una ventana con los títulos de los posibles cuestionarios a los que el usuario puede acceder, y que pueden haber sido introducidos por diferentes personas mediante el IDT como previamente se ha comentado. El usuario debe seleccionar uno de estos títulos e, inmediatamente, aparecerá una página con los dos módulos mencionados.

El primero de ellos, ME, muestra el cuestionario seleccionado en un formato estándar, en el que aparece cada pregunta con sus respectivas respuestas, de las cuales habrá que seleccionar siempre una y sólo una; y un botón que se pulsará cuando se haya finalizado la ejecución. Al pulsar este botón, aparecerá una ventana que dará la opción de cambiar alguna de las respuestas introducidas, o bien enviar los datos al servidor, lo que indicará que el cuestionario ha sido ejecutado. Una vez realizada la ejecución, los datos generados por el usuario irán a una base de datos residente en el servidor y exclusiva para cada cuestionario. Se entiende que a este módulo tendrán acceso los usuarios autorizados por el diseñador de los test (tipicamente, los alumnos).

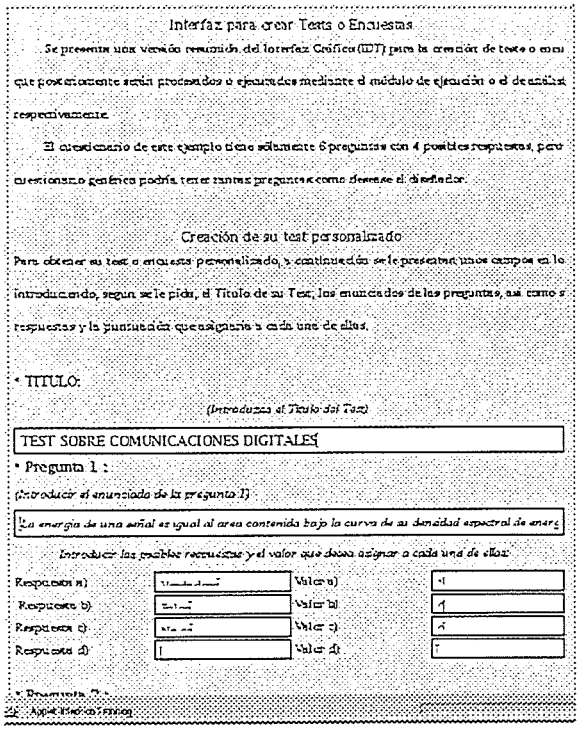


Fig. 2 : Interfaz de diseño del cuestionario.

El otro módulo es el de análisis, MA. A través de él, el usuario puede hacer peticiones sobre estadísticos de las ejecuciones del cuestionario recibidas hasta el momento.

El usuario seleccionará sobre el IDC la opción relativa a la información que desea recibir, a su vez el MA hará las correspondientes peticiones al SEA, que siempre estará esperando nuevas peticiones por parte del usuario. Al recibir una de estas peticiones, el SEA recogerá los datos oportunos de la base de datos, ejecutará la función correspondiente a la petición recibida, y una vez obtenida la información, se la enviará a través de los sockets de nuevo al cliente MA, que se le mostrará los resultados al usuario a través de la interfaz.

La información que un usuario puede obtener es la siguiente:

- Media de cada una de las preguntas. Es decir, de las puntuaciones obtenidas por los usuarios para una pregunta en concreto. Esto proporcionará al profesor, en principio, el grado de dificultad que la pregunta tuvo para los alumnos si se trata de un test; o bien, la respuesta preferida por los encuestados si el cuestionario es de tipo encuesta.

- Media de la ultima ejecución. Para un test, sería una medida de la nota que el alumno obtendría por su ejecución. En un sistema IAC, si el alumno pudiera acceder a esta información, tendría una forma rápida y automática de autoevaluación.

- Varianza de cada una de las preguntas. Proporciona al profesor el grado de conocimiento de los alumnos; si la varianza de una pregunta es pequeña, los alumnos tienen un conocimiento muy similar en esa materia, lo que podría significar que la pregunta les resultó fácil, en cuyo caso la media sería alta; o bien muy difícil, aparejado a una media muy baja. En el caso de las encuestas, serviría para saber si las personas encuestadas tienen parecidas o distintas preferencias en cuanto a esa materia.

- Histogramas. Para los llamados test, es interesante obtener los histogramas de los encuestados, es decir, la representación gráfica de cuantos encuestados tienen 0, 1, etc .. respuestas correctas, así el profesor tendría una representación gráfica de la proporción de sus alumnos que superan o no el test, e incluso de cuantos alumnos tienen una puntuación concreta. En cambio, para los cuestionarios de tipo encuesta, seria mas interesante obtener los histogramas por preguntas, es decir, para cada pregunta, un histograma que indique cuántos usuarios contestaron 'a', 'b', etc. Así se pueden medir las preferencias de los encuestados por preguntas o materias.

- Algoritmos de redes neuronales: algoritmo SOFM y algoritmo ART, con posibilidades de representación gráfica.

Este módulo puede tener un acceso restringido por medio de la introducción de alguna palabra de paso, o mediante un sistema de login y passwords de usuarios. En principio debe ser interesante para el diseñador del cuestionario, aunque en algunos casos también puede ser conveniente que el ejecutor tenga también acceso a este tipo de información.

Cada nueva ejecución del cuestionario incrementará la base de datos asociada a el, de forma que cuantas mas ejecuciones del cuestionario se produzcan, se podrán obtener unos estadísticos mas fiables, y los algoritmos de redes neuronales proporcionarán, asimismo, una información mas exacta.

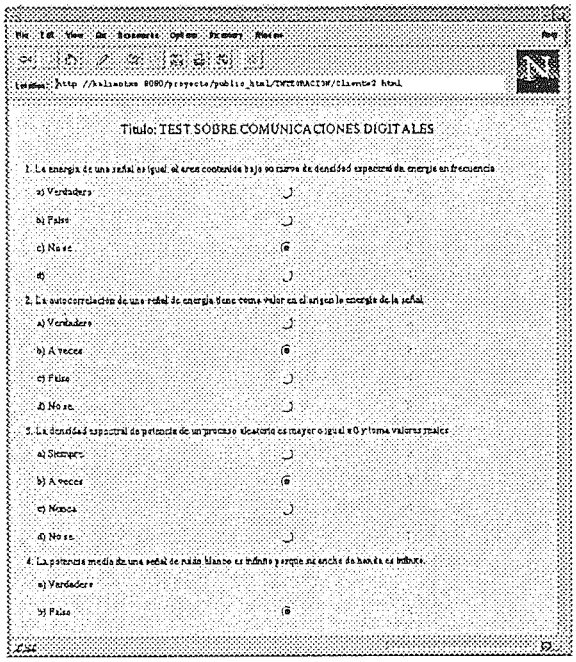


Fig. 3 : Ejemplo de cuestionario sobre comunicaciones digitales.

Al ser el servidor multi-cliente, varios usuarios pueden ejecutar simultáneamente el mismo cuestionario desde diferentes maquinas; lo que hace posible obtener un gran numero de ejecuciones que, como hemos visto, hará que los resultados proporcionados por la herramienta sean mucho mas fiables y útiles.

La información combinada de todos estos datos proporciona al diseñador del cuestionario una buena estadística de la marcha de las ejecuciones de su cuestionario. Además, la herramienta está totalmente abierta a ampliaciones especialmente orientadas a la flexibilización, lo que la hace especialmente adecuada para desarrollar sobre ella aplicaciones de sistemas ICA.

3. Modelado de Estudiante.

3.1. Cómo caracterizar al estudiante.

Se modela al estudiante para identificar el estado de aprendizaje en el que se encuentra, cuáles son sus lagunas o qué conceptos comprende y domina; ello permitira sugerir vías de instrucción, recomendar repaso de algunos temas o modificar el nivel de profundidad con que el instructor (humano o automático) expone cada tema.

Para comprender la utilidad del sistema desarrollado como herramienta de modelado del estudiante en sistemas AIC, es necesario describir brevemente el modelo neuronal de Harp.

Para modelar a un estudiante se procede del modo siguiente: en primer lugar, se plantean al estudiante cuestiones y problemas durante el mismo proceso de instrucción. La respuesta a cada cuestión se evalúa, recibiendo una puntuación de 0 (incorrecta) a 1 (correcta). Desde el punto de vista del modelo de Harp, un estudiante es un vector cuyas componentes son las puntuaciones obtenidas a cada una de las cuestiones planteadas.

Nótese que el valor medio de las componentes proporciona una calificación global al estudiante (es su "nota media"), pero no informa acerca de qué aspectos se dominan y cuáles no. El valor de cada componente es esencial en la identificación del estado de aprendizaje.

En el modelo de Harp, se dispone de una colección de estudiantes "típicos", cada uno de ellos caracterizado por su vector de puntuaciones, que llamaremos "patrón". El estudiante se asocia al patrón más parecido, y el sistema AIC responderá al estudiante del mismo modo que lo haría con su patrón asociado.

3.2. Redes Neuronales.

Los estudiantes "patrón" deben determinarse en algún momento anterior a la realización del cuestionario por parte del estudiante; por ejemplo, durante el diseño del sistema AIC: un experto (humano) en el dominio de conocimiento puede proporcionar "a priori" las puntuaciones típicas que obtendrían estudiantes en diferentes períodos de instrucción. El procedimiento supone que el experto conoce qué conceptos plantean más dificultades o qué distribuciones de puntuaciones son más frecuentes. Sin embargo, no siempre el experto tiene experiencia como evaluador, ni la determinación a priori de los patrones de respuesta típicos resulta fácil. Para resolver este problema, Harp obtiene los patrones empíricamente a partir de las puntuaciones obtenidas por una colección grande de estudiantes. Utiliza, para ello, un algoritmo que, a partir de una colección de vectores de puntuaciones, determine un familia de patrones que represente a todos los posibles estudiantes.

El problema de la determinación de los patrones, así planteado, es equivalente al de cuantificación vectorial, muy conocido en el ámbito del procesado de señal. Existen muchos algoritmos de cuantificación vectorial, cada uno con sus peculiaridades, y no nos detendremos aquí en su descripción; el lector interesado puede acudir a multitud de referencias, como [8]; diremos solamente que el programa informático que se describe en este artículo se implementaron dos técnicas diferentes:

1. El método de los mapas auto-organizados (SOFM, Self-Organized Feature Maps)

propuesto por Kohonen. Es el que utiliza Harp, porque no solamente obtiene los patrones de respuesta sino que los organiza en una estructura ordenada; se pretende con ello organizar los patrones en secuencias de aprendizaje, de modo que cada secuencia de patrones desde el (0,0,...,0) (el estudiante que no sabe nada) hasta el (1,1,...,1) (el estudiante que lo sabe todo) forma una "trayectoria" de aprendizaje típica; con ello, se puede averiguar qué conceptos suelen adquirirse con mayor facilidad, y cuáles menos.

2. El algoritmo ART (Adaptive Resonance Theory); a diferencia del algoritmo de Kohonen, el ART no fija previamente el número de patrones, sino que se establecen de forma automática, adaptándose a la dispersión de los vectores de puntuaciones.

La eficiencia de la cuantificación depende del volumen de datos que se disponga; por tanto, antes de poner en marcha el sistema es importante que un buen número de estudiantes hayan realizado el cuestionario. Si este no es posible, el proceso de cuantificación puede realizarse de forma adaptativa: inicialmente, los patrones son determinados por un experto humano; a partir de ese momento, cada vez que un estudiante-usuario rellena el cuestionario se actualizan los patrones, de modo que el sistema AIC puede adquirir, con el tiempo, experiencia sobre los tipos de estudiantes más frecuentes.

Si el sistema está implementado en una red abierta, tipo Internet, y la forma en la que el usuario accede al sistema no puede controlarse, el entrenamiento de la red plantea algunas cuestiones de seguridad y confianza, dado que no hay garantías de que el estudiante haya realizado el test sin ayuda o siquiera con interés. Un usuario malicioso podría generar vectores de puntuaciones al azar que confundirían al algoritmo SOFM o al ART, y falsearían el resto de datos estadísticos calculados por el sistema. Para evitarlo es necesario establecer algun sistema de claves o de restricciones de acceso.

Por otro lado, nótese que las posibles respuestas de un cuestionario tipo examen suelen ser del tipo "correcto" o "incorrecto", de modo que el vector de puntuaciones de cada estudiante solamente puede tener valores 0 y 1. En ocasiones, es conveniente agrupar las cuestiones por los conceptos que manejan, de modo que cada componente de un patron no mida una puntuación en una cuestión individual, sino la puntuación media de un grupo de cuestiones asociadas a uno o varios conceptos importantes de la asignatura. Esta facilidad, que permitiría reducir el tamaño de los vectores y facilitaría el entrenamiento de los mapas autoorganizados, no ha sido implementada todavía, pero se incorporará en versiones futuras.

4. Conclusiones

En este artículo se presenta un entorno informático basado en Java para el diseño, realización y evaluación automática de cuestionarios, que puede ser utilizado por un estudiante para su autoevaluación, por un instructor humano para obtener información acerca de la marcha global de un grupo de aprendizaje, o por un sistema AIC para determinar el estado de conocimiento de un usuario-estudiante. El sistema está concebido, en cualquier caso, para su utilización en una red informática, ya sea de ámbito local (un laboratorio, una escuela o una facultad, por ejemplo) o en un sistema abierto tipo Internet.

El sistema crecerá en el futuro incorporando un mayor número de herramientas de tipo estadístico (cálculo de percentiles, estadísticos de orden superior, otros cuantificadores vectoriales...) y flexibilizando los requerimientos sobre las encuestas (incorporando, por ejemplo, la posibilidad de agrupar las cuestiones por conceptos básicos, permitiendo respuestas múltiples en cada pregunta...). Su aplicación en un entorno educativo real es, quizás, la tarea pendiente más importante.

Referencias

[1] Bracey, G. "Computers in Education: What the research shows". *Electronic Learning*, Nov-Dec, (1982).

[2] Ragosta, M. "Computer assisted instruction and compensatory education: A longitudinal analysis". *Machine Mediated Learning*, **1**, 1 (1983).

[3] Lepper, M.R., "Microcomputers in education: Motivational and social issues". *American Psycologist*, **40**, 1 (1985).

[4] Cooper, L.R., "CAI with home-bound students proves successfull in model program". *THE Journal*, **18**, 68-69, (1991).

[5] Kulik, J.A. , Bangert, R.L., y Williams, G.W. "Effects of computer-based teaching on secondary school students", *Journal of Educational Psychology*, 75, 19-26 (1983).

[6] Wenger, E., *Nartifitial Ingelligence and Tutoring Systems*. Morgan Kauffmann Publishers, inc, (1987).

[7] Harp, S.A. , Samad, T., Villiano, M. "Modelling Student Knowledge with Self-Organized Feature Maps", *JIEEE Trans. on Systems, Man and Cybernetics*, 25, 5 (1995).

[8] Hayking, S., *Neural Networks: a comprehensive foundation*. Nueva Jersey: Prentice-Hall (1994).

Arquitectura Avanzada de Servidores WWW Autogestionados

DEPARTAMENTO DE ELECTRÓNICA Y SISTEMAS
FACULTAD DE INFORMÁTICA, UNIVERSIDAD DE A CORUÑA
CAMPUS DE ELVIÑA s/n, 15.071 A CORUÑA
Antonio López Fernández
Rocío Paradela Guerrero
Justo Hidalgo Sanz
Fidel Cacheda Seijo
Alberto Pan Bermúdez
Angel Viña Castiñeiras
Correo electrónico: {alf, rocio, justo}@cesat.es, {fidel, alberto, avc}@gris.des.fi.udc.es

Abstract:

Demand for Web-based aplications has undergone a spectacular increase. As a result, it's time to reconsider if present Web server models are capable of meeting the needs of those applications. This paper begins by analizing which are the demands of a dynamic web server with automatic data input and output , automatic administation systems, access statistics, and server data security.

Later, an architectural model as the one described above will be introduced. We will review its functional module breakdown (access, data, management, administration, and presentation modules) and how these modules interact.

Finally, there is a description of this architecture's implementation in an actual high-traffic web server. New technologies, such as Java -which allows the creation of cross-platform web servers, and the capability to run under UNIX or Windows NT- have been used in this implementation.

In order for the web-server to handle high-volume information efficiently, development of this application has been built over a relational database. The aforementioned architecture has been implemented in its entirety as part of an internet search engine named BIWE (http://biwe.cesat.es).

1. Introducción

Uno de los factores determinantes en la expansión de Internet, fue la aparición y difusión de los servidores WWW. En la actualidad los servicios WWW son los más usados en Internet. Esta popularidad ha obligado a muchas empresas y organismos a prestar sus servicios a través de estos medios. Este es el motivo por el cual el desarrollo de nuevas aplicaciones basadas en el Web ha experimentado un gran impulso. Un handicap importante a la hora de proporcionar sevicios a través de Internet ha sido la integración en este entorno de las aplicaciones ya existentes.

Uno de los principales problemas a la hora de ofrecer servicios basados en WWW ha sido el gran esfuerzo necesario para mantener actualizada la información publicada en los servidores Web. Un primer paso para reducir este esfuerzo es conseguir que las páginas HTML sean generadas total o parcialmente de forma dinámica.

La utilización de páginas HTML dinámicas tiene dos ventajas a la hora de administrar un servidor WWW. En primer lugar evita la reedición de páginas HTML cada vez que hay un cambio en la información que se publica en el servidor. Además desaparecen las páginas con información obsoleta.

Estas páginas HTML generadas dinámicamente obtendrían el contenido que tienen que publicar de un sistema lógico de almacenamiento de información (ficheros indexados, bases de datos relacionales, etc.). Esta información será presentada al usuario final en base a un formato preestablecido.

La primera técnica empleada para generar páginas HTML dinámicas fue la utilización de CGIs (*Common Gateway Interface*). La especificación CGI define el mecanismo mediante el cual se comunican un servidor HTTP y un programa. Utilizando CGIs se permite ejecutar programas en el servidor cuyo resultado se direcciona al cliente Web a través del servidor HTTP, efectuando únicamente una impresión por la salida estándar

Generalmente los CGIs suelen ser la acción o respuesta al envío de un formulario HTML al servidor WWW. La especificación CGI define como se deben pasar los campos del formulario HTML al CGI para que éste pueda procesarlos.

En servidores WWW con un volumen de información a tratar pequeño o mediano, el modelo más sencillo para la implementación de un servidor WWW dinámico, consiste en la utilización de CGIs. Estos, en función de los parámetros de entrada que reciben, acceden a un conjunto de ficheros de los que obtienen la información que han de presentar al cliente WWW. Esta información suele estar indexada mediante una serie de procedimientos periódicos para aumentar el rendimiento en el acceso a los datos.

Sin embargo, cuando los servidores WWW manejan grandes cantidades de datos, o información con un alto grado de interrelación, se

hace imprescindible almacenar ésta en un Sistema de Gestión de Base de Datos (SGBD), al no ser los sistemas de indexación tradicionales lo suficientemente óptimos. Por tanto los CGIs deben ser capaces de acceder al SGBD para poder publicar esta información en los servidores WWW. Además de las escasas prestaciones de rendimiento, los CGIs presentan serios problemas de seguridad. Todo esto hace inviable la construcción de una aplicación WWW que procese la información de un SGBD de forma eficiente y segura utilizando esta tecnología.

Para evitar las limitaciones de la tecnología CGI han surgido distintas alternativas:

- Extensión del servidor HTTP mediante APIs: esta alternativa consiste en el desarrollo de las aplicaciones Web extendiendo la funcionalidad del servidor HTTP[2]. Para esto se proporciona un API (Aplication Program Interface) con el cual el desarrollador puede programar la aplicación, generar una librería, y posteriormente enlazarla dinámicamente con el servidor HTTP. La desventaja de los servidores implementados con esta tecnología es que tanto las librerías desarrolladas como el servidor HTTP comparten el mismo espacio de direcciones. Esto significa que el fallo de una librería defectuosa puede llegar a bloquear el sistema entero.

- Gestor intermedio de ejecución: para evitar el problema de la ejecución en el mismo espacio de direcciones, algunos servidores HTTP utilizan un gestor intermedio de ejecución [1]. Este gestor se encarga de recibir las peticiones de ejecución, generadas por los navegadores Web, y asignárselas a las unidades de ejecución. Las unidades de ejecución, como su propio nombre indica, se encargan de ejecutar las librerías desarrolladas utilizando el API del servidor HTTP, en espacios de direcciones distintos.

La arquitectura de servidor WWW que se propone a continuación se basa en la segunda de las alternativas expuestas. Una vez definida la tecnología que se va a utilizar para el funcionamiento del servidor hay que desarrollar la arquitectura del mismo para luego pasar a su implementación.

Actualmente el Hardware y Software instalado en las empresas es muy diverso. A menudo han de convivir en la misma red corporativa, máquinas de distintos fabricantes y/o distinto Sistema Operativo. Por tanto, es importante que el sistema final desarrollado tenga un alto grado de portabilidad. Este ha sido un aspecto de especial importancia a la hora de elegir las soluciones tecnológicas sobre las que basa la arquitectura propuesta.

2. Descripción de la arquitectura

Se ha decidido utilizar la arquitectura modular compuesta por cuatro bloques básicos, que se muestra en la figura 1. A continuación se explica brevemente en qué consiste cada módulo:

- Módulo de presentación: interfaz Web entre el sistema y el usuario.
- Módulo de acceso a datos: engloba todas las operaciones que se realizan sobre los datos, es decir, inserción, borrado, actualización y consulta de datos.
- Módulo de gestión: información interna sobre datos diversos del Web (accesos, históricos, etc).
- Módulo de administración: conjunto de herramientas útiles para el administrador del servidor Web.

La divisisión en estos cuatro módulos se ha elegido por las siguientes razones:

- Se utiliza un planteamiento natural del problema.
- La modularidad no implica independencia absoluta. Los módulos interaccionan entre ellos y se puede reutilizar tanto el diseño como el código.
- Permite la codificación con la herramienta más apropiada para cada operación dentro del servidor Web, mejorando su eficiencia y fiabilidad.
- Cada módulo se puede dividir en submódulos de forma fácil y eficiente, llegando a la fase de implementación cómodamente.

2.1 Módulo de presentación

A partir de este módulo se crea la interfaz gráfica que actúa como intermediaria entre el usuario del sistema de información y el sistema de acceso a datos.

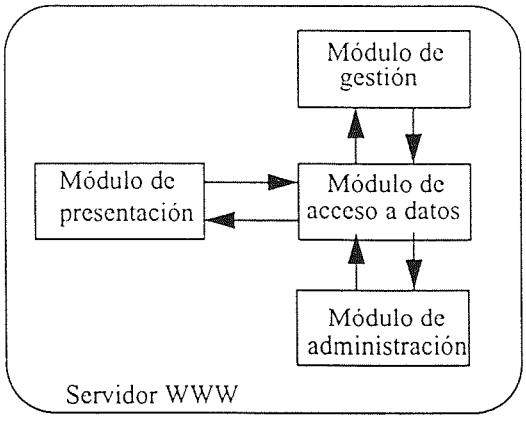


Figura 1: Diagrama de la Arquitectura

Tiene como característica principal su dinamismo. Las páginas HTML que van a formar parte de la interfaz no son estáticas sino que serán generadas y mostradas al usuario según las necesidades y peticiones del mismo.

En un servidor autogestionado ideal, en ningún momento se debería necesitar la intervención del ser humano. Obviamente, en la actualidad un servidor no es capaz por sí mismo de obtener, interpretar, decidir y efectuar cambios en el aspecto gráfico de la interfaz con las tecnologías existentes, sobre todo cuando se trata de hacer cambios que afectan al diseño gráfico del servidor como por ejemplo el cambio de la imagen corporativa de una empresa u organismo.

Teniendo en cuenta estas limitaciones, lo que sí sería deseable es que esta interfaz se basase en páginas Web generadas dinámicamente a partir de una estructura que podemos denominar *plantilla*. Para la organización de las plantillas se puede optar por una de estas dos opciones:

- La primera opción es tener una jerarquía orientada a objetos, de forma que la creación de la interfaz gráfica sea sencillamente la implementación de diferentes bloques tales que, ante el cambio de algún elemento, como por ejemplo un gráfico (el logotipo de la empresa), no sea necesaria la total reestructuración de las páginas y programas del servidor.

- Una segunda opción, como extensión de la primera, es almacenar las plantillas en la propia base de datos del servidor Web. Eligiendo esta segunda opción , se consiguen tres mejoras principales:

  - El cambio del aspecto gráfico del servidor Web, no involucrará en ningún caso la recodificación de los programas que lo componen. Para cambiar el aspecto gráfico de todo el servidor, bastaría con reemplazar estas plantillas en la base de datos, aplicándose inmediatamente los cambios realizados a todo el servidor.

  - Se consigue una consistencia absoluta en estos cambios ya que todas las páginas HTML que muestra el servidor se generan dinámicamente a partir de las mismas plantillas.

  - El proceso de cambio de estas plantillas puede ser a su vez almacenado en la base de datos, permitiendo la realización de diferentes estadísticas e históricos. Se puede pensar, por ejemplo , en la creación de estadísticas accesos a la publicidad existente en una página Web comercial, el tiempo medio de estancia de un producto o compañía, precios, etc.

2.2 Módulo de acceso a datos

Una vez creado el módulo de presentación, se requiere una interactividad con el usuario, darle el dinamismo necesario para que el cliente pueda comunicarse con el servidor Web. Se desea que los usuarios puedan consultar e incluso introducir información en el servidor WWW.

El módulo de acceso a datos se encarga de procesar las peticiones suministradas por el usuario del Web, obtener la información solicitada y proporcionarla al módulo de presentación para que éste último se la muestre al usuario

Este acceso a los datos ha de cumplir las siguientes propiedades:

- Ha de ser seguro, introduciendo, si es preciso, capacidad de identificación.
- La información ha de ser consistente en todo momento. Esta característica es esencial ya que esperamos y deseamos que tanto la introducción de datos como la consulta de los mismos pueda ser efectuada concurrentemente, sin producir los problemas típicos en estos casos, como pueden ser los bloqueos.
- La información ha de ser automáticamente comprobada tanto sintáctica como semánticamente.
- Tras la introducción de nueva información en el servidor, ésta ha de ser accesible inmediatamente, sin restringir el hecho de que más tarde pueda ser comprobada por un operario.

Con la información proporcionada por este módulo se creará una página HTML dinámica en la que el cliente podrá observar la información requerida, la respuesta a una consulta, la confirmación de que el sistema ha introducido sus datos correctamente, o incluso, mensajes de error, con posibles soluciones (información adicional a añadir, p.e.).

Tanto para la inserción de información en la base de datos como para el tratamiento de estos datos, se necesitan lenguajes de programación que se adecúen perfectamente a las necesidades de diseño impuestas en cada caso. Para el tratamiento de la información en una base de datos relacional, la referencia más clara es la utilización del lenguaje SQL con procedimientos almacenados.

Para el tratamiento externo de datos (tanto de

entrada como de salida) a este módulo, es necesario un lenguaje de alto nivel, multiplataforma... como es el caso de Java.

2.3 Módulo de gestión

El coste más importante de un servidor WWW convencional radica en el mantenimiento del mismo. La implementación de un servidor WWW normalmente tiene un coste relativamente bajo, tan sólo el desembolso inicial para la edición de las páginas HTML y el software necesario para publicarlas. El verdadero coste de tener un servidor WWW es el correspondiente a la persona o personas encargadas de mantener actualizada la información publicada en él.

Una de las labores más tediosas suele ser la gestión de la información de accesos al servidor. En sistemas WWW tradicionales, esta información suele ser almacenada por el software especializado del servidor Web en un fichero de texto plano, que ha de ser después procesado, ya sea a mano o mediante otra herramienta para generar unos informes; lo cierto es que sería más aconsejable prescindir de la rigidez y poca integrabilidad que estas herramientas suelen imponer. Para lograrlo, se ha desarrollado este módulo de gestión: el almacenamiento de la información de acceso se realiza en el SGBD, estableciendo una relación entre la información de gestión y el objeto gestionado. Esta información puede ser fácilmente utilizada por otros módulos y/o aplicaciones para generar informes ad-hoc sobre los accesos al servidor Web, ya sea mediante estadísticas, históricos, gráficas, etc. Además, esta capacidad puede ser integrada con otros servicios, como puede ser el correo electrónico: los informes serían entonces generados automáticamente por el servidor Web cada cierto tiempo previamente establecido, para después ser enviados automáticamente a la persona y/o compañía interesada.

Un ejemplo representativo es la publicidad insertada en una página Web, cuya vigencia va a depender del número de veces que un usuario externo accede a ella. Este estudio es necesario que se remita periodicamente, trabajo que el servidor puede realizar de una manera eficiente. Otro posible ejemplo es el control de accesos a determinada información total o parcialmente restringida.

2.4 Módulo de administración

El administrador ha de tener un conjunto de herramientas que le permitan tener un mayor control de la información almacenada en el SGBD que el que tiene un usuario normal. Por ejemplo, el administrador debe poder eliminar información, o modificarla, más allá de las restricciones que se puedan tener en condiciones normales, tales como que tan sólo el usuario poseedor de un registro determinado pueda eliminarlo.

Ésta es una parte muy importante en la actualización automática de datos; siempre ha de haber un control humano sobre aquellas inserciones que puedan afectar al entorno de una forma más o menos grave. El sistema automático se puede encontrar con muchos problemas para restringir información con diferencias semánticas entre sus diferentes campos (p.e. la dirección no existe en una ciudad determinada, o los comentarios respecto a un tema determinado no tienen nada que ver con ese tema en cuestión), que sólo un control manual puede evitar.

Aparte de la actualización anteriormente mencionada, el módulo de administración ha de permitir como mínimo las siguientes acciones:

- Borrado e inserción de datos restringidos al usuario normal.
- Modificación de la interfaz gráfica (nueva publicidad o logotipo, información temporal, etc).
- Conocimiento de información adicional de los datos que no es accesible al resto de usuarios.

Estas herramientas también han de permitir una actualización inmediata, de forma que los cambios se vean reflejados al instante en el servidor Web. De la misma forma, deben ser accesibles mediante Web, aunque de forma restringida, para que el administrador pueda realizar los cambios desde cualquier máquina que posea un navegador Web dentro de nuestra intranet, o desde Internet. Esto obliga a implementar un sistema de seguridad si cabe más efectivo, ya que con estas herramientas se puede conseguir un acceso total a la información que debe permanecer oculta para el resto de los usuarios.

3.Implementación

Una vez analizados los módulos en los que se va a dividir el servidor WWW y teniendo en cuenta los requisitos que sería recomendable reuniese un servidor Web, se llega a la conclusión de que se debe implementar con un lenguaje de alto nivel, como por ejemplo Java, y almacenar la información en un SGBD, utilizando el lenguaje SQL para acceso y manipulación de esta información.

El lenguaje Java tiene como principal característica que se puede ejecutar en varias plataformas sin tener que recompilar el código siendo únicamente necesario un intérprete de Java específico para cada plataforma. Esto nos permite

obtener aplicaciones Web multiplatafoma con relativa sencillez. Otras ventajas que ofrece Java y que resultan útiles a la hora de implementar un servidor de las características enunciadas anteriormente, son las siguientes :

- la utilización de las propiedades de herencia de clases y polimorfismo de Java, permite la implementación de la arquitectura modular descrita de forma secilla y facilmente extensible..

- la seguridad que proporciona Java es fundamental en cuanto a la privacidad de los datos que se manejan, autorización y autentificación de los usuarios a la hora de introducir cambios en la informacion que muestra el servidor y por último protección del servidor frente a las intromisiones en el sistema.

- la posibilidad de implementar mediante Java la multitarea resulta de gran utilidad a la hora realizar varias acciones concurrentemente, acelerando, por tanto, la ejecución

- los mecanismos de gestión de memoria que utiliza Java, facilita la labor de programación de las aplicaciones al desarrollador. También es útil el sistema de direccionamiento que tiene Java, las direcciones son simbólicas hasta el tiempo de ejecucion lo que facilita la seguridad.

Estas son las principales caracteristicas de Java y cubren los requisitos deseables para un servidor WWW : multiplataforma, seguridad en cuanto a los accesos al servidor y del contenido que este muestra al cliente, y multitarea.

Otra de las caracteristicas que debe tener un servidor Web es la consistencia de la información que almacena y muestra. Esto está garantizado organizando los datos de interés tanto del usuario como del administrador o propietario del servidor (estadísticas, históricos ...) en un SGBD relacional y accediendo a ellos a través del lenguaje SQL. Los SGBDs permiten almacenar procedimientos SQL que se ejecuten cuando se den unas determinadas condiciones establecidas por el administrador, facilitando la autogestión del servidor. La utilización de procedimientos almacenados en el SGBD permite reducir al máximo la información que han de intercambiar cliente y servidor Web, al permitir un preprocesamiento o filtrado de la información que solicite el cliente.

En los SGBD relacionales se puede proteger la información y los accesos a ésta estableciendo varios niveles de acceso y evitando bloqueos, algo fundamental en un servidor WWW de con un alto grado de concurrencia..

Es decir, que mediante la utilizacion de Java y SQL podemos implementar la arquitectura explicada anteriormente logrando un servidor WWW que se autogestione además de que cubra otros aspectos importantes para una aplicación de estas características.

Desde el punto de vista de las aplicaciones Web, cuando se trabaja contra un SGBD, un factor determinante es la forma en que se realiza la conexión con el mismo. Básicamente existen dos posibilidades, utilizar los métodos de acceso nativos (APIs) al SGBD proporcionados por el fabricante, o usar algún mecanismo estándar de acceso genérico a bases de datos (ODBC *Open Data Base Conectivity*, JDBC, etc).

A priori el uso de las APIs nativas para acceder a las bases de datos proporciona un mayor rendimiento y mejor aprovechamineto de las posibilidades que ofrece cada SGBD. Por otro lado el uso de mecanismos como ODBC o JDBC simplifica la implementación de aplicaciones que usen información proveniente de SGBDs de distintos fabricantes, permitiendo la realización de aplicaciones independientes de los SGBDs utilizados. Es necesario, por tanto, llegar a un compromiso entre rendimiento y flexibilidad.

Actualmente, el diseño modular de la arquitectura y su implementación utilizando clases de Java, nos permite efectuar facilmente mejoras mediante la especialización de las clases que proporcionan el acceso al SGBD. De este modo podemos conservar la funcionalidad de acceso al SGBD utilizando APIs nativas, para aquellas aplicaciones en las que el rendimiento de ejecución sea un requisito prioritario.

Una de las mejoras, en las que ya se está trabajando, consiste en extender el módulo de acceso a los datos, para permir la utilización de JDBC para acceder al SGBD en aquellas aplicaciones en las que sea necesario (figura 2).

El API JDBC 1.10 (JavaSoft) define un conjunto de clases Java para representar conexiones a bases de datos, sentencias SQL, conjuntos de resultados, información de la estructura de la base de datos, etc. Esto permite al programador de Java ejecutar sentencias SQL y procesar los resultados independientemente del SGBD utilizado. JDBC es el API primaria para el acceso a bases de datos relacionales desde Java. JDBC es en la actualidad un componente estándard de Java, y está incluido en el JDK 1.1.

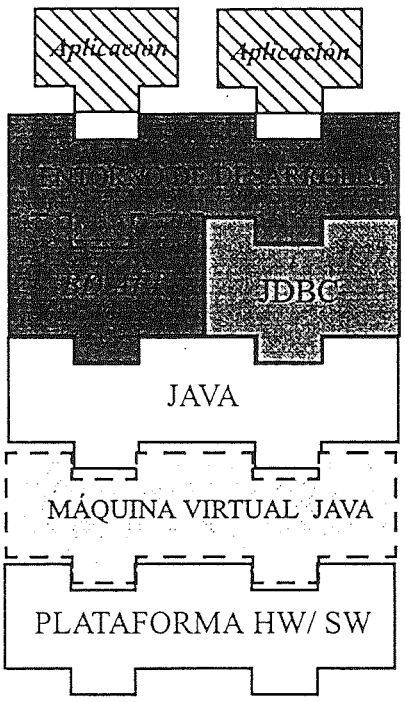


Figura 2: Capas de software para uso de JDBC

Gracias a esta mejora la implementación final de una aplicación WWW sería independiente de la plataforma *hardware* y *software* por la utilización de Java, e independiente del SGBD utilizado gracias a JDBC.

4. Conclusiones

La arquitectura propuesta supone un avance en el desarrollo y mantenimiento de aplicaciones Web, mediante la integración de tecnologías tales como SGBDs o Java. El objetivo de la arquitectura consiste en facilitar la implementación de aplicaciones Web autogestionadas. La principales aportaciones de estas aplicaciones son:

- automatización de los procedimientos de introducción de información,
- actualización de la información en tiempo real,
- generación automática de informes sobre el contenido y uso del servidor WWW.

Referencias

[1] M. Anand, E. chien, R. Condamoor, A. Mathaur, S. Adunuthula, S. Chou, and S. Nakhoda. "The Web Request Borker: A Framework for Distributed Web-Based Applications". Oracle Corporation, Redwood Shores, CA.

[2] "Netscape API Functions".Netscape Communications, Mountain View, CA, 1996.

Teletrabajo

Compartición de aplicaciones bajo una arquitectura replicada

JESÚS M. HERRERO (ROBOTIKER)
JAVIER OLIVER (ESIDE. Universidad de Deusto)
Correo electrónico: jesus@robotiker.es, oliver@deusto.es

Abstract:

At present there are available on the market several document conferencing systems which allow users to work together with the same applications at the same time. These systems follow a centralized architecture by which the output of one node is distributed to all conference sites. This method generates heavy network traffic. In this article a system based on a replicated architecture is presented. With this model input is generated at one of the conference sites and is distributed to all remote copies of the tool. This method produces significantly less traffic but presents serious problems in the maintenance of synchronism and consistency. These problems are analysed and several solutions are proposed.

1. Introducción

Durante años las herramientas y aplicaciones informáticas han sido utilizadas individualmente, sin que se hayan establecido relaciones con otras personas cercanas que empleaban sistemas similares [1]. Este aislamiento empezó a romperse con la aparición de la tecnología de red que interconectaba varios equipos informáticos y permitía la compartición de recursos como ficheros, programas o impresoras. Posteriormente surgió el correo electrónico que facilitaba la comunicación grupal. Estas tecnologías son los primeros ejemplos del soporte que los sistemas informáticos pueden dar al trabajo en grupo.

Según datos recogidos por DATAPRO, muchos procesos de trabajo son todavía lentos a pesar del grado de automatización existente [2]. Algunos expertos afirman que durante la década de los años 80 la industria americana no ha mejorado significativamente su productividad y sin embargo ha doblado sus inversiones en elementos hardware y software. Como ejemplo se presenta una aplicación de contratación de seguros de vida que tarda 22 días en generar la póliza. De ese tiempo, el proceso de aceptación o rechazo del seguro dura 17 minutos. El resto del tiempo se utiliza en procesos de interacción y comunicación entre el grupo de trabajo y personas externas [2]. Las herramientas informáticas individuales no pueden mejorar estos procesos. Deben ser abordados por sistemas que den soporte al trabajo en grupo y permitan que una tarea sea llevada a cabo por varias personas. La tendencia en la década de los 90 es pasar del concepto de la productividad personal al concepto de la productividad de la organización [3].

2. CSCW y Groupware

2.1 CSCW

Desde mediados de los años 80 empieza a utilizarse el término *Computer Supported Cooperative Work* (CSCW) (Trabajo Cooperativo Soportado por Ordenador) para definir el campo de investigación que aborda cómo los sistemas informáticos pueden dar soporte al trabajo grupal. Este área intenta superar la barrera del ordenador personal y aproximarse a los procesos reales de trabajo en los que existe una interacción continua entre diferentes personas.

El término CSCW fue utilizado por primera vez en 1984 por Greif del MIT y Cashman de DEC, quienes organizaron un seminario al que asistieron unas 20 personas de diferentes disciplinas interesadas en analizar cómo se organizan los grupos y cómo llevan a cabo sus trabajos. Dos años después, en 1986, ya se celebró la primera conferencia sobre CSCW en Estados Unidos y en 1989 tuvo lugar la primera conferencia europea [4].

Como todo nuevo campo de investigación su ámbito de trabajo, sus límites, su definición e incluso su propio nombre son objeto de discusión. Greif, creadora del término, afirma que lo escogieron para poder referirse a los aspectos relacionados con un grupo de personas trabajando conjuntamente con la ayuda de sistemas informáticos. Y reconoce también que no esperaban que se hiciera tanto hincapié en cada una de las palabras que forman el término y sus posibles implicaciones.

Bannon presenta el término CSCW como un concepto amplio y lo describe como [5]:

El campo que cubre todo lo relacionado con el soporte informático de las actividades en las cuales están involucradas más de una persona.

Según Bannon, en este foro deben participar investigadores de diferentes especialidades. Este carácter multidisciplinar del CSCW es defendido por la mayoría de los autores [4] [6].

Por otra parte, Ellis propone otra definición general [7], pero haciendo más hincapié en el aspecto sociológico del trabajo cooperativo, así afirma que:

El CSCW investiga cómo trabajan los grupos e intenta descubrir la forma en que la tecnología y especialmente los ordenadores pueden facilitar ese trabajo.

La conjunción de la potencia de los ordenadores y las nuevas formas de comunicación electrónica deben mejorar, según Ellis, la interacción entre las personas y posibilitar la creación de sistemas que integren el procesamiento de la información y las actividades de comunicación. No obstante, Ellis ya avanza que la principal dificultad para desarrollar sistemas eficientes va a estar en los aspectos sociales y de organización.

En CSCW se integran dos tareas fundamentales y complementarias como son el análisis de cómo trabajan las personas y los grupos, y el desarrollo de prototipos y sistemas informáticos que intentan facilitar los procesos de trabajo realizados por varias personas.

2.2 Groupware

Relacionado con CSCW aparece otro término: Groupware. Tampoco existe una única definición del mismo. Fue utilizado por primera vez por Johnson-Lenz en 1982, y por tanto antes que el propio concepto de CSCW. Hacía referencia a los procesos y los procedimientos que llevaban a cabo los grupos de trabajo junto con los sistemas informáticos utilizados [8]. Posteriormente Johansen en 1988 restringió la definición únicamente a los sistemas informáticos que daban soporte a un grupo de trabajo [9]. Este último planteamiento es el seguido por Ellis, cuya definición es hoy en día la más utilizada dentro de la comunidad que estudia el trabajo cooperativo [7]:

Sistemas que dan soporte a grupos de personas involucradas en una tarea común (o en un objetivo común) y que proporcionan una interfaz a un entorno de trabajo compartido.

En esta definición, los elementos tarea común y entorno de trabajo compartido resultan claves a la hora de determinar si un sistema es un producto groupware. Por ejemplo, no se consideran productos groupware los sistemas multi-usuario como las bases de datos ya que quienes los utilizan no comparten un objetivo común. Por otra parte, esta definición incluye sistemas de correo electrónico en los que el concepto de cooperación es más débil.

2.3 Relación entre CSCW y Groupware

Nos encontramos con dos términos, CSCW y Groupware, que para algunos autores no son más que dos nombres para un mismo campo de investigación, que además también recibe otras denominaciones como *Coordination Technology, Group Decision Support Systems (GDSS), Electronic Meeting Systems (EMS), Collaboration Technology*, etc [10] [11].

El término Groupware fue utilizado inicialmente para hacer referencia a los procesos grupales y los sistemas informáticos que les daban soporte. Esta definición puede aplicarse actualmente al término CSCW. La definición de Groupware más extendida en estos momentos es la propuesta por Johansen y Ellis que incluye únicamente los sistemas informáticos que dan soporte al trabajo en grupo.

Se puede, por tanto, considerar al CSCW como el área de investigación que aborda el trabajo cooperativo y su soporte informático, y al Groupware como el área que estudia y desarrolla sistemas informáticos para dar soporte al trabajo cooperativo [3] [6] [11]. De forma clara y escueta esta relación queda descrita en el título del libro de Marca y Bock: *"Groupware: Software for Computer Supported Cooperative Work"* [12].

Siguiendo esta relación entre CSCW y Groupware, Kremar distingue tres campos de investigación dentro del área general de CSCW [11]:

1. Estudio de los grupos de trabajo. Se encarga de profundizar en cómo los grupos de personas se organizan y llevan a cabo una tarea, teniendo en cuenta los factores que intervienen: la coordinación, la comunicación o la colaboración. Es un área que requiere la participación de diversas disciplinas como la sociología, la psicología o las teorías de organización.

2. Desarrollo de herramientas que dan soporte al trabajo cooperativo. Es el concepto de Groupware que se utiliza en este artículo.

3. Evaluación de los sistemas. Estudia los criterios necesarios para evaluar los sistemas de ayuda al trabajo en grupo. Se debe analizar asimismo el impacto que provoca su introducción en una organización y los cambios que es necesario llevar a cabo en los hábitos adquiridos de trabajo.

El campo de investigación de CSCW requiere un esfuerzo interdisciplinar. Ellis utiliza la palabra interdisciplinar en lugar de multidisciplinar para resaltar que los análisis y las contribuciones de las diferentes disciplinas (incluyendo los usuarios finales) deben ser integradas y no únicamente tenidas en cuenta [7]. El éxito de los sistemas Groupware no depende en este momento tanto del desarrollo de nuevas tecnologías como de la comprensión del funcionamiento de los grupos de trabajo y del diseño de aplicaciones que satisfagan realmente los requisitos de los usuarios que forman parte de esos grupos.

3. Clasificación de los sistemas

Existen diversos criterios de clasificación de los sistemas Groupware. Algunos se basan en colocar todos los sistemas en un espectro multidimensional definiéndose diferentes dimensiones como *tarea común* o *entorno de t*rabajo compartido [7]. Los sistemas ocuparán un lugar en ese espacio dependiendo del grado de cumplimiento de estos requisitos.

Otro criterio ampliamente utilizado consiste en una tabla espacio-temporal que crea 4 cuadrantes en función de las coordenadas: mismo tiempo (cooperación sincrónica) o diferente tiempo (cooperación asíncrona) y mismo lugar o diferente lugar [9]. Cada sistema se colocará en un cuadrante.

En este artículo se va a presentar una clasificación que responde a las funcionalidades básicas que presentan los sistemas.

1. Sistemas de Mensajes. Para la mayoría de los autores se trata de los sistemas Groupware de mayor éxito [6] [7]. Se pueden distinguir los sistemas tradicionales de correo electrónico (ej.: cc:Mail, Miscrosoft Mail) en que el destinatario es uno o varios usuarios y las conferencias sobre diferentes temas (ej.: News, BBS). En los primeros se mantiene un buzón individual por cada usuario donde es depositada la información mientras que en los últimos el usuario se conecta a una repositorio central para acceder a la información.

2. Sistemas de Conferencia en Tiempo Real. Estos sistemas permiten a un grupo de usuarios que se encuentran en diferentes lugares interactuar simultáneamente a través del ordenador. En la actualidad esta comunicación puede ser a través de diferentes medios como el vídeo, el audio y los documentos. En este artículo nos vamos a centrar posteriormente en aquellas aplicaciones que permiten la creación de documentos que pueden ser modificados en tiempo real por cualquier usuario que participa en la conferencia.

3. Sistemas de Soporte a Reuniones. En general se trata de sistemas que proporcionan un apoyo para organizar y desarrollar una reunión. Se pueden distinguir varios tipos como los Sistemas de Agenda Electrónica que se encargan de determinar la hora óptima para desarrollar una reunión gestionando las diferentes agendas individuales de las personas involucradas. Existen en el mercado gran variedad de ejemplos como *Network Scheduler* [13] o *SCHEDULE+* de Microsoft [14]. Otro tipo son los Sistemas de Soporte al Desarrollo de las Reuniones *(Electronic Meeting Systems)* que se instalan en una sala que normalmente cuenta con una pantalla de proyección y unos equipos informáticos conectados entre sí. Proporcionan diferentes ayudas para el desarrollo de la reunión como sistemas de votación, de recolección de propuestas o de ayuda en la toma de decisiones. Algunos ejemplos son los sistemas *CoLab* desarrollado por Xerox o el proyecto *NICK* del MCC [15].

4. Sistemas de Creación Compartida de Documentos. Permiten la cooperación asíncrona en la creación de un documento por varias personas en diferentes secuencias de tiempo. El ejemplo más característico es el sistema *Quilt* desarrollado por Bell [16].

5. Sistemas Workflow. Se encargan de dividir los procesos en una secuencia ordenada de tareas, encaminan el flujo de información de una tarea a otra y determinan el papel que deben tomar las personas involucradas en cada tarea. El sistema *Action Workflow System* de Action Technologies es un ejemplo de este tipo de sistemas [13].

En esta clasificación funcional puede que un sistema pertenezca a más de un grupo. Por ejemplo, los sistemas de creación compartida de documentos se pueden considerar un tipo específico de sistemas workflow.

4. Sistemas de Conferencia en Tiempo Real

Los Sistemas de Conferencia en Tiempo Real permiten a un grupo de usuarios, que pueden estar en diferentes localizaciones físicas, interactuar simultáneamente a través del uso de aplicaciones informáticas.

Por medio de la utilización compartida de estas aplicaciones se puede crear, visualizar o modificar un documento con un procesador de textos o una hoja de cálculo, se puede editar conjuntamente un gráfico o se puede discutir y tratar de corregir en grupo un código fuente.

Dentro de los Sistemas de Conferencia en Tiempo Real existe un nivel de clasificación atendiendo a la arquitectura de los mismos distinguiéndose entre sistemas de arquitectura centralizada y sistemas de arquitectura replicada.

4.1 Sistemas de Arquitectura Centralizada

En estos sistemas se ejecuta sólo una copia de la aplicación cooperativa. Los eventos de entrada generados por cualquier usuario son enviados al nodo donde se encuentra la aplicación. Este nodo genera la salida la cual se distribuye al resto de usuarios (Fig. 1). La cooperación se basa en la distribución de las salidas. Estas salidas son las ventanas que reflejan el estado actual de la aplicación.

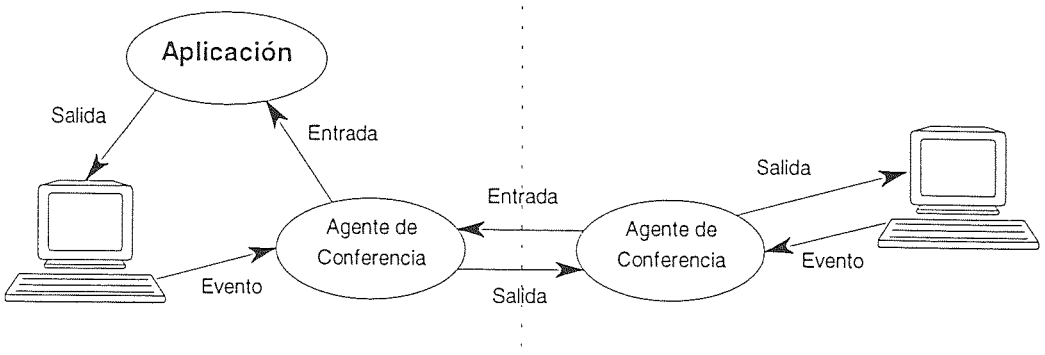


Fig. 1 - Sistema de Arquitectura Centralizada

La principal ventaja de este modelo es su consistencia, ya que al distribuirse constantemente las ventanas que reflejan el estado de la aplicación se garantiza que todos los usuarios visualizan siempre la misma información. Pero, a su vez, presentan importantes inconvenientes como el alto tráfico de red o el alto tiempo de respuesta que pueden presentar.

En estos momentos existen bastantes aplicaciones comerciales que siguen este modelo. Algunos ejemplos son: *FarSite* de DataBeam, *ProShare* de Intel, *TalkShow* de Future Lab y *Smart 2000* de Smart Technologies [17]. En algunos casos estas aplicaciones se presentan como productos independientes y en otros se integran en sistemas de videoconferencia de sobremesa.

4.2 Sistemas de Arquitectura Replicada

Estos sistemas se caracterizan porque se ejecuta una copia de la aplicación en cada nodo que participa en la conferencia. Los eventos de entrada generados en cualquier nodo son distribuidos al resto de usuarios participantes. Las salidas son reproducidas en cada nodo a partir de las entradas recibidas. Este modelo basa la cooperación en la transmisión de las entradas y en la reproducción de las mismas en cada nodo (Fig. 2).

La principal ventaja de estos sistemas es el reducido tráfico de red que generan comparados con los anteriores. También el tiempo de respuesta es menor ya que la salida se obtiene localmente. Por otra parte resulta difícil mantener la consistencia ya que hay que garantizar que la secuencia de entradas que se procesan en cada nodo produce siempre las mismas salidas. Este hecho presenta serias dificultades. Además el tratamiento erróneo de un solo evento provoca ya que esa aplicación no mantenga el mismo estado que el resto, por lo que la cooperación se ve seriamente afectada.

Estas dificultades son las razones por las que estos sistemas han sido desarrollados hasta la fecha por equipos de investigación y no han pasado todavía el umbral de la comercialización. Algunos de los proyectos involucrados en el desarrollo de este tipo de sistemas son: *MMConf* [18], *VConf* [19], *Dialogo* [20], *Rapport* [21] y *MERMAID* [22].

Existe un continuo debate dentro de este área sobre la viabilidad de los sistemas replicados. Algunos autores se decantan de manera preferente por los sistemas centralizados por su consistencia, pero otros consideran necesario seguir profundizando en el estudio de los sistemas replicados para poder aprovechar las importantes ventajas que presentan:

* *"En general, creemos que los beneficios del modelo de arquitectura replicada son de la suficiente importancia como para continuar profundizando en ellos"* [20].

* *"Uno puede preguntarse por qué todavía se presta atención al modelo replicado. La respuesta tiene que ver con el concepto de latencia, que es el tiempo que un paquete de datos permanece en la red"* [23].

* *"Todavía no existe en estos momentos una respuesta clara sobre cual de los modelos de arquitectura centralizada o replicada es el más adecuado para las aplicaciones Groupware"* [23].

5. Descripción del sistema *DocuLAN*

DocuLAN se enmarca dentro de los sistemas de conferencia de documentos en tiempo real. Está construido siguiendo un modelo de arquitectura replicada. Por tanto, en cada nodo se va a lanzar las aplicaciones sobre las que se va a cooperar. El sistema va a permitir que las aplicaciones que se utilizan de forma individual en el entorno *MS-Windows 3.x* puedan ser utilizadas por un grupo de personas simultáneamente. Varios proyectos han abordado el desarrollo de sistemas replicados, pero todos ellos han sido desarrollados sobre el entorno UNIX.

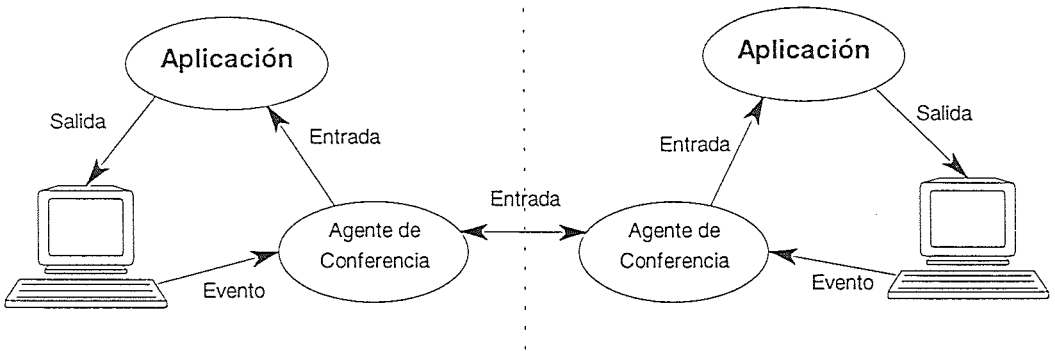


Fig.2 - Sistema de Arquitectura Replicada

El sistema se ha separado en varios módulos independientes. Incluso es posible modificar uno de los módulos para adaptarlo a un entorno concreto sin afectar al resto del sistema. De esta forma, por ejemplo, se pueden utilizar diversos protocolos de comunicación como TCP/IP y RDSI adaptando únicamente el módulo de comunicaciones al protocolo seleccionado. En la Fig. 3 se recogen los diferentes módulos del sistema *DocuLAN*.

Antes de comenzar con una descripción más detallada de cada uno de los módulos resulta conveniente introducir algunos términos que se van a utilizar en la descripción del sistema.

* Sesión Cooperativa: Es la propia conferencia de documentos en tiempo real en la que participan varios usuarios.
* Aplicaciones Cooperativas: Son las aplicaciones que se utilizan en una sesión cooperativa.
* Moderador: Es el usuario que inicia la Sesión Cooperativa.
* Participantes: Son los usuarios que toman parte en una Sesión Cooperativa, exceptuando el Moderador.
* Testigo: Es el elemento cuya posesión determina el usuario que puede escribir sobre las aplicaciones cooperativas.
* Usuario Activo: Es el usuario que en un momento dado posee el Testigo.
* Usuario Pasivo: Es el usuario que no posee el Testigo y, por tanto, no puede actuar sobre las aplicaciones cooperativas.

5.1 Módulo de Gestión de las Aplicaciones

Este módulo se encarga de todas las tarea relacionadas con el lanzamiento y la identificación de las aplicaciones cooperativas. En una primera fase el usuario que inicia una sesión cooperativa, el Moderador, selecciona las aplicaciones y ficheros sobre los que desea cooperar. A continuación transfiere esta información a los usuarios que van a tomar parte en esa sesión, los Participantes.

En una segunda fase se deben abrir las aplicaciones cooperativas en cada uno de los nodos. Para garantizar el mantenimiento del sincronismo es imprescindible que una misma aplicación se abra con el mismo estado en cada uno de los nodos. Esta condición constituye una dificultad seria ya que, cada vez en mayor medida, las aplicaciones cuentan con numerosas opciones de configuración que influyen en su apariencia externa. Para solucionar este problema se toma como referencia las opciones de configuración del Moderador y así cuando se transfiere la información de la sesión también se transfieren los ficheros de configuración (normalmente con extensión *ini*) a los Participantes. De esta forma, se garantiza que la misma aplicación se abre con el mismo estado en cada uno de los nodos. Una vez finalizada la sesión se restauran las opciones de configuración que se utilizaban en cada nodo.

Otro de los factores que es necesario gestionar para garantizar el mismo estado inicial es la resolución de las pantallas de los usuarios. Para ello el sistema *DocuLAN* adopta como resolución cooperativa la menor de las resoluciones de los usuarios que toman parte en una sesión (ej. 640x480). *DocuLAN* impedirá que cualquier aplicación, aun pemitiéndoselo la pantalla de ese nodo, sobrepase el espacio de la resolución cooperativa. Además inicialmente el sistema gestiona que las ventanas cooperativas se coloquen en el mismo lugar y con el mismo tamaño en cada una de las pantallas. Todo este proceso se realiza de forma transparente al usuario.

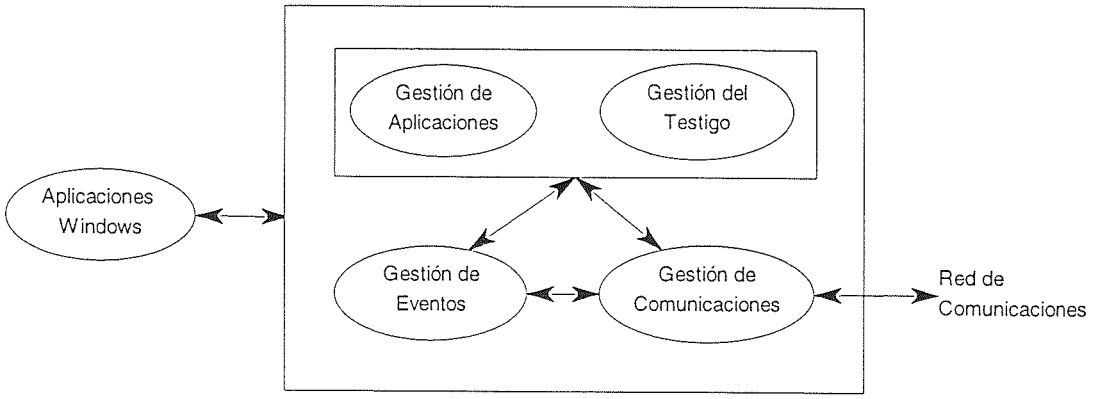


Fig. 3 - Descripción del sistema *DocuLAN*

Este módulo permite, asimismo, la realización del trabajo privado, que es el que realiza un usuario con aplicaciones que no forman parte de la sesión cooperativa. Este trabajo privado se puede realizar mientras no se generen eventos cooperativos. Si llega un evento de este tipo, debe ser procesado inmediatamente, para lo cual la ventana cooperativa correspondiente debe ser previamente activada.

Para garantizar el mantenimiento del sincronismo y por tanto de la cooperación durante el desarrollo de la sesión se han deshabilitado algunas funcionalidades como la acción de minimizar las ventanas cooperativas.

Una de las principales dificultades en los sistemas de arquitectura replicada es la posibilidad de permitir a nuevos usuarios sumarse a una sesión cooperativa ya iniciada. Para permitir esta acción, sería necesario reconstruir toda la historia de los eventos generados a partir del estado inicial de la sesión, lo que implica grandes consumos de recursos. No obstante, en *DocuLAN* se ha implementado una solución que no implica el almacenamiento de todos los eventos, aunque impone unas restricciones. Consiste en que el nuevo participante realiza una petición de entrada a una sesión en curso. El Moderador decide sobre su incorporación y en caso afirmativo, el mismo Moderador guarda los cambios realizados en los archivos, cierra la sesión y la reinicia inmediatamente incorporando al nuevo usuario. Este proceso provoca lógicamente una interrupción de la cooperación, pero por otra parte resuelve uno de los principales inconvenientes de la arquitectura replicada.

5.2 Módulo de Gestión del Testigo

Este módulo es responsable de otorgar a los usuarios el permiso correspondiente para poder trabajar sobre las aplicaciones cooperativas.

En este sistema, existe un único testigo en cada sesión y por tanto sólo un usuario puede en cada instante escribir sobre las aplicaciones cooperativas. El resto de usuarios podrá solicitarlo u obtenerlo inmediatamente dependiendo de la política de gestión del testigo que se utiliza.

El sistema ha sido comprobado con cuatro políticas de intercambio:

1. Dirigida por el Moderador. El Moderador decide el usuario que va a poseer el testigo en cada instante.
2. Dirigida por el Poseedor del Testigo. El usuario que posee el testigo se lo pasa a otro usuario.
3. Por peticiones. El usuario que desea el testigo realiza una petición, la cual es almacenada en una cola del tipo *First Come First Served*. Cuando el poseedor actual libera el testigo, éste pasa al primer usuario de la cola.
4. Por prioridades. Es similar a la política anterior, excepto que las peticiones son ordenadas en la cola dependiendo de la prioridad asignada previamente a cada usuario.

No se puede afirmar cuál es la política óptima. En cada caso, dependiendo de las características de los participantes y de los objetivos de la reunión habrá que seleccionar la más conveniente.

5.3 Módulo de Comunicaciones

Este módulo es responsable de la transmisión de la información entre los participantes en una sesión cooperativa. Esta información puede estar constituida por los archivos de datos, los archivos de configuración, los eventos cooperativos e información de gestión de la sesión.

Hay que destacar que el diseño modular del sistema permite utilizar diferentes protocolos de

comunicaciones sin tener que modificar el resto del sistema. El sistema *DocuLAN* ha sido evaluado con tres módulos diferentes de comunicaciones:

1. Módulo punto a punto sobre TCP/IP. Permite la cooperación entre dos usuarios. Por su sencillez ha sido el módulo más utilizado para estudiar todos los temas relacionados con la consistencia y el mantenimiento del sincronismo.
2. Módulo multipunto con servidor sobre TCP/IP. Este módulo fue utilizado en el proyecto ESPRIT MMTCA [24]. Incluía un servidor que proporcionaba una mayor seguridad en el envío de los datos y permitía establecer una conferencia entre más de dos usuarios.
3. Módulo punto a punto sobre RDSI. Permite la cooperación entre dos usuarios conectados a través de la Red Digital de Servicios Integrados (RDSI).

5.4 Módulo de Gestión de los Eventos

Este módulo es la base de la cooperación ya que se encarga de la captación de los eventos en el nodo donde se originan y su reproducción en el resto de nodos participantes. Estas tareas se lleva a cabo fundamentalmente a través del mecanismo de los filtros (*hooks*). Los filtros permiten identificar y capturar los eventos antes de que sean procesados en el entorno *MS-Windows* y permiten en algunos casos la modificación de sus parámetros o su propia eliminación, suprimiendo ese evento de la cola del sistema [25] [26].

En el nodo activo se captan los eventos cooperativos y se envían al módulo de comunicaciones para su transmisión al resto de usuarios. En los nodos pasivos se reciben los eventos a través del módulo de comunicaciones y se pasan al modulo de gestión de eventos, donde son almacenados en una cola de entrada. La función del filtro de reproducción toma los eventos de esta cola y los reproduce sobre la aplicación cooperativa correspondiente.

Para garantizar la cooperación este módulo es responsable del mantenimiento del sincronismo. Debe garantizar que los mismos eventos y en el mismo orden sean procesados en cada nodo y debe garantizar que estos eventos producen la misma salida. Para ello se han implementado una serie de mecanismos que se describen a continuación.

a) Mecanismo de eliminación de eventos no cooperativos

Se encarga de desactivar, en los nodos que no poseen el testigo, los eventos que van dirigidos a las aplicaciones cooperativas. Si estos eventos no fuesen desactivados, se tendría una política libre de intercambio de testigo, por la cual todos los usuarios podrían actuar simultáneamente sobre las aplicaciones cooperativas.

b) Mecanismo de pinchar y arrastrar

Uno de los principales problemas para el mantenimiento del sincronismo ha sido ocasionado por la secuencia de acciones de pulsar, arrastrar y liberar un botón del ratón. En la mayoría de los casos estas acciones no ocasionan ningún problema, como por ejemplo, al mover una ventana. Pero hay determinadas situaciones en la que esta acción es problemática. Esto ocurre en las aplicaciones gráficas (ej. *MS-Paintbrush*) cuando se dibuja una línea, un rectángulo u otra figura. Esta acción se lleva a cabo pulsando inicialmente el botón, arrastrándolo hasta completar la figura y liberándolo. Durante el tiempo que dura esta operación la ventana sobre la que se dibuja posee el control total del entorno *MS-Windows* y durante ese tiempo ninguna otra ventana puede procesar mensajes que tenga en espera. Este hecho impide a la propia aplicación *DocuLAN* procesar más mensajes y se bloquea el sistema. Este comportamiento se debe a que el entorno *MS-Windows 3.x* no es un entorno de multitarea real.

Para poder solucionar este problema, se ha implementado un mecanismo por el cual los eventos comprendidos entre la pulsación y la liberación del botón del ratón únicamente pueden empezar su reproducción cuando todos ellos se encontraran ya en el propio nodo.

c) Mecanismo de pulsación doble

La introducción del mecanismo anterior resuelve un problema, pero crea otro nuevo en la acción de doble pulsación del botón del ratón. Esta pulsación doble provoca una acción determinada si el intervalo entre ambas pulsaciones es menor que un valor configurable en cada entorno *MS-Windows*.

De la forma en que inicialmente se ideó el mecanismo anterior, cuando en el nodo pasivo se recibía el evento de liberación del ratón se transmitía toda la secuencia a la cola general de eventos. No se tenía en cuenta la posibilidad de que tras esa primera pulsación se produjera inmediatamente otra similar que provocaría en el nodo activo una pulsación doble, mientras que en los nodos pasivos se traduciría como dos pulsaciones individuales consecutivas. Ambas acciones no provocarían los mismos resultados en las aplicaciones cooperativas por lo que al perderse el sincronismo, se perdería la consistencia y la cooperación.

La solución ha consistido en introducir un control adicional en la cola en la que se almacenan la secuencia de eventos de las pulsaciones del ratón. Cuando se recibía el evento de liberación del ratón se espera un tiempo para comprobar que no llegaba

otro evento de pulsación dentro del periodo máximo de pulsación doble.

Para que este mecanismo sea válido y en general para evitar la pérdida del sincronismo, durante una sesión cooperativa todos los nodos deben poseer el mismo valor del tiempo máximo de pulsación doble. Este ajuste es realizado por el sistema al inicio de la sesión pasando el valor del Moderador al resto de usuarios. Una vez finalizada la sesión, el valor antiguo es restaurado en cada nodo.

d) Mecanismo sobre situaciones no deterministas

Una situación no determinista se produce cuando partiendo del mismo estado inicial y la misma secuencia de entradas se producen resultados diferentes. Esta situación aparece en la sesión cooperativa cuando los eventos son transferidos correctamente, pero se obtienen resultados diferentes en cada nodo.

El ejemplo más característico y que se ha abordado en este sistema es la utilización de las barras de desplazamiento (*scrolling*). Por ejemplo, si se toma la barra de desplazamiento vertical de un procesador de textos y se mueve el documento una o varias líneas presionando de forma continua las flechas aparecen las situaciones no deterministas.

En el nodo activo, que posee el testigo, al presionar de forma continua la flecha de desplazamiento el documento se moverá un número de líneas que dependerá del número de mensajes de desplazamiento generados. Este número de mensajes dependerá del tiempo que se mantenga presionada la flecha de la barra de desplazamiento.

Al introducir el factor tiempo se produce una situación no determinista en la sesión cooperativa, ya que el número de mensajes que se pueden generar en cada nodo depende de la capacidad de proceso del ordenador que puede ser diferente en cada uno de ellos.

En *DocuLAN* se ha introducido una primera solución de carácter restrictivo. Se ha transformado la opción de pulsación continua de la flecha de desplazamiento. Esta opción va a ser equivalente en todos los casos a la pulsación y liberación inmediata de la flecha. Independientemente del tiempo que se mantenga pulsado el botón, el resultado será el movimiento de una sola línea.

En la Universidad de Paisley en Escocia también se han realizado trabajos en este sentido. Han abordado las situaciones no deterministas en una conferencia de documentos haciendo uso del mismo módulo de cooperación y han obtenido una solución no restrictiva del problema de la presión continua de la barra de desplazamiento [27].

e) Mecanismo de mejora del rendimiento

Para aprovechar al máximo las ventajas de este modelo, entre las que destaca el bajo tráfico de red que genera, se ha incluido también un mecanismo por el cual no se transmiten aquellos eventos que no proporcionan una información significativa. De esta forma se eliminan eventos del movimiento de ratón sin que por ello se pierda el sincronismo ni la cooperación.

Con la inclusión de todos estos mecanismos se alcanza un grado aceptable en el mantenimiento del sincronismo y de la consistencia entre las aplicaciones cooperativas.

6. Comparación con otros sistemas

Como ya se ha indicado, las principales ventajas potenciales de los sistemas replicados es el bajo tráfico de red que generan y el reducido tiempo de respuesta que presentan. En [28] se ha realizado un estudio completo comparando estos factores entre *DocuLAN* y dos de los sistemas de arquitectura centralizada más conocidos como son *FarSite Document Conferencing* de DataBeam y *ProShare Personal Conferencing* de Intel.

Los resultados muestran que la reducción del tráfico de red en *DocuLAN* depende directamente de la relación entre el volumen de datos de salida y al volumen de datos de entrada. Cuanto mayor sea el primero mejores resultados se obtendrán en *DocuLAN*. Por ejemplo, en una aplicación gráfica, la operación de mover una imagen se realiza por medio de un número relativamente pequeño de eventos, que constituyen el volumen de datos de entrada. Mientras que la salida que se genera implica un mayor volumen que dependerá del tamaño y complejidad de la imagen y de la aplicación utilizada. En una operación como la anterior y haciendo uso de la aplicación *CorelDraw* se obtienen reducciones del volumen de información que se transmite del 90%, comparando los datos de *DocuLAN* por una parte y *FarSite* y *ProShare* por otra. Todas las pruebas se han realizado en una red Ethernet con protocolo TCP/IP.

Cuando la diferencia entre el volumen de datos de entrada y de salida disminuye también decrece la mejora que se obtiene. Pero en general las ventajas son bastante significativas. Por ejemplo, en la escritura de un breve texto de 30 palabras con el procesador de textos *MS-Word* se obtiene una mejora media del 80%.

Unicamente se iguala el volumen de tráfico de red entre *DocuLAN*, *FarSite* y *ProShare* en determinadas operaciones que se realizan con

aplicaciones como *Notepad* o *PaintBrush* cuyo interfaz de usuario es bastante sencillo. Estas operaciones que implican un reducido volumen de salida son, por ejemplo, la escritura sobre *Notepad* o dibujar un sencillo gráfico en un documento vacío en *PaintBrush*.

El tiempo de respuesta va a depender de la infraestructura de red que se utilice y del volumen de datos que se genere en cada operación. En general, en un sistema replicado el usuario no percibirá un tiempo de respuesta mayor al que tendría si estuviera trabajando de forma individual con las mismas aplicaciones. En estos sistemas los eventos se reproducen en cada ordenador y por tanto la salida se genera localmente. El único tiempo de espera se produce por la transmisión de los eventos de entrada a cada nodo que participa en una sesión cooperativa.

En cambio, la arquitectura centralizada implica un mayor tiempo de respuesta ya que se debe transmitir la salida que normalmente implica un mayor volumen de información. Pero hay una situación especialmente negativa. En estos sistemas las aplicaciones cooperativas únicamente se lanzan en un nodo, en el Moderador. Cuando posee el testigo un usuario Participante, se producen los tiempos de respuesta más largos. Este usuario genera un evento y para poder visualizar su salida debe esperar tres acciones: los eventos son transmitidos al Moderador, se reproducen dichos eventos en el Moderador y se transmite la salida de la aplicación a los Participantes.

Esta es la situación más problemática ya que el Participante es consciente de que ha generado unas entradas y por tanto está esperando a ver los resultados. Estos se retrasarán un tiempo que dependerá del volumen de salida de esa aplicación. Por ejemplo, este retraso puede alcanzar los 45 segundos al mover una imagen de 100 KB dentro del *PaintBrush*.

En estos momentos existen en el mercado numerosos sistemas de videoconferencia de sobremesa, fundamentalmente sobre RDSI, que integran la comunicación de vídeo y audio con sistemas de conferencia de documentos sobre arquitectura centralizada. En ellos al existir una limitación sobre la capacidad del canal de comunicaciones que se dedica a la conferencia de documentos se obtienen comportamientos poco eficaces en el tiempo de respuesta que dificultan en gran medida la cooperación.

7. Conclusiones

Con este artículo no vamos a solventar el debate sobre los sistemas de arquitectura centralizada y replicada, pero intenta reforzar la idea de aquellos que piensan que hay que seguir considerando la validez de los sistemas replicados. Estos sistemas presentan unas limitaciones iniciales importantes como la necesidad de trabajar sobre el mismo entorno para poder cooperar o que las aplicaciones deban estar correctamente instaladas en cada uno de los nodos. Posteriormente, durante la cooperación, como ya se ha descrito en el artículo, existe un nivel de dificultad en el mantenimiento del sincronismo y de la consistencia de la sesión.

Los sistemas centralizados no cuentan con estas limitaciones y es la razón por la que se pueden encontrar ya numerosos productos en el mercado que siguen este modelo. Pero a pesar de su seguridad también cuenta con limitaciones. Sobre todo el tiempo de respuesta puede alcanzar cotas inaceptables. Incluso el usuario puede tener la percepción de que ha habido algún problema en una operación y puede repetirla antes de que visualice los resultados de la realizada anteriormente.

En determinadas situaciones o entornos las ventajas potenciales que presentan los sistemas replicados pueden ser muy aprovechables, especialmente cuando se trabaja con aplicaciones en las que la diferencia entre el volumen de datos de salida y el volumen de datos de entrada es considerable. Esta circunstancia se produce normalmente en las aplicaciones gráficas. Tampoco se puede hacer una selección exacta de las aplicaciones más optimas ya que esa diferencia entre el volumen de datos también depende de las operaciones realizadas en cada aplicación.

En algunas situaciones puede resultar más sencillo solventar alguna de las limitaciones iniciales de estos sistemas: mismo entorno de ejecución y aplicaciones instaladas en todos los nodos. Aquellas compañías que cuentan con varias delegaciones en diferentes lugares pueden tener necesidad de hacer uso de estos sistemas de cooperación. En esta situación es más probable que los usuarios que deseen trabajar de forma conjunta cuenten con el mismo entorno y hagan uso de las mismas aplicaciones. Esta situación cuenta además con otra ventaja proporcionada por el hecho de que los usuarios mantienen ya algún vínculo al pertenecer a la misma empresa. Según los estudios realizados resulta muy difícil una cooperación entre personas sin ningún vínculo de unión anterior.

Referencias

[1] Rodden, T. *Technological Support for Cooperation.* Department of Computer Science, Lancaster University, UK (1991).

[2] Crooks-Bile, A. "Groupware and Workgroups: Overview". *DATAPRO, Desktop Software & Solutions*, 3501, Groupware, pp. 1-7 (1994).

[3] Marshak, R.T. "Examining Groupware What Is It? Why Use It? What's Going On?" *Workgroup Computing Report*, vol. **17**, no. 2, pp. 3-18 (1994).

[4] Grudin, J. "CSCW: History and Focus". *Computer,* IEEE Computer Society, vol. **27**, no. 5, pp. 19-26, May (1994).

[5] Bannon, L.J., Schmidt K. "CSCW: Four Characters in Search of a Context". *Studies in Computer Supported Cooperative Work.* Eds. J.M. Bowers and S.D. Benford. Elsevier, Netherlands, pp. 3-16 (1991).

[6] Kling, R. "Cooperation, Coordination and Control in Computer-Supported Work". *Communications of the ACM*, vol. **34**, no. 12, pp. 83-88, December (1991).

[7] Ellis, C.A., Gibbs, S.J., Rein, G.L. "Groupware: Some Issues and Experiences". *Communications of the ACM*, vol.**34**, no. 1, pp. 38-58, January (1991).

[8] Johnson-Lenz, P., Johnson-Lenz, T. "Groupware: The process and impacts of design choices". *Computer-Mediated Communication Systems: Status and Evaluation.* E.B. Kerr, S.R. Hiltz. Academic Press, N. Y. (1982).

[9] Johansen, R. *Groupware: Computer Support for Business Teams.* The Free Press, N.Y. (1988).

[10] Palmer, J.D., Fields, H.A. "Computer-Supported Cooperative Work". *Computer,* IEEE Computer Society, vol. **27**, no. 5, pp. 15-17, May (1994).

[11] Kremar, H.A.O. "Computer Supported Cooperative Work - State of the Art". *Human Aspects in Computing: Design and Use of Interactive Systems and Information Management.* Ed. H.-J. Bullinger, Elsevier Science Publishers B.V., pp. 1.113-1.117 (1991).

[12] Marca, D., Bock, G. *Groupware: Software for Computer-Supported Cooperative Work.* IEEE Computer Society Press (1992).

[13] Cummings, J. "Defining Groupware and Other Imposible Tasks". *Business Communications Review*, vol. **24**, no. 2 pp. 35-39, February (1994).

[14] Grudin, J., Palen, L. "Why Groupware Succeeds: Discretation or Mandate?" *Proceedings of the European Conference on Computer Supported Cooperative Work (ECSCW'95).* Eds. H. Marmolin, Y. Sundblad, K. Schmidt. Stockholm, Sweden, pp. 263-278, Kluwer Academic Press, September 10-14, (1995).

[15] Rodden, T. *A Survey of CSCW Systems.* Department of Computer Science, Lancaster University, UK (1991).

[16] Fish, R., Kraut, R., Leland, M., Cohen, M. "Quilt: A Collaborative Tool for Cooperative Writing". *Proceedings of the Conference on Office Information Systems*, Palo ALto, Calif. ACM, New York, pp. 30-37, Mar. 23-25 (1988).

[17] Labriola, D. "Desktop Videoconferencing". *PC Magazine*, pp. 221-254, April 25 (1995).

[18] Crowley, T., Milazzo, P., Baker, E., Forsdick, H., Tomlinson, R. "MMConf: An Infrastructure for Building Shared Multimedia Applications". *Proceedings of the Conference on Computer Supported Cooperative Work (CSCW'90),* Los Angeles, California, pp. 329-342, ACM Press, October 7-10 (1990).

[19] Lantz, K.A. "An experiment in integrated multimedia conferencing". *Proceedings of the Conference on Computer Supported Cooperative Work (CSCW'86),* Austin, Texas, December (1986).

[20] Lauwers, J.C., Joseph, T.A., Lantz, K.A., Romanow, A.L. "Replicated Architecture for Shared Window System: A Critique". *Proceedings of the Conference on Office Information Systems,* ACM Press, pp.303-311, March (1990).

[21] Ahuja, S.R., Ensor, J.R., Lucco, S.E. "A Comparison of Application Sharing Mechanism in Real-time Desktop Conferencing System". *Proceedings of the Conference on Office Information Systems*, pp. 238-248, Boston, April 25-27 (1990).

[22] Watabe, K., Sakata, S., Maeno, K., Fukuoka, H., Ohmori, T. "Distributed Multiparty Desktop Conferencing System: MERMAID". *Proceedings of the Conference on Computer Supported Cooperative Work (CSCW'90),* Los Angeles, California, pp. 27-38, ACM Press, October 7-10 (1990).

[23] Greenberg, S., Marwood, D. *Real Time Groupware as a Distributed System: Concurrency Control and its Effect on the Interface.* Research Report 94/534/03 (1994).

[24] Crowe, M.K. (Ed.). *Cooperative Work with Multimedia.* Research Reports ESPRIT Project 6310, MMTCA, Vol. **1**. Springer-Verlag (1994).

[25] Richter, J.M. *Windows 3.1: A Developer's Guide.* Prentice Hall. M&T Publishung Inc. ISBN 0-13-960543-6 (1992).

[26] Marsh, K. *Microsoft Windows Hooks.* Microsoft Developer Network Technology Group. July (1996).

[27] Tian, S. "Nondeterminate Behaviour: One of the Challenges to Integrating Single User Applications into Groupware Systems". *Computing & Information Systems Department Journal*, pp. 25-31. University of Paisley, Scotland (1995).

[28] Herrero, J.M. *Sistema de Conferencia de Documentos bajo una Arquitectura Replicada.* Tesis en preparación. Facultad de Informática, Universidad de Deusto (1997).

Sistema de Trabajo Cooperativo Soportado por Ordenador para la Enseñanza de Escritura usando el Paradigma de Papel Electrónico

O. M. González[†], M.J. Verdú[††], Y. A. Dimitriadis[††], J.L. Barrio[†††] y M.T. Blasco[†††]

DPTO. DE ORGANIZACIÓN Y GESTIÓN DE EMPRESAS[†]
EUEE, Paseo Prado de la Magdalena, s/n, 47005, Valladolid[†]
DPTO. DE TEORÍA DE LA SEÑAL Y COMUNICACIONES, E INGENIERÍA TELEMÁTICA[††]
ETSIT, C/ Real de Burgos s/n, 47011, Valladolid[††]
DPTO. DE DIDÁCTICA DEL LENGUAJE Y LITERATURA[†††]
FACULTAD DE EDUCACIÓN, C/ Hernández Pacheco 1, 47014[†††]
UNIVERSIDAD DE VALLADOLID

Correo electrónico: oscarg@emp.uva.es, marver@tel.uva.es, yannis@tel.uva.es, tbq@wamba.cpd.uva.es,y jlb@wamba.cpd.uva.es

Abstract:

In this paper we present a new system based on electronic pen for teaching written composition in a collaborative way. The objective of the system is to enhance the process of teaching and learning composition taking advantage of two factors. On one hand, we exploit the natural and simple interaction among professor and students derived by the use of hand-written gestures, annotations and comments due to the use of a pen interface. On the other hand, we support collaboration for the tasks of writing (brainstorming, planning, editing and revising) through the communication based on computer networks. The foundation of communications is a protocol of transmission of graphics events, that provides a more efficient form of information storage, capacity for registering temporal information besides the spatial one for the interaction that take place during a session and possibility for a posterior analysis and processing of all registered information. This system is part of our work toward the creation of virtual classes, and its aforementioned advantages make it suitable for use as a support of teaching "Techniques of Writing" in the Faculty of Education, University of Valladolid, during two academic courses. Current work includes the implementation of an alternative prototype based on multi-agent platform, within a Web-Java environment, as well as a mechanism for recognition of the most common hand-written gestures.

1. Introducción

El aumento del uso de ordenadores en los centros de enseñanza se ha acompañado por los avances en el campo de *redes de ordenadores* y de técnicas generales de comunicaciones digitales, que progresivamente han permitido la integración de distintos tipos de información, tales como voz, datos, imagen o vídeo. Entonces, ordenadores personales, estaciones o servidores, con capacidad de multimedia y aislados, se han podido incorporar en redes LAN o WAN.

Tecnologías subyacentes, tales como RSDI o ATM, junto con el progreso de la Internet basada en TCP/IP, han permitido la reciente explosión de la telemática en diversos campos, con especial énfasis en problemas de educación. Un amplio abanico de redes de comunicaciones pueden apoyar las aplicaciones de telemática en enseñanza, desde las clásicas redes telefónicas conmutadas, X-25, hasta satélites, Internet, RSDI o conexiones sin hilo de ordenadores móviles.

Típicas aplicaciones recientes de telemática en educación incluyen [1]

- *Conferencia electrónica*, que permite la comunicación asíncrona entre los componentes de un curso (profesores, tutores, alumnos).
- *Correo electrónico*, de una forma similar a la conferencia electrónica, ya que permite tanto la comunicación uno-a-uno como la comunicación muchos a muchos (mediante *listas de distribución*).
- *Tablones de anuncios* o Bulletin Board Services (BBS), o *Videotexto*, con posibilidad de acceso a bases de datos educativas, software educativo, intercambio de mensajes y experiencias en temas de educación, etc.
- *Acceso a cursos*, permitiendo el flujo de información y de material educativo entre las instituciones educativas y los alumnos.
- *Creación de un aula virtual* o *grupos virtuales* (cuyos miembros se encuentran distantes geográficamente y participan en instantes de tiempo diferentes), que intercambian mensajes o acceden a un fichero con toda la actividad registrada.

Sin embargo, la enseñanza se entiende principalmente en el entorno de una clase, donde hay interacción social entre los distintos componentes: profesor, alumnos y entorno. El aprovechamiento de este aspecto en el contexto de la aplicación de nuevas tecnologías en la enseñanza sería imposible sin el reciente desarrollo del campo de *CSCW* [2] *(Computer Supported Cooperative Work o Trabajo Cooperativo Soportado por Ordenador)*.

Las investigaciones en el campo de Aprendizaje Colaborativo Soportado por Ordenador o CSCL (*Computer Supported Collaborative Learning*) [3] dirigen sus pasos a entender y proporcionar soporte tecnológico para el aprendizaje colaborativo y cooperativo. El *groupware educacional* se está haciendo cada vez más viable debido a la introducción cada vez más generalizada de redes LAN y WAN, lo que permitirá su uso tanto en aulas electrónicas como en aprendizaje a distancia.

Por otro lado, si queremos simular el entorno natural de un aula debemos pensar en el concepto *de aulas electrónicas* con infraestructuras *móviles* que proporcionen la flexibilidad propia de este tipo de entornos educativos. Y por supuesto, muy unida a esta idea de portabilidad está la necesidad de que los ordenadores sean de tamaño reducido. Esto nos hace pensar en las nuevas formas de interfaz hombre-máquina basadas en escritura a mano o voz, que en conjunción con las conexiones de red sin hilos serán la base de las futuras aulas y, por qué no, de la vida profesional del futuro.

En este artículo se presenta un sistema CSCW basado en lápiz electrónico para la enseñanza y aprendizaje de la composición de textos. El sistema permite la interacción entre alumnos (grupos de 3) y la supervisión e interacción del profesor. El sistema, llamado PENCACOLAS (PEN Computer Aided Composing cOLlAborative System), ha sido el fruto de la colaboración entre investigadores de distintas Facultades y Escuelas de la Universidad de Valladolid durante los años 1995, 1996 y 1997.

El sistema está siendo utilizado por alumnos de la Facultad de Educación como complemento de la asignatura *"Técnicas de Escritura"*, en la cual se les enseña a realizar textos manuscritos utilizando lápiz y papel. El sistema no es simplemente otro sistema de aprendizaje utilizando el ordenador sino que también es una herramienta capaz de analizar las posibles interacciones que se dan en el proceso de la generación de un texto de forma colaborativa, así como el producto final generado por los estudiantes (textos, revisiones, esquemas, etc.).

La configuración actual del sistema es fruto de la experiencia acumulada por profesores de la Facultad de Educación y de las observaciones y problemas surgidos en estos dos cursos. El interfaz utilizado por el sistema está basado en la utilización de un lápiz electrónico.

En este artículo se presenta el sistema, haciendo una breve reseña de los interfaces basados en lápiz electrónico, pasando a continuación a describir las distintas funcionalidades de PENCACOLAS, su uso, los problemas surgidos y las soluciones aportadas e implementadas por los autores. Por ello, el artículo está estructurado de la siguiente forma: en el primer apartado se presentan los sistemas CSCW y las interfaces de lápiz electrónico, a continuación se presentan las características propias del aprendizaje de la escritura de forma colaborativa, después se describe el sistema implementado y su funcionamiento, los distintos problemas y soluciones surgidas durante los dos años de experimentación, se analiza la posibilidad del uso de sistema en una red móvil, y por último se presenta las conclusiones y el trabajo en curso.

2. CSCW y las interfaces basadas en "lápiz electrónico" en el aprendizaje de la Escritura

Debido a la complejidad inherente al proceso de composición, en el transcurso del aprendizaje de cómo componer la mayoría de las veces es necesario deshacerse del trabajo realizado y comenzar de nuevo. Esto significa que el alumno que aprende a componer debe acabar rompiendo muchas páginas para volver a reescribirlas, puesto que es posible que determinados fragmentos del texto no comuniquen suficientemente bien la idea que pretendía transmitir. Cuando se emplea lápiz y papel, este hecho suele ser bastante tedioso y frustrante puesto que supone la reelaboración completa del ejercicio [4]. Empleando el ordenador y un procesador de textos se supera en parte esa dificultad. En primer lugar, porque en la mayoría de los casos no es necesario volver a escribirlo todo de nuevo, sino que bastará con modificar aquellos fragmentos con los que el autor no esté contento para conseguir una mejora considerable del texto. Y segundo, el alumno se vuelve más activo en el proceso de componer ya que sabe que lo que está escribiendo no es de ningún modo definitivo y puede cambiarlo, mejorarlo o eliminarlo en cualquier momento sin mayor problema.

Pero es más, los ordenadores pueden permitir el almacenamiento de los distintos borradores y de las modificaciones, proporcionando mecanismos para la visualización y modificación de todos los productos intermedios, favoreciendo la idea de que un texto escrito es algo que *está siendo construido* y modificado según las necesidades del autor.

Por otro lado, el proceso de aprendizaje de composición mediante el ordenador se puede ver relentizado, al menos en sus fases iniciales, debido a la no familiaridad de los alumnos con los

dispositivos de entrada y de salida tradicionales. Por eso en un contexto educacional toman un interés muy especial las *interfaces basadas en lápiz electrónico* [5]. Es un hecho que, tradicionalmente en las aulas, las ideas y el conocimiento se han transmitido usando lápiz y papel.

Las nuevas interfaces basadas en lápiz electrónico son ideales para que la transición del aprendizaje tradicional al aprendizaje mediante el ordenador sea más suave, y que el alumno interactúe de forma más natural con el ordenador. Además, el lápiz electrónico es superior a otros dispositivos de entrada en muchos aspectos, entre los que se encuentra la posibilidad de utilizar gestos manuscritos, muy útiles en las típicas operaciones de insertar, borrar..... de los editores de texto. Por lo tanto, queda claro que, un ordenador con una interfaz basada en lápiz electrónico, con el software adecuado, puede ser un medio muy eficaz en el proceso de aprendizaje de composición, así como en el proceso de composición en sí.

En la escritura colaborativa hay mucha *información metatextual* que es exteriorizada por los autores (colecciones de ideas, 'outlines', planes y anotaciones). Debido a su naturaleza, cuando esta información se expresa de forma escrita se utilizan flechas, círculos, tachones, etc. Está claro que estas operaciones son especialmente costosas si se realizan con una interfaz basada en teclado y ratón y que la forma más natural e ideal de realizarlas es utilizar una interfaz basada en lápiz electrónico, ya que:

- Proporcionan una interacción más natural con el ordenador y son fáciles de usar para los alumnos, dado que han usado el lápiz desde niños.
- Su tamaño es pequeño, por lo que no son obstáculo para la deseable portabilidad de las "futuras" aulas electrónicas.
- Ofrecen la potencialidad de los gestos, muy útiles y eficaces en las operaciones típicas de edición.
- Facilitan la comunicación entre los colaboradores, ya que son ideales para expresar la información metatextual.
- Proporcionan información *prosódica* [6] de la escritura, es decir, cuando un alumno usa este tipo de interface, no sólo queda registrado lo que escribió sino "cómo" lo escribió, aspecto que con el teclado no se puede registrar.

Por tanto, a parte de que la utilización de interfaces basados en pen aportan una forma natural y sencilla de interaccionar los alumnos con el ordenador, en nuestro sistema la utilización de dicho tipo de interface nos da la posibilidad de almacenar en ficheros de eventos gráficos, no sólo lo que los alumnos escribieron, sino cómo lo hicieron (información prosódica), y qué notaciones y gestos manuscritos realizaron.

Por otra parte, si se quiere ofrecer un medio para la colaboración se debe introducir algún mecanismo de comunicación. La utilización del ordenador y de las redes de ordenadores como soporte para la escritura colaborativa es uno de los objetivos más perseguidos por los investigadores de CSCW [7][8][9]. Y en particular, se le está empezando a dar énfasis a la composición colaborativa en el contexto de un aula, durante el proceso de aprendizaje [10].

3. El aprendizaje de la escritura de forma colaborativa

La escritura colaborativa es un instrumento pedagógico que puede ayudar a los estudiantes a improvisar la negociación y cooperación (procesos muy utilizados en la vida real) así como la adquisición de conocimiento y las habilidades en el proceso de la escritura [11].

Muchos sistemas basan la colaboración en la escritura en un modelo individual en el cual el estudiante genera un documento propio que luego es revisado por otros compañeros, de tal forma que estas revisiones puedan servir para modificar el documento original si el autor lo desea. Nuestro sistema, PENCACOLAS, rompe esta forma clásica de colaboración ya que los usuarios del sistema van pasando por las distintas fases trabajando, de forma individual en unas fases, y de forma colaborativa en otras teniendo que llegar a un acuerdo sobre las ideas generales y un esquema común que serán la base de sus composiciones individuales.

Por lo tanto, es en estas fases cooperativas, donde los usuarios aprenden a colaborar para lograr una meta común, negocian y discuten los diferentes puntos de vista sobre los temas propuestos, ayudándose mutuamente cuando ven que alguno se desvía de la idea inicial o simplemente se atasca. En definitiva colaboran en el aprendizaje tanto del proceso de composición como en la tarea de trabajar coordinadamente y cooperativamente (una de las desventajas fundamentales de la introducción del ordenador en las aulas es que el uso de estos puede potenciar el aislamiento de los usuarios a la hora de trabajar)

En una herramienta diseñada para la enseñanza de la escritura de forma colaborativa se deben tener en cuenta los puntos que acontinuación se enumeran (cuyas soluciones adoptadas por PENCACOLAS son descritas en la sección 5).

P1. *Consciencia del espacio de trabajo compartido:* cada alumno o co-autor debe tener consciencia de lo que él está haciendo y lo que significa en el contexto global, y de lo que hacen los demás [3].

P2. *Implementación de roles*: algunos investigadores [7][8] están a favor de la implementación de *roles*, con accesos restringidos según el rol (autor, revisor, editor...) que se controlan y coordinan mediante un mecanismo de autorizaciones. Por otro lado, hay otros [9][10] que propugnan que no debe haber restricciones técnicas sobre el trabajo de los autores, sino que debe proporcionarse un pequeño soporte para permitir el protocolo social. En lo que sí que coinciden es en que siempre deben existir *áreas privadas* en las que los usuarios trabajan de forma individual, teniendo los demás acceso restringido a estas ares (ningún acceso, acceso de lectura, etc...).

P3. Debe proporcionar *un soporte adecuado para la comunicación* entre los miembros del grupo. Hay que proporcionar mecanismos de comunicación para todos esos casos de interacción. Son necesarios mecanismos de comunicación tanto síncronos como asíncronos. Entre los primeros se pueden destacar la 'videoconferencia', 'enlaces de audio', utilizadas en GROVE [7], voz y superficies de discusión compartidas, junto con mecanismos de telepunteros para señalar, hacer gestos.... Y entre los mecanismos asíncronos, el modelo en el que más coinciden los investigadores es el *modelo de anotaciones* [8][9][12].

P4. Las herramientas deben ser apropiadas al nivel de experiencia de los usuarios, y hay que evitar la distracción. En aquellos casos en los que no se necesite realimentación inmediata se podría implementar *algún mecanismo para que los usuarios descubran las acciones de los otros de forma gradual*, de tal manera que no les provoque distracciones ni les interrumpa bruscamente en su trabajo [7][8]. Por todo esto se deberán estudiar las reacciones de los alumnos, por ejemplo, ante los telepunteros, y procurar usarlos en aquellos casos que proporcionen más beneficios que distracción.

P5. *El papel del profesor*: no hay que olvidar, que el *profesor* es parte del sistema dentro del cual el proceso de aprendizaje tiene lugar. Si la nueva tecnología tiene efectos positivos o negativos sobre el profesor, dichos efectos se reflejarán en el aprendizaje que tiene lugar [13].También hay que considerar, por tanto, el papel del profesor, intentando ayudarle en su tarea, siempre teniendo como fin último la mejora en la calidad de la enseñanza y del aprendizaje.

Cuando un sistema ha tenido en cuenta todos los aspectos anteriores y permite trabajar de forma colaborativa, o lo que es lo mismo el aprendizaje en grupo, el sistema aportará las siguientes ventajas [14] que han sido observadas durante los dos años de experimentación con alumnos:

- Los estudiantes aprenden de los demás.
- Cuando hay desacuerdos, las relaciones sociales llevan a encontrar la solución.
- Las interacciones verbales (y escritas) para resolver los puntos de vista conflictivos conducen al aprendizaje.
- Los estudiantes no pueden estar pasivos, como mínimo realizan acciones de visualización y posicionamiento dentro de las ventanas de trabajo.
- Al sentir la presencia de los demás se favorece y estimula el aprendizaje.

4. Descripción del Sistema CSCW implementado

Para poder estudiar toda la problemática inherente a los sistemas CSCW utilizados en la enseñanza hay que experimentar con estos en un entorno real. PENCACOLAS fue diseñado para estudiar estos problemas y posibilitar el aprendizaje de la composición. A continuación se describen las fases implementadas en PENCACOLAS así como el hardware y software subyacente en el sistema.

4.1 Fases implementadas

PENCACOLAS incorpora todas las fases típicamente presentes en el proceso de generación de una composición. Dichas fases se describen a continuación brevemente y en el orden en que se realizan:

1. *Brainstorm Individual*: Esta es una fase corta en el tiempo, de pocos minutos, y en ella los alumnos escriben las ideas que les van surgiendo relacionadas con el tema propuesto por el profesor.
2. *Brainstorm Cooperativo*: es una fase que dura unos 30 minutos en la cual los alumnos negocian las ideas comunes que se han de seguir cuando se escriba el texto individual. En esta fase se trabaja en una misma ventana compartida en la cual todos pueden escribir de forma simultánea. Es esta característica la que hace que aparezcan algunos de los problemas típicos del trabajo cooperativo.
3. *Esquema Individual y Cooperativo:* En estas dos fases se debe generar un esquema común que será el "esqueleto" a seguir en el texto final. La

duración de esta fase suele ser de unos 45 minutos.

4. *Composición*: En esta fase, cada alumno genera un documento individual en el cual debe reflejar las ideas generadas-consensuadas por todos y debe seguir el orden del esquema común desarrollado entre todos.

5. *Revisión:* En esta última fase, cada alumno puede revisar los documentos de los demás. Además, el usuario tiene la posibilidad de solapar las revisiones hechas en sesiones anteriores por los demás alumnos a su texto (si estas revisiones se han realizado). De esta forma puede ir secuencialmente visualizando las distintas revisiones y decidir que puntos tendrá en cuenta a la hora de re-escribir el texto.

4.2 Elementos que forman PENCACOLAS

Como se puede observar en la Figura 1, PENCACOLAS actualmente está formado por los siguientes elementos:

- *Ordenadores*: una Sun Sparcstation 10 y tres ordenadores personales , a los que denominaremos PC-1, PC-2 y PC-3, donde, PC-1 tiene un procesador Pentium, PC-2 un procesador 80486 y PC-3 es notebook NEC Versa-E con procesador 80486.
- *Sistemas Operativos*: Solaris 2.4 para la Sparcstation, Windows for Pen Computing instalado sobre Windows 3.11 para el PC-2 y el PC-3, y Windows 95 para el PC-1.
- *Plataformas para el interface de Pen*: una tableta digitalizadora transparente Wacom PL-100V con un lápiz sin cable para el PC-2 (dispositivo de entrada/salida), una tableta digitalizadora transparente Wacom PL-300V para el PC-1 con un pen sin cable (dispositivo de entrada/salida), y un display Touchpen con un pen con cable conectado al puerto serie del PC-3 (entrada/salida).

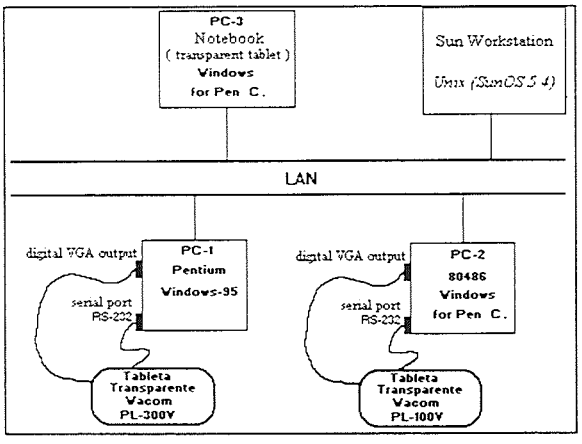


Figura 1: Elementos de PENCACOLAS

Como plataforma de comunicación se usa la Red Ethernet con TCP/IP, usando Berkeley sockets. La visualización se hace usando x-windows con distintos servidores X para cada ordenador.

El sistema permite la interacción tanto del ratón como del pen, sin embargo, está pensado y diseñado para potenciar el beneficio que aporta el usar un pen para escribir texto de forma natural.

Con respecto a las plataformas de software usadas, no se han observado problemas al utilizar distintos entornos.

La configuración presentada en la Figura 1 ha variado desde su concepción inicial en la cual la tableta Wacom PL-300V ha sustituido a una tableta opaca Calcomp 23120 que tenía un pen con cable. Este cambio ha propiciado que el sistema sea más cómodo desde el punto de vista de comunicación hombre-máquina ya que los usuarios no están obligados a observar la pantalla mientras escriben encima de la tableta. Además el lápiz que incorpora la tableta digitalizadora transparente Wacom PL-300V no tiene cable con lo cual se aporta más libertad de movimientos y los usuarios escriben de forma natural.

Con respecto al software, comentar que es lo suficientemente flexible para que el sistema pueda adaptarse fácilmente y resolver los problemas surgidos de las observaciones de los usuarios en condiciones "reales", de tal manera que se han ido generando distintos prototipos que se adaptaban a las necesidades de estos. Este elemento es muy importante en entornos de comunicación hombre-hombre (CSCW) ya que muchas dinámicas del grupo no se conocen a priori, especialmente en el uso de la nueva tecnología.

A continuación se describen los distintos problemas que ha habido que abordar para que el sistema cumpliera de forma eficaz la tarea para la que fue diseñada. Con el diseño del sistema CSCW se ha pretendido, por un lado estudiar los problemas propios de los entornos cooperativos, y por otro, estudiar el proceso educativo y de aprendizaje.

5 Trabajando con PENCACOLAS en el aula real: problemas y soluciones

Durante los dos años de utilización del sistema en la asignatura "*Técnicas de Escritura*" han ido surgiendo distintos problemas tanto del tipo funcional, como informático o telemático. Como se comentó anteriormente, esto ha derivado en la continua actualización del prototipo de tal forma que se adaptara a los requisitos necesarios.

Algunos problemas han sido funcionales, propios de la colaboración como son:

- la organización del trabajo entre todos los alumnos.
- la aparición de roles asumidos por los distintos alumnos: escribano , moderador etc..
- la forma de organizar el debate y como llegar a negociar las ideas interesantes, etc...

Otros problemas han sido problemas "informáticos", la mayoría surgidos en las fases cooperativas, ya que en estas fases se trabaja en una misma ventana compartida en la cual todos pueden escribir de forma simultánea y el sistema tiene que implementar mecanismos de control de concurrencia. Y otros problemas han sido "telemáticos", como la carga excesiva de la red por el tamaño de los ficheros generados y su consecuente relentizamiento. Veamos a continuación las soluciones que se han ido dando a los problemas más interesantes, mencionados en la sección 3.

5.1 Consciencia del espacio de trabajo compartido (P1)

Debido al relativamente pequeño tamaño de las tabletas electrónicas, los alumnos pueden ver solamente parte del trabajo que están realizando en las distintas fases, con lo cual surgen dos problemas fundamentales:

1. ¿Cómo puede un alumno saber en qué parte no-visible del espacio de trabajo se encuentra el resto del grupo trabajando en un momento dado?
2. ¿Cómo se puede saber donde está localizado el puntero de los demás usuarios para evitar problemas de concurrencia [15] (como escribir simultáneamente en la misma zona)?
3. Al ser una sistema de aprendizaje de la composición con varias fases y varios alumnos trabajando a la vez, ¿Cómo puede el usuario moverse de forma eficaz entre las distintas ventanas que le aparecen? y ¿Cómo se organizan las ventanas para que estas no molesten durante el aprendizaje?

Para evitar el primer problema, PENCACOLAS presenta siempre unos marcadores en la parte superior de las ventanas de trabajo en los cuales muestra (con un marcador por persona) en qué página se encuentra trabajando cada usuario. De esta manera se resuelve simultáneamente el problema de gestión de espacio y se contribuye a incrementar la conciencia de trabajo en grupo.

Con respecto al segundo problema, el sistema tiene implementado un sistema de múltiples punteros en las fases cooperativas, de tal forma que mostrando punteros de distinta forma a la del usuario, siempre se tiene conciencia de donde están situados los demás.

Por último, en cuanto al problema de las múltiples ventanas, los usuarios tienen la posibilidad de iconizar aquellas ventanas que no están utilizando y ver a tamaño reducido las ventanas, como se muestra en la Figura 2, de los demás usuarios (a las que sólo pueden acceder para visualizar el contenido) con lo cual pueden gestionar en todo momento qué ventanas quieren tener visibles y cuáles no.

Con respecto a este modo de trabajar, los distintos alumnos que manejaron el sistema indicaron que aunque era la primera vez que manejaban un sistema de ventanas (muchos nunca habían manejado un ordenador), el interfaz les pareció cómodo y no tenían problemas para moverse entre ellas.

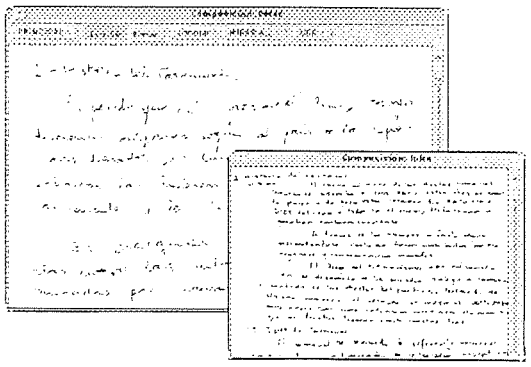


Figura 2: Ventana propia y de otro alumno

5.2 Implementación de roles (P2) y organización del trabajo en las fases cooperativas

Cuando los alumnos trabajan en las ventanas compartidas hay que determinar cuál es la forma más eficaz de trabajar colaborativamente. Así, durante las sesiones realizadas surgieron dos problemas distintos que los alumnos debieron resolver. El primero se refiere a la adopción o no de roles por parte de los usuarios, es decir:

- ¿Es necesario que alguien tome el papel de moderador?
- En caso de que alguien sea el moderador, ¿quién debe serlo, el profesor o un alumno?

El segundo se refiere a cómo organizar el debate de ideas en sí. Algunas soluciones posibles

para la organización de un debate de ideas de estas características son:

1. Utilizar la enumeración de las ideas, es decir, cada alumno pone en la ventana cooperativa la idea que le interesa que figura de los demás. Por ejemplo Idea 1 de Cesar, Idea 2 de Oscar, etc...

2. Un alumno va escribiendo idea a idea cada una de las que han surgido en las fases individuales y los demás califican con OK o NO OK si desean que aparezcan o no. Aquí aparece el rol de escribano.

3. Se va discutiendo las ideas por usuario, es decir, primero sólo se discuten las ideas de uno, después las del otro etc...

4. Se escribe a la vez, separando con una raya tres zonas de la ventana de trabajo, de tal manera que cada usuario sólo utilice esa zona para exponer las ideas que más le interesa y después iniciar el debate.

En cuanto a la cuestión de quién modera la discusión (si es que alguien la debe moderar), la utilización del envío de mensajes y las anotaciones manuscritas fueron los mecanismos más usados durante las distintas pruebas. Así el profesor era el que adoptaba siempre el papel de moderador evitando de esta manera que los alumnos se perdieran o despistaran en la tarea concreta que estaban realizando.

Respecto al modo de organizar las ideas, la solución más adoptada durante los dos cursos académicos fue la escritura en paralelo (sin una división visible de la ventana de trabajo), de tal manera que cada uno escribía la idea que más atractiva le había parecido, y después debajo de esa idea los demás alumnos daban su opinión o ponía otras ideas derivadas de la anterior (o nuevas) como se puede observar en la Figura 3.

Los alumnos opinaban que, de hecho, lo que más les gustaba del sistema era que les ayudaba a concentrarse ya que les obligaba a leer las ideas de los demás. Además, varios alumnos comentaron que al poder trabajar a la vez en la misma ventana se abrían las posibilidades de colaboración real ya que trabajando colaborativamente de forma tradicional (con lápiz y papel) realmente siempre hay alguien que toma el rol de escribano y los demás se dejan influenciar por éste.

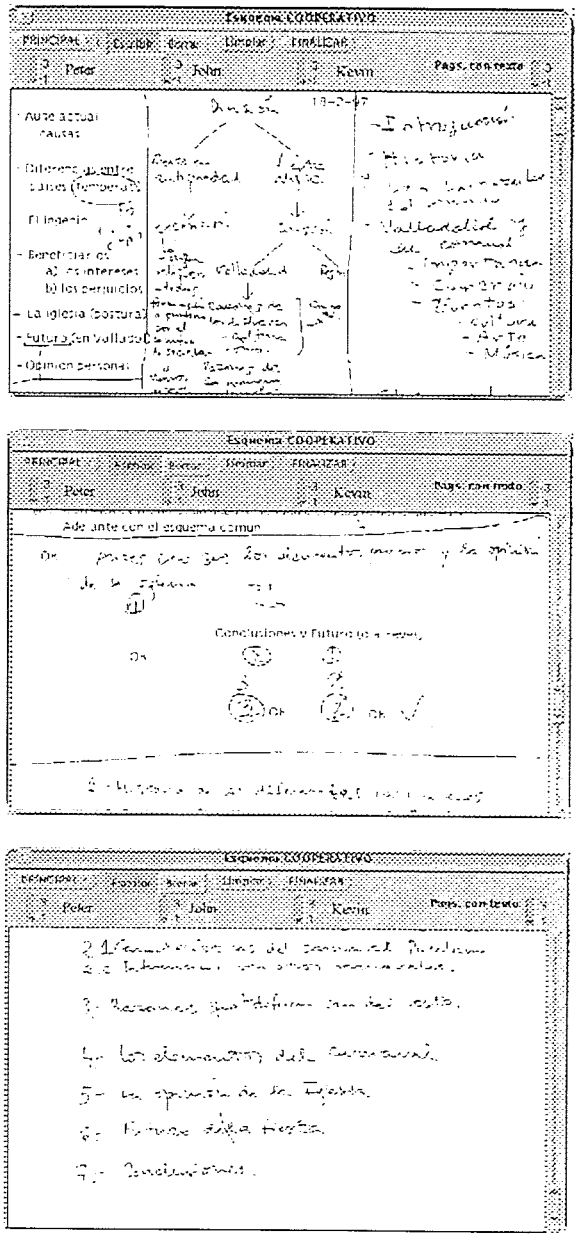


Figura 3: Secuencia de las tres zonas de trabajo de un esquema cooperativo

5.3 Soporte adecuado para la comunicación (P3)

La comunicación entre los usuarios del sistema se puede realizar por los siguientes medios:

- *Mediante el envío de mensajes*: el alumno o profesor puede desplegar un menú en el cual selecciona a quién quiere mandar el mensaje (alguien en concreto o todos a la vez) y escribe el mensaje. Durante toda la sesión el sistema almacena en un fichero la secuencia de mensajes enviados por los usuarios que contiene los siguientes campos:

Fase	Tipo	Emisor	Mensaje	Destinatario

El tener en un fichero almacenada toda la conversación de cada sesión hace que sea posible un

análisis posterior por parte del profesor. De este análisis pueden sacar mucha información útil como por ejemplo:

1. Qué alumno/s son los que necesitan más ayuda (esto se puede observar leyendo el campo "Emisor" y el texto enviado), por lo que denota que este alumno no tiene las ideas claras, no entiende bien el proceso.
2. Cuáles son las fases en las que se producen más mensajes (viendo el campo "Fase"). Esto puede suponer que los alumnos pasen más tiempo intentando entender o pidiendo ayuda que escribiendo, con lo cual puede presuponerse un problema de "no entender" el objetivo concreto de la fase.
3. Los roles que van tomando los alumnos. Dependiendo del texto del mensaje puede observarse si el alumno manda una opinión, una sugerencia o una orden.., etc..

- Otra posibilidad es el empleo de la voz y la utilización de anotaciones manuscritas. El usar la voz en un aula cerrada sería molesto, sin embargo al utilizar anotaciones en las ventanas compatidas los alumnos no interrumpen sus trabajos (algo inevitable cuando se usa el lápiz y el papel)

5.4 Almacenamiento de la información generada por la interface basado en pen

Como se ha comentado anteriormente, el sistema va almacenando en ficheros de eventos toda la información generada en el proceso de aprendizaje. Los ficheros generados son ficheros de eventos gráficos, cuyo contenido son registros con los siguientes campos:

Tipo de Evento (1 byte): Indica el tipo de evento que se ha generado.

Argumento (1 byte): Determinados eventos necesitan de algún tipo de argumento. Es opcional.

ID (1 byte): Identificador del usuario que generó el evento.

Posición X (4 bytes): Coordenada X

Posición Y (4 bytes): Coordenada Y

Los tipos de eventos gráficos que se almacenan en los ficheros y que son generados por el lápiz (que en este sentido funciona igual que el ratón) son los siguientes:

- *Dibujar*. Se produce cuando un alumno desplaza el lápiz sobre la tableta y está activada la opción de escritura, su efecto es el de dibujar líneas en negro. No necesita argumento.
- *Borrar*. Es lo mismo que el anterior pero cuando el sistema tiene activado el botón de "borrar", su efecto es que borra algo ya escrito. No necesita argumento.
- *Carácter*: Para escribir un carácter El argumento es el código ASCII del carácter a teclear.
- *Borrar Carácter*: Para suprimir el carácter anterior. El argumento es el código ASCII del carácter a borrar.

Observando por tanto la información contenida en los ficheros, el profesor puede saber de forma secuencial qué acciones fueron realizadas por los distintos usuarios en las fases implementadas. Además, en las fases cooperativas, pueden analizar, gracias al campo ID de los registros del fichero, quién es el que generó cada evento en concreto. Esto es muy importante para analizar la calidad del proceso de aprendizaje

5.5 Almacenamiento y Recuperación del trabajo de sesiones anteriores

Otro problema que surgió en los prototipos iniciales de PENCACOLAS fue el problema típico que aparece cuando hay que generar muchos ficheros de datos, a saber, la posibilidad de duplicidad de nombres de ficheros. Se tuvo que implementar un mecanismo eficaz que pusiera nombres distintos a los múltiples ficheros de datos que se van generando de forma automática. Mediante este mecanismo PENCACOLAS crea para cada fase individual ficheros de eventos con nombres diferentes y únicos del tipo: *oscar_12-6-97_1.b*, *oscar_12-6-97_1.e* en el que se indica el usuario, la fecha, la fase y la sesión; en las fases cooperativas nombres del tipo *coop_12-6-97_1.b*; y en las fase de revisión *oscar_a_luis_12-6-97_1.r* donde se indica primero el revisor y después al que se revisa

Si un usuario se encuentra en una fase y decide cargar un fichero ya existente de una sesión o día anterior, el sistema automáticamente presenta el listado de aquellos ficheros que se corresponden sólo a esa fase concreta y el usuario utilizando el botón del lápiz electrónico selecciona el fichero deseado sin tener que teclear el nombre, tarea que puede provocar errores. Como se puede observar, estos son los problemas típicos que aparecen, por ejemplo, en la gestión de bases de datos.

El utilizar esta forma de seleccionar los ficheros a cargar sin tener que teclear los nombres

surgió como demanda explícita de los usuarios y presenta las siguientes ventajas:

1. Los usuarios no deben "memorizar" el nombre del fichero que tienen que teclear, evitando posibles errores.
2. El sistema sólo presenta los ficheros que pueden cargarse en una fase concreta, es decir, sólo presenta ficheros con extensión ".b" en la fase de brainstorm, ".c" en revisión etc... evitando la acumulación de nombres de ficheros innecesarios a cargar.
3. Posibilita la agrupación de ficheros por sesiones, fases, usuarios, días etc... para un análisis posterior por parte del profesor.
4. El sistema no tiene que comprobar que el fichero cuyo nombre ha sido introducido por el usuario existe, siempre existen.

Si un alumno no quiere empezar desde cero una fase y carga un fichero de otra sesión o día, automáticamente se genera una copia con los datos de ese fichero y todo lo nuevo que se genere durante la sesión en curso. Con lo cual nunca se destruye la información del fichero original y se puede tener una secuencia histórica del trabajo realizado.

5.6 Tamaño de cada ventana de trabajo

Como resultado de la experimentación, los alumnos demandaron más espacio para trabajar dentro de la fase de Composición ya que el texto que generaban llenaba todo el espacio designado para esa tarea antes de que hubieran acabado. Los motivos por los cuales surgió este problema fueron dos:

- La naturaleza propia de la fase de *Composición*: para escribir un texto siempre se necesita más espacio que para enumerar unas ideas o plantear un esquema, y,
- La tendencia que tiene la gente a escribir más grande de lo normal cuando utilizan una tableta digitalizadora con un lápiz electrónico.

Una vez adaptado el sistema a esta necesidad se observó que los ficheros generados eran de menor tamaño que cuando había menos espacio en la ventana de trabajo. La causa de este fenómeno es que cuando los usuarios tienen poco espacio para escribir, se ven obligados a borrar y rehacer constantemente el texto, con lo cual hay párrafos que son borrados y sobrescritos varias veces. Como para el sistema el evento "borrar" es equivalente al evento "dibujar" (en cuanto a espacio de almacenamiento), y la no visualización de lo escrito anteriormente suele implicar una mayor tendencia a hacer más borradores implicaba el almacenamiento de mayor número de eventos.

Por lo tanto al resolver éste problema se generan ficheros de menor tamaño y como consecuencia de esto el sistema se relentiza menos cuando realiza operaciones de carga o grabado de ficheros de eventos (tareas que se realizan de forma automática y casi constante en una sesión).

6 El empleo de la voz. Posibilidad de localizaciones remotas

Hay que resaltar como dato curioso la escasa utilización de la voz cuando los usuarios están trabajando en una sesión con el sistema. Es sorprendente, y así se lo hemos notado en las diferentes pruebas que los alumnos realizaban, que aunque estaban situados en la misma mesa, separados por centímetros los alumnos y el profesor, no empleaban la voz para hacer comentarios, sugerencias o simplemente preguntas.

Preguntados después de las sesiones por el motivo de este comportamiento, explicaban que como sabían que los demás usuarios podían observar lo que estaban escribiendo preferían escribir con el lápiz la frase que normalmente utilizarían con la voz para que los demás lo leyeran (algo parecido a la utilidad "talk"). Esta forma de comunicarse les hacía perder menos la concentración en lo que estaban haciendo.

En este sentido, algunos autores [14] resaltan que la interacción no-verbal puede ser sustituida perfectamente por medio del lenguaje de signos escritos. Además el utilizar anotaciones "habladas" se diferencia de las "escritas" en que estas últimas suelen llevar implícitos mensajes más directos (incluso hostiles) [11], y los alumnos (como así lo comentaron) siempre tenían una referencia visual de los comentarios que les hacían los demás (ya que lo podían leer en cualquier momento), no teniendo que memorizar las instrucciones dadas por otros.

Es muy importante para nosotros el haber comprobado que la necesidad de hablar no era un requisito fundamental para el buen funcionamiento del sistema por lo cual perfectamente podría utilizarse para realizar la misma tarea pero en localizaciones remotas. Además, cuando los usuarios hablan va almacenando todo el proceso en ficheros y el profesor puede analizar las conversaciones que se han tenido durante las distintas fases sin más que ir recuperando dichos ficheros.

7. Utilización del sistema en localizaciones remotas, Web y Java en PENCACOLAS

Una de las ventajas del uso de ordenadores con interface basada en lápiz electrónico es la movilidad de las llamadas "aulas virtuales" o "clases electrónicas", como la utilizada en la Universidad de Duke, con ordenadores portátiles conectados por medio de infrarrojos [16], y en la Universidad de California, en Berkeley, con el proyecto *Infopad*, proyecto de investigación relacionado con el desarrollo de hardware, software y soporte para redes móviles que permitirán la ubicuidad, acceso sin cable de datos multimedia en tiempo real con redes de alta velocidad usando un terminal portátil barato [17].

Uno de los mayores obstáculos en el uso de redes móviles sin cable en las aplicaciones del mundo real es su bajo ancho de banda, comparado con las redes fijas. Por ejemplo, el modelo de Apple E-mate proporciona alrededor de 115 Kbps., en vez de un máximo de 10 Mbps. en una típica LAN Ethernet. Con vistas a testear la posibilidad de usar nuestro sistema CSCW dentro de un contexto de ordenadores móviles de una clase electrónica, hemos medido el trafico generado por el sistema. En la fase de composición durante la cual la actividad alcanza la cota más alta, el tráfico fue siempre inferior a 12 Kbps como se puede observar en la Figura 4. Hay que apuntar que estas medidas se refieren a la información útil incluida en los eventos gráficos y de coordinación de nuestro protocolo, incluyendo la información añadida por los protocolos TCP/IP, pero excluyendo el tráfico ocasional producido por la gestión de ventanas. en cualquier caso, estas medidas garantizan el uso de nuestro sistema con hardware de red de bajo-costo. Incluso existe suficiente margen para un canal de voz si la aplicación lo requiriera.

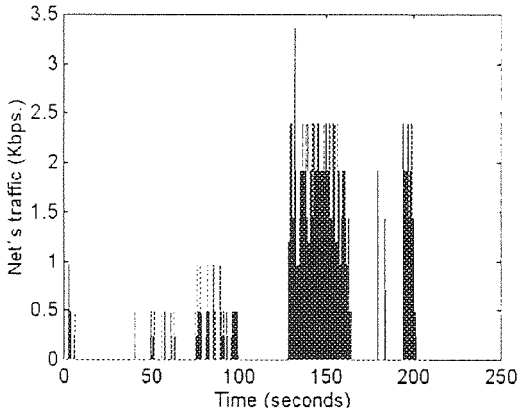


Figura 4: Tráfico de red generado por el sistema

Además de nuestros datos experimentales del uso de pen-computers en un contexto educativo, la posibilidad de usar redes móviles han sido demostrados con herramientas que manejan información multimedia, como son el conjunto de herramientas llamadas Coral [18].

Como consecuencia de la viabilidad de poder usar el sistema en localizaciones remotas, nuestro grupo de investigación está desarrollando actualmente otro prototipo que haga posible la utilización del sistema en aulas remotas utilizando la plataforma Web y el lenguaje JAVA, así como tecnología de agentes.

La principal ventaja del uso de Web es que su fácil, universal e intuitivo acceso a redes TCP/IP (Internet e Intranets) ha impulsado una explosión en su uso en distintas aplicaciones. La educación puede beneficiarse, por ser una plataforma de comunicación y búsqueda de información [19].

Por otro lado, se están estudiando las posibilidades de integrar Web en entornos de colaboración síncrona y asíncrona [20][14], aunque todavía restan muchos avances para conseguir este objetivo. Nuestro objetivo es poder evaluar el impacto del uso de Web, comprobando los problemas y ventajas en un sistema con la funcionalidad de PENCACOLAS, y compararle con el prototipo basado en software a medida. En este sentido, hacemos especial hincapié en la posibilidad de integrar trabajo síncrono y asíncrono, dentro de una consideración global de la asignatura de "Técnicas de Escritura". En ella un alumno crea una carpeta a lo largo de todo el cuatrimestre, como medio de concienciar el carácter recursivo de la escritura, de comunicación con el resto de compañeros y profesores y de evaluación global.

Además, se está implementando una tecnología de agentes como soporte de la arquitectura distribuida de nuestro sistema. Nuestra aproximación para integrar PENCACOLAS y la plataforma multiagente MAST [21] [22] ha sido la aplicación de una metodología llamada MAS-CommonKADS [23]. En concreto, se aplica esta metodología para todas las fases de desarrollo de la aplicación multiagente. Así, se consiguen ventajas propias de ingeniería de software, tales como la orientación a objetos, modularización, facilidad de programación con el uso de un lenguaje declarativo de agentes, así como se facilita la incorporación de una base de conocimientos y la comunicación entre agentes. La aplicación final en Java se integra cómodamente al entorno Web.

Otro trabajo que se está realizando es el análisis de los gestos manuscritos más utilizados para poder introducir en el sistema un mecanismo de reconocimiento de gestos que haga que ciertas tareas propias de la edición y revisión de textos puedan automatizarse al traducir el propio sistema

el gesto realizado con la acción deseada por el usuario.

8. Conclusiones

En este artículo se ha presentado un sistema CSCW basado en una interfaz de lápiz electrónico para la enseñanza y aprendizaje de la composición de textos. El sistema permite la interacción entre alumnos (grupos de 3) y la supervisión e interacción del profesor. Con el sistema, llamado PENCACOLAS (PEN Computer Aided Composing cOLlAborative System), se han logrado los siguiente objetivos:

- Posibilitar el paso del alumno por las distintas fases que subyacen en el proceso de composición de un documento (brainstorm, planificación, edición y revisión).
- Dotar, tanto al profesor como a los alumnos, de una interfaz que les permita visualizar el trabajo de los demás e intervenir en ciertas circunstancias.
- Facilitar la interacción con el ordenador de forma que sea lo más natural posible y aprovechar la posibilidad de trabajar en un entorno de red.
- Facilitar la comunicación entre los participantes, mediante gestos y anotaciones, para que cooperen en la generación del documento.

El desarrollo e implementación del sistema ha sido posible gracias a la colaboración estrecha de un grupo de investigadores especializados en áreas de conocimiento tan diferentes como son la educación (profesores de la Facultad de Educación), las telecomunicaciones (profesores de la ETSI de Telecomunicación) e Informáticos, grupo que lleva colaborando en este proyecto educativo desde el año 1995.

Además, con PENCACOLAS se ha logrado por una lado introducir en el aula real un sistema CSCW con una interface novedosa para los alumnos y a la vez realizar un estudio profundo de los problemas y características propias del trabajo cooperativo soportado por ordenador, implementando soluciones y testeando estas en un entorno de trabajo real.

Para lograr estas metas, el sistema ha sido utilizado por alumnos de la Facultad de Educación de la Universidad de Valladolid, durante los dos últimos años académicos (95-96 y 96-97), y como resultado de los comentarios, sugerencias y observaciones realizados por ellos y por parte del profesorado el sistema ha ido evolucionando hasta llegar a su concepción actual.

Agradecimientos

Los co-autores de este artículo desean agradecer a todos los alumnos de la Facultad de Educación de la Universidad de Valladolid el esfuerzo y la paciencia demostrado al acceder a utilizar nuestro sistema. También queremos agradecerles los comentarios, sugerencias y problemas planteados que han hecho que el sistema haya mejorado. Agradecemos también al Experto en Aplicaciones de Wacom, Sr. Bailey por su apoyo e interés en nuestro sistema. Apoyo, e interés demostrado con la donación de materia y su visita a nuestra Escuela.

Referencias

[1] Martín, P., "Aplicaciones de la telemática en escenarios educativos", *Technical Report, Gabinete para la Aplicación de las Tecnologías a la Educación (GATE)*, Universidad Politécnica de Madrid, Madrid, (1995).

[2] Grudin J., "Computer-Supported Cooperative Work: History and Focus", *IEEE Computer*, **27**, 5, 19-26, (1996).

[3] Gutwin, C., Stark, G., and Greenberg, S. "Support for Group Awareness in Educational Groupware." *Proceedings of the 1995 Conference on CSCL'95*, Indiana University, Bloomington, Indiana, disponible en http://www-cscl95.indiana.edu/cscl95/, (1994).

[4] Alonso, C. "El ordenador y el tratamiento de la información", *Cuadernos de Pedagogía*, no. 230, 14-18, (1994).

[5] Kimura, T. D., "Pen-Based User Interface", *Proceedings of 1992 IEEE Workshop on Visual Languages*, 168-173, (1992).

[6] Kimura, T. D., "A Pen-Based Prosodic User Interface for Schoolchildren", *IEEE Multimedia*, **4**, 3, 48-55, (1996).

[7] Ellis, C.A., et al., "Groupware: Some issues and experiences", *Communications of the ACM*, **34**, 1, 38-58, (1991).

[8] Michels, S., "The cooperator: Co-writing, look and feel!", *Master thesis, INFOLAB*, Tilburg, , disponible en http://infolabwww.kub.nl:2080/w3thesis/, (1995).

[9] Koch, M., "Design issues and model for a distributed multiuser editor", *Computer Supported Cooperative Work*, 3: 359-378, (1995).

[10] Mitchell, A., Posner, I., and Baecker, R. "Learning to write together Using Groupware", *Proceedings of the Computer Human Interaction*, CHI'95, disponible en http://www.acm.org/sigchi/chi95/Electronic/documents/papers/, (1995).

[11] Chen Ch., Rada R., "Collaborative Authoring Dynamics", In Groupware and Authoring, Editado por Roy Rada, Academic Press, 45-65, (1996).

[12] Davis, J. R., and Huttenlocher, D. P., "Shared Annotation for Cooperative Learning", *Proceedings of the 1995 Conference on CSCL'95*, disponible en http://www-cscl95.indiana.edu/cscl95/1, (1995).

[13] Moran, T.P., et al. "Implicit Structures for Pen-Based Systems Within a Freeform Interaction Paradigm", In *Proceedings of the Computer Human Interaction, CHI'95*, (1995).

[14] Hiltz, S. R., and Turoff M., "Asynchronous Learning. Networks: The Theory and Practice of Collaborative Learning Online, *Tutorial Notes of CSCW'96*, (1996).

[15] Greenberg, S. and Marwood D., "Real Time groupware as a distributed system: Concurrency control and its effects on the interface", *Research Report 94 534 03*, Department of Computer Science University of Calgary, (1994).

[16] Oakley II, B. "Computers and Networks in Engineering Education. One Educator's perspective of where are today and where are going to be tomorrow" *IEEE Education Society Newsletter*, 1-6, (1994).

[17] Narayanaswamy, S., et al. "Application and Network Support for Infopad", *IEEE Personal Communicactions*, **3**, 2, (1996).

[18] Minneman, S., et al., "A Confederation of Tools for Capturing and Accesing Collaborative Activity", *Proceedings of the Third ACM Multimedia Conference and Exibition, Multimedia'95*, 523-534, (1995).

[19] Graves W. H., "Why higher education needs an advanced Internet", *IEEE Computer*, **29**, 11, November 1996, 93-98, (1996).

[20] Bentley R., et al. "Supporting Colalborative Information Sharing with the WWW: The BSCW Shared Workspace System", *Proceedings of the Fourth International World Wide Web Conference*, (1995).

[21] Velasco J.R., et al., "Multiagent-based Control Systems: A Hybrid Approach to Distributed Process Control", *Control Eng. Practice*, **4**, 6, 839-845, (1996).

[22] Iglesias C. et al., "MIX: A general purpose multiagent architecture.", In M.Wooldridge, K.Fischer, P.Gmytrasiewicz, N.R. Jennings, J.P. Muller, and M.Tambe, editors,"*INTELLIGENT AGENTS II: Agent Theories Architectures, and Languages*", volume 1037 of Lecture Notes in Artificial Intelligence, Springer Verlag, 251-266, (1996).

[23] Iglesias C., et al., "A methodological proposal for multiagent systems development extending commonkads", en B.Gaines and M.Musen, editors, *Proceedings of the 10th Banff Knowledge Acquisition for Knowledge-Based Systems Workshop*, **1**, 1-17, (1996).

Trabajo Cooperativo en el Sector de la Construcción

Juan Pérez Sainz de Rozas
Dpto. de Tecnologías de la Información
LABEIN
Parque Tecnológico, 101. 48170 - ZAMUDIO
Correo Electrónico: juan@labein.es

Abstract:

Along the phases of a building design and construction process there are a lot of working groups, like the client, the management team, the architecture team, ..., involved in different tasks, each one with different level of activity and responsibility. Nevertheless, the building process is not always a sequential activity, but usually the initial project is modified during its development. This changes imply asynchronous interaction among different working teams that could be place at remote locations. The Information Technologies provide the required mechanisms to facilitate this type of interaction. In general, this kind of mechanisms are called CSCW (Computer Supported Cooperative Work).

1 Introducción

Durante las fases de diseño y construcción de edificios existen diferentes equipos, como el cliente, la dirección de obra, el estudio de arquitectura, etc, que toman parte en el proceso con diversos grados de intervención y de responsabilidad.

Ahora bien, el proceso de construcción de un edificio no es necesariamente un proceso secuencial, sino que el proyecto inicialmente realizado sufre modificaciones hasta que se concluye la construcción del mismo. Estas modificaciones pueden surgir como consecuencia de problemas encontrados por los diferentes equipos de trabajo durante la construcción o como mejoras sugeridas por alguno de ellos. Dichas propuestas de modificación deben ser consensuadas entre las partes implicadas, con el agravante de que generalmente estos equipos de trabajo se encuentran en lugares distantes geográficamente.

Actualmente las Tecnologías de la Información ofrecen mecanismos como las herramientas de trabajo cooperativo o groupware orientadas a resolver los problemas derivados de la interacción entre equipos de trabajo que no coinciden en el espacio (situados en lugares distantes) o en el tiempo (disponibilidad temporal incompatible).

En esta presentación se comentará la repercusión de éstas herramientas en el sector de la construcción, el cual a pesar de jugar un papel fundamental en la economía del país se encuentra muy atomizado y en niveles bajos de automatización.

2 ¿Qué es el Trabajo Cooperativo?

Dada la novedad del concepto resulta imposible encontrar una definición universal, habiendo infinidad de ellas, pero detrás de todas ellas subyace la idea de un conjunto de herramientas que facilitan la colaboración de varias personas para desarrollar una tarea o alcanzar un objetivo común. Por lo tanto, la gama de herramientas o aplicaciones es amplísima, desde los sistemas de directorios en red (NFS - Net File Systems) hasta las aplicaciones más avanzadas de video-conferencia o workflow.

Como reflejo de esta confusión, el mismo término Trabajo Cooperativo, o Groupware, se entremezcla con el de CSCW (Computer Supported Cooperative Work). En general, se tiende a utilizar CSCW para referirse a la ciencia encargada de analizar el comportamiento de los grupos de trabajo y Trabajo Cooperativo para referirse a las herramientas software y hardware que soportan las técnicas de CSCW. En este artículo se utilizará el término Trabajo Cooperativo indistintamente.

Realmente lo que definiría si una aplicación forma parte del flujo de Trabajo Cooperativo o no es el uso que se haga de la misma. Por ejemplo, si el correo electrónico se utiliza sólo para envío de mensajes de forma indiscriminada en toda la empresa no se debería considerar parte de un flujo de Trabajo Cooperativo, pero si por el contrario se definen alias, listas de distribución, etc. si que podría ser considerado como tal.

En el Trabajo Cooperativo hay tres componentes básicos: la comunicación, la colaboración y la coordinación.

- ***Comunicación***, entendiendo por tal el simple *intercambio* de información.

- ***Colaboración***, entendiéndose por ello el compartir la información, como puede ser un documento de referencia o una base de datos. Es decir, implica compartir un *espacio* físico o virtual.

- ***Coordinación***. Es un concepto similar al de colaboración, pero en realidad es mucho más que eso ya que implica definir la colaboración en el *tiempo*, como por ejemplo al definir un proceso en una determinada empresa.

El Trabajo Cooperativo consiste en la fusión de estos tres conceptos. Su finalidad no es otra que utilizar adecuadamente la capacidad de cada uno de ellos, al objeto de generar abundantes beneficios no sólo para la empresa sino también para sus clientes.

Ahora bien, ¿A que se debe el auge actual de esta tecnología? Hace 13 años, en 1984, Paul Cashman e Irene Grief organizaron un workshop de carácter multidisciplinar entre personas cuyo interés común era la actividad de los grupos de trabajo. El objetivo del workshop era analizar como la tecnología disponible en aquel momento podía soportar las actividades relacionadas con el trabajo en grupo. De este workshop surge el término Computer Supported Cooperative Work o CSCW. A pesar de que la idea surge en los Estados Unidos, su difusión a nivel mundial fue inmediata,.

Pero, ¿por qué surge la idea en 1984? Repasemos brevemente la historia de la informática. A mediados de los años 60 se comenzaron a desarrollar aplicaciones de gran éxito, como las de asignación de plazas en aviones o las de gestión de nóminas. Estas aplicaciones tenían todas en común unas especificaciones de requisitos muy claras. A mediados de los años 70, ante el avance de los miniordenadores se comienza a vislumbrar la posibilidad de ofrecer soporte informático a las tareas más sofisticadas relacionadas con tareas de grupo y corporativas, con lo que nace la idea de la *Oficina Automática*. Si bien se alcanzaron grandes éxitos en aplicaciones orientadas a un usuario único, como los procesadores de texto o las hojas de cálculo, en las aplicaciones de grupo no se alcanzaron éxitos significativos, ya que el gran problema era que se desconocían los requisitos de este tipo de sistemas. A fin de dar respuesta a estas necesidades surge el Trabajo Cooperativo, que representa un punto intermedio entre las grandes aplicaciones corporativas corriendo en mainframes de los años 60 y las aplicaciones monousuario de los años 70-80. Las primeros sistemas surgieron del mundo de la gestión empresarial y estaban orientados a la ayuda en la toma de decisiones de altos ejecutivos, ya que su elevado precio solo se justificaba en este tipo de aplicaciones. Sin embargo, conforme los costes han ido bajando el uso de estas aplicaciones se ha generalizado, ampliándose su aplicación como soporte a grupos de trabajo de muy variada naturaleza.

Actualmente la tendencia hacia la cooperación en el ámbito de las empresas es más fuerte que nunca debido tanto a factores socioeconómicos como técnicos. Por un lado la globalización de la economía que se ha producido a lo largo de los años 90 exige un incremento constante de la competitividad, por lo que se debe prestar especial atención a aspectos como productividad o calidad, lo cual ha derivado en procesos de reingeniería, tanto a nivel de actividades como de estructura de la empresa. Consecuentemente cada vez las empresas están menos jerarquizadas, los empleados tienen mayor autonomía, y en cada momento es necesario disponer del personal más adecuado, el cual no tiene porque estar en la misma oficina. Por otra parte, la técnica ha permitido reducir el coste de la inversión de forma drástica, el crecimiento de las redes de comunicación ha sido, y seguirá siendo, espectacular. Todo ello ha creado el ambiente propicio para poder decir que el Trabajo Cooperativo será la tecnología de esta década.

3 Tipología de las Aplicaciones

A la hora de clasificar una aplicación de Trabajo Cooperativo la tipología más conocida es la de Robert Johansen [1], que se basa en criterios de lugar y tiempo, diferenciando 4 tipos:

- **Mismo Tiempo/ Mismo Lugar.**
  En este tipo de aplicaciones lo fundamental es proporcionar herramientas de asistencia a reuniones cara a cara. Estas herramientas van desde las más sencillas, tipo retroproyectores, hasta las más sofisticadas, en las cuales cada asistente desde un ordenador interactúa con el resto de asistentes, por ejemplo, aportando ideas o contribuyendo a la elaboración de un documento.

- **Mismo Tiempo/ Distinto Lugar.**
  Este tipo de aplicaciones es habitual en las empresas con varias delegaciones repartidas en un área geográfica limitada, en las cuales se desea poder celebrar reuniones entre personas repartidas por todo el mundo. Un concepto clave en este tipo de aplicaciones es WYSIWIS (What You See Is What I See). Esto permite definir un espacio de trabajo compartido por todos los asistentes a la reunión. Los sistemas de videoconferencia permiten oír y ver lo que sucede en diferentes lugares. Un complemento ideal a estos sistemas son las aplicaciones de pizarra compartida, que permiten definir una pizarra en un ordenador y que esta sea visualizada desde varios puestos, y ejecución compartida, que permite a varias personas compartir la ejecución de una aplicación, tanto desde el punto de vista de visualización como de control de la misma.

- **Distinto Tiempo/ Mismo Lugar.**
  Este tipo de aplicaciones es habitual en las empresas que trabajan a relevos, como las factorías o los hospitales. En este caso las aplicaciones más típicas son las de *memoria*, como las aplicaciones de tipo *tablón de anuncios* o *biblioteca*.

- **Distinto Tiempo/ Distinto Lugar.**
  Este tipo de aplicaciones es habitual en las empresas con varias delegaciones repartidas por todo el mundo. En esta caso son fundamentales aplicaciones de tipo *mensajería* o *workflow*.

Por otra parte, cabría hacer una clasificación en base al tamaño del grupo al que se orienta la aplicación, diferenciando entre Trabajo Cooperativo para grupos de trabajo pequeños frente a aplicaciones corporativas.

Conforme los PC´s y las workstations se han ido integrando en redes, LAN o WAN, los pequeños grupos de trabajo han ido cobrando relevancia como mercado, por lo que aplicaciones monousuario plenamente consolidadas se han visto potenciadas por funcionalidades de Trabajo Cooperativo orientadas a dar soporte en las comunicaciones y la coordinación de sus actuaciones. Además las empresas de telecomunicaciones están haciendo un gran esfuerzo en potenciar las redes de comunicación de Banda Ancha. En este tipo de aplicaciones los aspectos relacionados con la comunicación son los más potenciados, siendo secundarios los temas de cooperación y coordinación, ya que en general hay un objetivo común muy claro, por lo que todos los miembros del equipo de trabajo tienen una actitud cooperante.

Por otra parte, las aplicaciones corporativas rodando sobre mainframes o miniordenadores llevan décadas rodando, por lo que en el entorno de los IS ya están familiarizados con los aspectos *sociales* de las empresas. Conforme las redes de comunicaciones se han ido desarrollando y los costes de los grandes sistemas se han ido reduciendo, el ámbito de aplicación de los IS se ha ido ampliando, con lo que los sistemas de ayuda a la decisión, inicialmente reservados a la alta dirección, se han ido generalizando, dando lugar a los sistemas corporativos de trabajo en grupo, más baratos y flexibles. Este tipo de sistemas presta especial atención a los temas de cooperación y coordinación, ya que a nivel global de empresa los problemas fundamentales son la coordinación de esfuerzos y la resolución de conflictos resultantes de objetivos contrapuestos. Este tipo de sistemas suelen ser consistir en desarrollos internos y se centran en cubrir necesidades no resueltas por las herramientas comerciales.

4 Herramientas en el Mercado

En una sección anterior se describen los 3 componentes fundamentales desde el punto de vista metodológico del Trabajo Cooperativo, pero hay un cuarto componente que es fundamental para el Trabajo Cooperativo, los entornos de desarrollo, los cuales se componen de utilidades y herramientas que facilitan las tareas de desarrollo de las aplicaciones.

Dado que las aplicaciones de Trabajo Cooperativo son aplicaciones que deben convivir con otras de uso específico ya existentes en la empresa, por ejemplo, aplicaciones de control de gastos o de edición, su interoperabilidad es fundamental. Por otra parte, los procesos en cada empresa presentan unas particularidades que hacen necesario configurar y programar la *toolkit* comercial para que satisfaga las necesidades específicas de cada aplicación.

Para cubrir estas necesidades han surgido entornos de desarrollo integrado que permiten el diseño y la codificación de aplicaciones en una única herramienta o en su defecto son integrables con lenguajes standard (Visual C++, Visual Basic, Delphi, etc).

Dentro de las herramientas aplicadas en Trabajo Cooperativo se diferencian 2 tipos fundamentales: herramientas síncronas y herramientas asíncronas.

Las herramientas síncronas están orientadas a facilitar el trabajo en grupo de varias personas en el mismo instante de tiempo, siendo un componente fundamental de las mismas las funcionalidades de videoconferencia y compartición de ejecución de aplicaciones.

Por el contrario, las herramientas asíncronas están orientadas a facilitar el trabajo en grupo de varias personas en diferentes instantes de tiempo, siendo un componente fundamental de las mismas las funcionalidades de base de datos compartida, mensajería y control del workflow.

Como se puede observar, ambos tipos de herramientas son complementarias, estando determinada la relevancia de cada tipo de herramienta en un proyecto de Trabajo Cooperativo en función de los requisitos de cada proyecto.

4.1 Herramientas Asíncronas

La gama de herramientas existentes en el mercado es muy amplia, siendo las siguientes las más significativas:

- **Lotus Notes.**

  Notes ha dominado el mercado del Trabajo Cooperativo durante años gracias a su amplio soporte de plataformas (soporta prácticamente todas las plataformas del mercado), los potentes formularios o plantillas y la capacidad de replicación de bases de datos de documentos.

  Notes se construye a partir de una base de datos orientada a objetos. Se pueden enlazar objetos entre sí, incrustar documentos, hojas de cálculo y gráficos, y controlar el acceso a ellos mediante "listas de control de acceso" o ACIs. Estas bases de datos se pueden replicar en servidores remotos siempre que se desee, de forma que cuando los usuarios añaden, borran o modifican documentos, el sistema sincroniza todos los cambios.

  Respecto a sus funciones de mensajería, Notes permite ligar documentos a e-mails mediante 3 mecanismos diferentes: la inclusión del documento en el mail, la anexión del documento y la inclusión de una referencia al documento, con lo que se evitan problemas de espacio en disco. Además, los usuarios pueden convertir los e-mail en tareas de workflow integrando la nueva función TASKS con Lotus Organizer.

  En base a LotusScript, lenguaje propio de programación orientado a objetos, se pueden

programar *agentes*. Los agentes se disparan por un determinado evento y dan origen a una serie de acciones. Por ejemplo, el recibir un determinado correo puede hacer que se ejecuten una serie de tareas de forma automática, como hacer un forward del mensaje, archivarlo o replicar la base de datos.

Respecto a su integración con Internet, ya se pueden incluir enlaces a páginas Web en el correo y documentos de Notes. Además, InterNotes Web Publisher (Notes 3.x) convierte las plantillas, los documentos, las vistas y las bases de datos de Notes en Hypertext Markup Language (HTML) para publicar en el Web. Por otra parte, InterNotes Web Navigator (Notes 4.0) permite a un servidor Notes actuar como Browser de la Web, pudiendo guardar páginas Web en la base de datos de Notes y acceder a Internet desde cualquier puesto, con la única condición de que el servidor disponga de TCP/IP.

- **Microsoft Exchange.**

La potencia de ésta herramienta reside en el correo electrónico. Presenta una relativa limitación, ya que el servidor de Exchange sólo corre sobre Windows NT. Sin embargo, los clientes ya corran bajo Macintosh, Win 95 o Win 3.1x. Desde el punto de vista del sector de la construcción esto no es un grave problema, ya que el entorno Windows es el dominante. Además, aunque Exchange Server funciona sólo bajo Windowa NT, se comunica bien con otras redes del tipo TCP/IP, NetBIOS, IPX/SPX, y AppleTalk.

Dado que Exchange se basa en una herramienta de correo electrónico, por lo que se compone de buzones de correo privados y folders públicos (buzones que no pertenecen a usuarios específicos), que son similares en ciertos aspectos a las bases de datos de Notes. Los propietarios de los folders determinan el acceso de los usuarios.

Como correo electrónico Exchange 4.0 es un buen producto, que soporta conexiones a través de Microsoft Mail, X.400 e Internet y es compatible con POP3. Sin embargo, los usuarios de Exchange a la hora de replicar sus mensajes o bien eligen todos los folders o seleccionan los ficheros a replicar de uno en uno.

En lo referente a lenguajes de programación se cuenta con Visual Basic, Visual C++ y controles ActiveX. Sin embargo, no dispone de *agentes*.

Respecto a la integración con Internet, la versión 4.5 de Exchange es compatible con HTTP (Hypertext Transfer Protocol, protocolo empleado para intercambiar documentos por Internet, entre cliente y servidor Web), con POP3 (Post Office Protocol, protocolo de mensajería usado habitualmente en internet). Esto permite a los clientes de correo de internet usar servidores Exchange.

Como conclusión, si se quiere un avanzado sistema de correo y se tiene la plataforma necesaria para soportarlo Exchange es una opción muy válida, pero no se debe olvidar que solamente dispone de una parte de las funciones de Trabajo Cooperativo de Notes. Ahora bien, Microsoft está en el camino de hacer un buen producto y dada la capacidad de inversión y presencia en el mercado de Microsoft es previsible que las distancias se acorten rápidamente.

- **WorkCenter.**

A pesar de que es una herramienta minoritaria en el mundo del trabajo cooperativo, de cara al sector de la construcción se debe tener muy en cuenta, ya que es un desarrollo de Autodesk, que a su vez es el fabricante de AutoCAD, el standard de facto en el mundo de la construcción. Sólo soporta la red local Novell, pero tampoco es un problema importante en el sector de la construcción, ya que Novell es la LAN más frecuente. WorkCenter está totalmente integrado con AutoCad y posee herramientas para la gestión de la documentación y el control del flujo de trabajo.

Tiene un sistema de acceso sencillo a los documentos que evita sobreescrituras, borrados o errores de ubicación de archivos. Cada documento se cataloga en unas tarjetas descriptivas que permiten incluir información del contenido, nombres de archivo, número de documento, numero o código de revisión, fecha de las aprobaciones y revisiones, y cualquier dato considerado como relevante. Permite la compactación de archivos y crear copias de seguridad.

Con las llamadas carpetas inteligentes se pueden emplear una gran variedad de formas de catalogación. Esto es una alternativa a los directorios en árbol de DOS.

Con los datos descriptivos se pueden realizar búsquedas de cualquier documento, o documentos asociados, o guardar los criterios de búsqueda para una utilización posterior. Para visualizar un archivo no es necesario contar con el programa original con el que se crearon los archivos.

El sistema retiene todas las revisiones aprobadas, manteniendo la historia de todos los cambios realizados en el diseño. Además permite comparar revisiones.

La capacidad de control del flujo de trabajo del programa permite a los diseñadores enviar la información exacta a quien corresponda. Los responsables del sistema pueden, además, realizar seguimiento del trabajo.

A parte de estas herramientas comerciales, desde 1995, las Intranets empezaron a cuestionar el software comercial para Trabajo Cooperativo. Tienen como ventajas el ser mas baratas, fáciles de configurar y el hecho de que los navegadores Web necesarios para acceder a ellas operan prácticamente en todas las plataformas. Como respuesta a esto, Lotus y MS dotaron a sus productos con la posibilidad de utilizar las comunicaciones de bajo coste de Internet.

Aunque hoy en día, Notes y Exchange son superiores a Internet/Intranet, los fabricantes de productos Internet están intentando igualar las funcionalidades de los paquetes de Trabajo Cooperativo. Por ejemplo, Netscape preve adjuntar Collabra Share (conferencia + algunas capacidades de Notes) a su navegador Netscape Navigator.

Aunque Notes, actualmente, no tiene competencia, Internet esta siendo considerada como una seria alternativa. La escasez de seguridad ha sido la crítica habitual a la información compartida mediante Internet, pero la aparición de los Secure Socket Layers (SSL), los firewalls y la encriptación ha restado importancia a este problema.

Los Newsgroups de Internet tienen *discussion features* y funcionalidades similares a las de Notes y Exchange. Donde internet diverge es en el manejo de documentos. Newsgroup es una forma adecuada para enviar mensajes y *attachments* para no hay forma de organizarlos a excepción de la línea que explica el *subject*.

Como puntos débiles de los *Web sites* se puede citar el que no se pueden usar para gestión de documentos de grupo. Una vez que el documento está publicado se puede leer o descargar fácilmente, pero es imposible editarlo. No hay ninguna base de datos para almacenar u organizar documentos: sólo se dispone del árbol de directorios del sistema operativo, con documentos dispersos entre directorios y unidos mediante *hypertext links*.

Además, no es posible la replicación entre diferente *sites*. Se puede propagar el mensaje a múltiples *sites* pero la replicación de *Web sites* es mucho más difícil . Se pueden transferir ficheros de un sitio a otro, pero no hay ningún mecanismo para sincronizar los cambios hechos en múltiples *sites*.

Tanto Lotus como Microsoft trabajan para llevar sus aplicaciones de Trabajo Cooperativo hacia Internet, pero no está claro como podrán competir con productos diseñados específicamente para Internet.

4.2 Herramientas Síncronas

Son herramientas orientadas a la celebración de reuniones electrónica. Casi todas incluyen la funcionalidad de video-conferencia. De entre la amplia gama existente, se ha considerado que las más interesantes son las basadas en PC, ya que generalmente ofrecen las siguientes funcionalidades:

- Selección de resolución de 352*288 puntos a 20 $^{imágenes}/_{segundo}$ ó 176*144 puntos a 30 $^{imagenes}/_{segundo}$. Esto permite ajustar en cierta medida la calidad de la imagen a las necesidades del momento. Por ejemplo, si se desea ver un documento es más importante la resolución que la velocidad de refresco.

- Entrada auxiliar de video, lo cual permite conectar una segunda cámaras analógica (la cámara principal suele ser digital y está orientada a recoger la imagen del interlocutor).

- Incorporan señal de audio.

- Conexión a RDSI de 1 a 3 acceso básicos, lo que significa un ancho de banda de 128 Kbits a 384 Kbits. Algunas herramientas también son conectables a una LAN, pero en estos casos las prestaciones dependen totalmente del nivel de carga de la red.

- Kit de desarrollo para integrarlas con otras aplicaciones Windows.

- Transferencia de ficheros entre los interlocutores.

- Compartición de aplicaciones, lo que permite ejecutar una aplicación en uno de los puestos y que los resultados sean visibles tanto en el puesto que ejecuta la aplicación como en el de su interlocutor. Además, ambos puestos pueden controlar la aplicación.

- Los precio se sitúan en torno a las 250.000 ptas.

Como referencias comerciales se pueden citar ARMADA Cruiser100, INTEL ProShare 200, SAGEM Meet-Me PC y Vtel.

En caso de desear más información sobre herramientas, en la dirección WEB *http:/wwwschlichter.informatik.tu-muenchen.de/cscw/yp/YP-index-type.html* hay un índice de todo tipo de herramientas relacionadas con trabajo cooperativo. Si sólo se desea información sobre herramientas de videoconferencia en la página WEB *http://www3.ncsu.edu/dox/video/features.html* existe un sumario de ellas clasificadas por plataformas que las soportan.

5 Problemática del Sector de la Construcción

La construcción de un edificio generalmente consta de tres partes:

- ***Planteamiento Inicial***, que es la fase en la cual el cliente define **cuáles** son sus necesidades.

- *Desarrollo*, que es la fase en la cual un equipo técnico, generalmente liderado por una Ingeniería o un Estudio de Arquitectura, define **cómo** se van a cubrir las necesidades del cliente. En esta fase a parte de las Ingenierías y de los Estudios de Arquitectura también intervienen las Empresas de Servicios Auxiliares, como las instaladoras de calefacción y aire acondicionado, electricidad, etc.

- *Construcción*, que es la fase en la cual se realiza la contratación de la obra y la construcción. En esta fase normalmente interviene una empresa que realiza las labores de control de obra, 1 ó varias empresas constructoras y un gran número de proveedores.

Sin embargo, la construcción de un edificio, tal y como se mencionó anteriormente, no es necesariamente un proceso secuencial, sino que, el proyecto inicialmente realizado puede sufrir una serie de modificaciones hasta que se concluye la construcción del mismo. Estas modificaciones pueden surgir en cualquiera de las fases del proyecto bien como consecuencia de problemas encontrados por los diferentes grupos durante la construcción o bien como mejoras sugeridas por alguno de ellos. Dichas propuestas de modificación deben ser consensuadas entre las partes implicadas antes de modificar el proyecto y ejecutarlo.

Este sector presenta la particularidad de que concentra un porcentaje muy alto de PYMEs, por lo que a lo largo del proyecto son infinidad el número de empresas y proveedores que han intervenido en el mismo. Por otra parte, este fenómeno irá en aumento, ya que la creciente demanda de calidad y complejidad en los procesos hace que cada vez las empresas estén más especializadas.

Por otra parte, la necesidad de buscar nuevos mercados está llevando a las empresas a abrir delegaciones y realizar trabajos en lugares remotos, teniendo que competir con empresas europeas y americanas.

Todo esto implica que se necesita disponer de herramientas que por un lado permitan paralelizar al máximo la actividad de los equipos de trabajo que intervienen en el proyecto, a fin de evitar los tiempos muertos de espera y reducir la duración del proyecto, y por otro que permitan trabajar en conjunto equipos de trabajo situados en distintos puntos geográficos, lo que permitiría elegir en cada momento el mejor equipo de trabajo posible sin que sea un factor determinante donde se va a desarrollar el mismo. La solución a estos problemas viene de manos una de las áreas de las Tecnologías de la Información que ha experimentado mayor auge en los últimos años, el Trabajo Cooperativo.

Uno de los aspectos básicos del Trabajo Cooperativo son las comunicaciones, ya que normalmente los equipos de trabajo se encuentra físicamente separados, desde en despachos diferentes hasta en países diferentes. Además, en el mundo de la construcción un aspecto muy importante es la movilidad del equipo de trabajo, ya que muchos de los cambios del diseño surgen en la misma obra, la cual no siempre está en lugares fácilmente accesibles ni *física* ni *electrónicamente*.

El siguiente aspecto básico a considerar son las interfaces entre los diferentes módulos que intervienen en el proceso. Por ejemplo, los planos de obra civil se deben transferir a las empresas especializadas en instalaciones de servicios, como aguas, electricidad o aire acondicionado, por lo que los sistemas de CAD utilizados por cada uno de ellos deber ser capaz de *entenderse* con los de el resto, de lo contrario se deberían redibujar los planos, con el consiguiente sobrecoste y dificultad de actualización de los mismos. Actualmente ya existen una serie de estándares que permiten operar a las empresas, pero dado que en el futuro la interacción entre equipos de trabajo diferentes será cada vez mayor, y por lo tanto el número de sistemas informáticos diferentes en juego también crecerá, se considera que una línea de trabajo muy importante son los protocolos STEP (ISO 10303. **ST**andard for the **E**xchange of **P**roduct model data) y el software asociado a ellos (arquitecturas, herramientas, etc) para intercambio de datos entre aplicaciones, los cuales ya están muy impuestos en la industria del automóvil y de la aeronáutica y comienzan a estarlo en la de la construcción.

Otro aspecto a tener en cuenta son los temas relacionados con la gestión del flujo de información. Dado que en una obra de cierta envergadura el número de personas involucradas en el desarrollo de la misma es muy amplio, todos los temas de mensajería y flujo de documentos son muy importantes y complejos, siendo necesario mantener en todo momento la coherencia entre los datos usados por cada uno de los miembros del equipo de trabajo. Una de las características principales es que los datos no sólo fluyen en una dirección, si no que las modificaciones a que se somete un dato, por ejemplo un plano, hacen que el flujo se invierta ya que el autor inicial debe ser consciente del cambio realizado, aprobarlo y redistribuirlo.

Igualmente, como consecuencia de la interacción entre diferentes equipos de trabajo sitos en lugares distantes, otra de las características principales es la visualización de la misma información en varias pantallas a la vez, de forma que si se discute sobre un cambio en un plano todos los implicados puedan ver el plano y los cambios propuestos en tiempo real.

6 Principales Proyectos

A continuación se hace un breve resumen de los proyectos y experiencias más significativas desarrolladas

en el área del Trabajo Cooperativo en el sector de la construcción.

Una de las componentes fundamentales de los proyectos de Trabajo Cooperativo es la capacidad de acceso a la información desde diferentes lugares. En esta línea de trabajo se encuentran los proyectos BRICC, MICC y COMMIT.

Proyecto BRICC

Es un proyecto RACE, cuyo objetivo es aprovechar las facilidades que ofrecen las redes de comunicación de banda ancha para agilizar el proceso de toma de decisiones ante el gran número de imprevistos que surgen durante el desarrollo de un proyecto de construcción. Para ello se basa en la gran capacidad de transmisión de datos, audio y video de las redes de banda ancha.

Un factor fundamental para alcanzar este objetivo es mejorar las comunicaciones entre las oficinas técnicas y las instalaciones a pie de obra, ya que esto reduce los desplazamientos desde la oficina a la obra, con la consiguiente reducción de costes y la mayor agilidad en el proceso de toma de decisiones.

Los factores clave para el éxito de este tipo de proyectos son:

- Acceso integrado a todas las bases de datos pertinentes (datos CAD del proyecto y de los suministradores, normativa, proyectos similares, etc).

- Herramientas hipertexto para ligar la información.

- Incremento en las velocidades de transmisión de datos y abaratamiento de las telecomunicaciones, ya que cuando hay posibilidad de disponer de alta velocidad ésta es cara.

- Controles de seguridad en el acceso a la información.

Proyecto MICC

Es un proyecto ACTS cuyo objetivo es dotar a las empresas del sector de la construcción de comunicaciones móviles que le permitan utilizar sus propios sistemas de información desde cualquier emplazamiento dentro de Europa. Para ello se trata de integrar las redes de cables y las de radio en una única red virtual. Así mismo se trata de desarrollar un dispositivo de comunicación personal móvil con acceso a servicios básicos (walkie-talkie, teléfono, e-mail, ...) y acceso a bases de datos en tiempo real.

Proyecto COMMIT

Su objetivo es recopilar los resultados de proyectos anteriores, como el proyecto *ICON*, a fin de modelar el flujo de información a lo largo del ciclo de vida del proyecto y el impacto de los cambios. Para ello se propone el modelo de gestión de la información *CIMM*, el cual aborda los problemas relacionados con el **trabajo en grupo**, como las versiones, las notificaciones, confidencialidad de los documentos, etc.

La necesidad de esta herramienta surge del hecho de que cada vez intervienen más agentes en el proceso de la construcción, los cuales trabajan de forma concurrente desde diferentes localizaciones geográficas y utilizando tecnologías y herramientas heterogéneas.

Se va a desarrollar en *C++* utilizando la herramienta comercial *ORBIX*, que se adhiere al standard *CORBA* definido por la OMG (Object Management Group).

Otro componente fundamental son los modelos de datos que permiten compartir la semántica de la información entre diferentes aplicaciones. En esta línea de trabajo se encuentran los proyectos ROCOCO, VEGA y GOAL.

Proyecto ROCOCO

Fue uno de los primeros proyectos en trabajar en esta línea, ya que comenzó en 1989. Es un proyecto ESPRIT cuyo objetivo era mejorar la productividad mediante la aplicación de TI a la monitorización y control del proceso productivo., atendiendo a aspectos tales como la gran variedad de suministradores, diseño paralelo, etc). Para ello se presta especial atención a los temas siguientes:

- Generación de modelos de referencia para todos los aspectos de los procesos de producción que definan claramente los principales flujos de datos y los procesos de toma de decisiones.

- Rápida realimentación y captura de datos.

- KBS para planificación dinámica de recursos.

Proyecto VEGA

Es un proyecto ESPRIT cuyo objetivo es desarrollar una plataforma que integre todas las actividades de la empresa, tanto los aspectos técnicos como los de gestión. Para ello se basará en las herramientas de trabajo en grupo, las plataformas de sistemas distribuidos y los emergentes estándares de modelos de datos. Las dos líneas principales de trabajo son la utilización de modelos de datos *STEP* y el diseño de una arquitectura que integre el standard *CORBA* con el standard de acceso a la información *SDAI*

(*Standard Data Access Interface)* definido en **STEP**.

Proyecto GOAL

Es un proyecto ESPRIT cuyo objetivo era desarrollar una herramienta de gestión de información y ayuda a la toma de decisiones (PMIDSS - Project Managenent Information and Decision Support System) en grandes proyectos en los que intervienen varias organizaciones. GOAL se compone de 2 partes:

- Open GOAL Model, que es un modelo de datos orientado a objetos para almacenar y transmitir información.

- GOAL toolkit, que son una serie de herramientas de gestión orientadas a proyectos interempresariales.

El carácter interempresarial de los proyectos implica los siguientes condicionantes:

- Necesidad de adaptar GOAL a multitud de aplicaciones y tipos de organizaciones e integración con otras herramientas.

- Capacidad para definir la política de calidad del proyecto.

Otros proyectos se centran más en modelar el proceso de la construcción y analizar la interacción entre las tareas y el flujo de información que estas generan. En esta línea se encuentran los proyectos COSMOS, CICC y ELSEWISE.

Proyecto COSMOS

Su objetivo es desarrollar una herramienta de gestión basada en un modelo del proceso de construcción que permita reducir los tiempos de desarrollo.

La necesidad de esta herramienta surge del hecho de que en los proyectos de construcción cada vez hay una mayor tendencia a la especialización, por lo que cada vez intervienen más agentes en el proceso, con lo que pasa de ser un proyecto único a ser un conjunto de subproyectos, cuya coordinación es cada vez más compleja. Por otra parte hay una presión constante a reducir los tiempos de desarrollo, para lo cual este tipo de herramientas resulta de vital importancia.

Para alcanzar este objetivo se abordarán las siguientes tareas:

- Definición teórica del modelo del proceso de construcción.

- Desarrollo de la herramienta que permita particularizarlo a un proyecto concreto, incluyendo la posibilidad de linkar herramientas de optimización

- Modelos de cada uno de los módulos del proceso de la construcción, con sus relaciones.

- Implementación de modelos de reducción de tiempos mediante la ordenación adecuada de las tareas, desarrollo de sistemas integrados de información, etc.

Como resultado del proyecto surge la herramienta ***PROMO***, que es un editor y browser de modelos de procesos.

Está desarrollada en Paradox for Windows, es compatible con IDEF0, permite ver el proyecto desde varios puntos de vista y genera información en varios formatos (ASCII, HTML, Design Structure Matrix y PlanMan File).

Esta herramienta permite hacer accesible el Sistema de Calidad a toda la empresa vía Intranet, particularizarlo para un proyecto concreto, coordinación de los participantes en el proyecto, ligar documentos relevantes a los hitos del proyecto, etc.

Proyecto CICC

Es un proyecto ACTS cuyo objetivo es demostrar que con unas comunicaciones y un interfaz de usuario adecuados se puede obtener el mismo rendimiento trabajando en grupos separados geográficamente que en un emplazamiento único. Para ello se deberá dotar el sistema de los siguientes mecanismos:

- Modelo de proyecto único on-line para la coordinación del diseño.

- Bases de datos distribuidas para todos los documentos del proyecto.

- Realidad Virtual (VR) para simulación de actividades en el emplazamiento.

- Generación de documentos multimedia para propuestas de ingeniería.

- Buscador de Personas e Información con comunicaciones multimedia, a fin de que todos tengan la sensación de trabajar en la misma oficina.

Proyecto ELSEWISE

Es un proyecto ESPRIT cuyo objetivo es analizar los flujos de información que se producen en los grandes proyectos de ingeniería (***Large Scale Engineering - LSE***) y definir las herramientas de

informáticas y los modelos de datos necesarios (denominados en su conjunto ***Product Data and Information Technology - PDIT***) para las industrias europeas involucradas en proyectos de este ámbito, los cuales se caracterizan por ser proyectos de gran volumen de capital y largos en el tiempo pero con muy poca holgura. Además, demandan gran cantidad de recursos humanos altamente cualificados y presentan una organización muy compleja, ya que hay un conjunto de empresas independientes que deben actuar como una empresa virtual única.

Como experiencia práctica se planteó la aplicación de este tipo de tecnología a la empresa constructora Johnson Johnson Crabtree Architects de Nashville, especializada en construcción de centros hospitalarios, ya que suele trabajar con entorno a 50 subcontratistas, con el consiguiente problema de que la gente reciba la información adecuada en plazo y también la remita a tiempo.

El coste de la puesta en marcha del sistema fue de 60.000 $, más 1.500 $ por subcontratista (8.500.000 ptas + 210.000 ptas por subcontratista).

Con el nuevo sistema los documentos no circulan, sino que están en la base de datos que se replica periódicamente en/contra los puestos de trabajo (incluso puestos móviles), de forma que no se pierde el tiempo en circulación de documentos y todo el personal dispone automáticamente de la última versión. Además se incluyen funcionalidades de agenda para concertar reuniones, etc. Como resultado, el diseño y construcción de un ambulatorio de 60.000 m^2 se redujo en 4 meses (pasó de 12 a 8 meses).

7 Conclusión

La utilización de técnicas de Trabajo Cooperativo hoy en día ya es una realidad a nivel industrial y presenta un potencial enorme en el sector de la construcción, ya que permitiría disponer para cada proyecto del equipo de trabajo más adecuado, así como optimizar su desarrollo, ya que las tareas se podrían paralizar más y la detección y corrección de problemas de diseño se realizaría durante las fases iniciales del proyecto, siendo los problemas resueltos por las personas más capacitadas.

Referencias

[1] Saadoun, M ., "El proyecto groupware", *Ediciones Gestión 2000 S.A.*, **Enero 1997**, (1997).

[2] Grenier, R., Metes, G., "Going Virtual", *Editorial Prentice Hall*, (1997).

[3] Coleman, D., "An Overview of Groupware", *Editorial Prentice Hall*, (1995).

[4] "Groupware", http://www.netconnections.co.uk/spnetcon/2152.htm

[5] Grudin, J., "CSCW: Its History and Participation", http://www.ics.uci.edu/~grudin/CSCW.html

[5] Pfeifer, J., Koo, K., Yin. P., "Electronics Meetings - C S C W & G D S S ", http://ksi.cpsc.ucalgary.ca/courses/547-95/yin/groupware.html

[6] "SDAI: ISO 10303 Industrial Automation Systems Product Data Representation and Exchange Part 22 - Standard Data Access Interface", http://www.nist.gov/sc4/tools/nist

[7] Mowbray, T., Zahavi, R., "The Essential CORBA: System Integration Using Distributed Objects", *Editorial Wiley*, (1995).

[8] "ORBIX. Iona Technologies Ltd", http://www.iona.com/Orbix/Orbix.html

Análisis, Diseño y Desarrollo de un Centro de Teletrabajo

Claudio Feijóo González, Luis-Alfonso Serrano, Luis Castejón, Jorge Pérez
Grupo de Tecnologías de la Información y las Comunicaciones
E.T.S.I.Telecomunicación. Universidad Politécnica de Madrid
Ciudad Universitaria 28040 Madrid
grupo@gtic.ssr.upm.es

Abstract:

Telework Centres gather the basic advantages of Teleworking and avoid up to a certain extent their drawbacks. Hence, Telework will be one of the main new services provided by the Global Information Society. In this communication, the guidelines for the design and development of a Telework Centre are presented. The CENSATE (Centro Satélite de Teletrabajo) has been created taking into account the point of view of the user, therefore evolving from a functional description to a technical architecture. Finally some examples of teleworking activities within CENSATE are introduced.

1. Introducción al Teletrabajo

El desempleo es uno de los problemas más importantes que tienen planteados las sociedades modernas. En la Unión Europea este problema este problema se agrava particularmente, siendo especialmente preocupante en España, que cuenta con unos índices de desempleo desorbitados. Por ello, es necesario encontrar nuevos medios que permitan crear, mantener y mejorar las oportunidades de empleo. El teletrabajo es una respuesta a este reto. Consiste en la organización flexible del trabajo, de manera que no se requiera la presencia física del trabajador en el empresa durante un periodo representativo de su horario laboral.

El teletrabajo tiene **ventajas** indudables para el trabajador que lo hacen indispensable en la futura Sociedad de la Información. En concreto ofrece las posibilidades de crear, mantener y mejorar el empleo para todas aquellas personas que carecen de las oportunidades de tener un trabajo presencial tradicional. Por supuesto, ello ha sido posible debido a la disponibilidad de equipamiento, software y redes de comunicaciones con capacidad suficiente para interaccionar con la información a distancia y manejarla localmente de manera eficaz. Los principales **inconvenientes** con los que se enfrentan los que trabajan a distancia son el **aislamiento** y la **inseguridad laboral**. El trabajador aislado tiene mucho más difícil hacer una carrera profesional de la manera tradicional dentro de una organización, puesto que una carrera se compone de presupuestos profesionales objetivos, pero también de interacciones informales propias de la dinámica de grupos.

Para disminuir en lo posible los inconvenientes anteriores se desarrollan los **centros de teletrabajo**. Un centro de teletrabajo es básicamente una oficina satélite potencialmente muy distante de la nodriza pero que depende en muchos aspectos de ella. Es una situación intermedia entre la oficina tradicional y el trabajo en el domicilio que reúne las ventajas fundamentales del teletrabajo y evita en gran medida sus posibles inconvenientes.

2. Centro de Teleservicios: CENSATE

El **CENSATE** es un centro de teletrabajo considerado en su visión más amplia, pasando a constituir un **centro de teleservicios**, donde se consideran con especial atención todos los aspectos relacionados con **el apoyo, soporte y enlace entre teletrabajadores**. Aparte de las funcionalidades clásicas y ventajas del teletrabajo, el centro alberga **capacidades** para:

- ofrecer la necesaria formación general y específica a los teletrabajadores.
- posibilitar la actualización continua de los conocimientos necesarios para el teletrabajo.
- conectar a los teletrabajadores con las organizaciones destinatarias de su trabajo mediante las redes de comunicaciones más baratas y eficaces.
- compartir recursos caros, inabordables económicamente de otra manera.

- mantener los mecanismos de comunicación formal e informal habituales entre trabajadores, reduciendo su aislamiento y facilitando la solución de problemas y la conciencia de colectivo unido.

- disponer de elementos de seguridad que mantengan la información y el control sobre la misma.

- efectuar las pruebas de equipamiento, software y redes que mantengan las ventajas competitivas de los teletrabajadores asociados al centro.

- detectar nuevas necesidades de información que pudieran constituirse en futuras oportunidades de empleo.

Así, gracias a las facilidades proporcionadas por las nuevas tecnologías de la información y las comunicaciones, un centro de teleservicios puede ser una de las aplicaciones clave, por su potencial de creación y renovación de empleo, dentro de lo que se ha venido en llamar la Sociedad de la Información.

Por supuesto, este centro sirve de enlace entre los teletrabajadores y el resto de agentes implicados, es decir:

- **empresas** y organizaciones con necesidades de trabajo relacionado con el acceso y uso de información.

- **proveedores** de redes de comunicaciones, software, hardware y servicios asociados a la información.

- **gabinetes** de apoyo en cuestiones laborales, jurídicas, organizativas, administrativas, contractuales, ...

- organizaciones dedicadas a la **formación.**

En este sentido, el proyecto contará con representantes de todos estos agentes, de forma que se garantiza el éxito de sus funciones de enlace y como generador de oportunidades.

3. CENSATE. Innovación y Nuevas Aportaciones

La **innovación** del proyecto reside fundamentalmente en establecer las estructuras de apoyo y de enlace necesarias para mejorar la seguridad laboral y la comunicación en el teletrabajo. Así se eliminan los problemas de inseguridad y aislamiento que constituyen el principal inconveniente del teletrabajo autónomo, mientras que se aprovechan las oportunidades que representa como medio de fomento y mantenimiento del empleo.

La **estructura de apoyo** está basada en un **centro de teleservicios**, el **CENSATE**, que permite ofrecer un medio para la **formación** general y específica de los teletrabajadores, un procedimiento continuo de **actualización de los conocimientos** necesarios para el desempeño del teletrabajo, medios físicos para la **conexión** de los teletrabajadores con las empresas y organizaciones destinatarias de su trabajo de la forma más barata y eficaz posible, una forma de **compartir recursos** caros entre los teletrabajadores, un mecanismo de enlace y **comunicación entre los teletrabajadores**, los elementos de **seguridad** necesarios para que la información objeto del trabajo este asegurada continuamente y este accesible solamente para las personas y organizaciones adecuadas, una plataforma de **pruebas** de nuevo equipamiento, y un observatorio de **detección de necesidades** para ofrecer toda la capacidad de las tecnologías de la información y las comunicaciones a nuevas posibilidades de interacción con la información y, por tanto, de empleo. Todos estos objetivos serían inalcanzables para un teletrabajador aislado.

El proyecto contempla también que, eventualmente, el centro de teleservicios pueda convertirse en una **empresa virtual** que ofrezca como **servicios** las capacidades de los teletrabajadores asociados.

4. Objetivos Generales

Este proyecto pretende definir y evaluar el plan de desarrollo imprescindible para la creación de un centro de teleservicios. Profundiza en el análisis de áreas de teletrabajo, competencias, carencias formativas y de seguridad, viabilidad económica, etc.

Así mismo se diseña e implementa un centro de teletrabajo con todos los servicios asociados, funcionalidades, equipamiento, software, redes, interconectividad, aplicaciones avanzadas, etc. Posteriormente se evaluarán estos resultados para optimizar el rendimiento del centro.

5. Funcionalidades del CENSATE. Hitos Relevantes Alcanzados

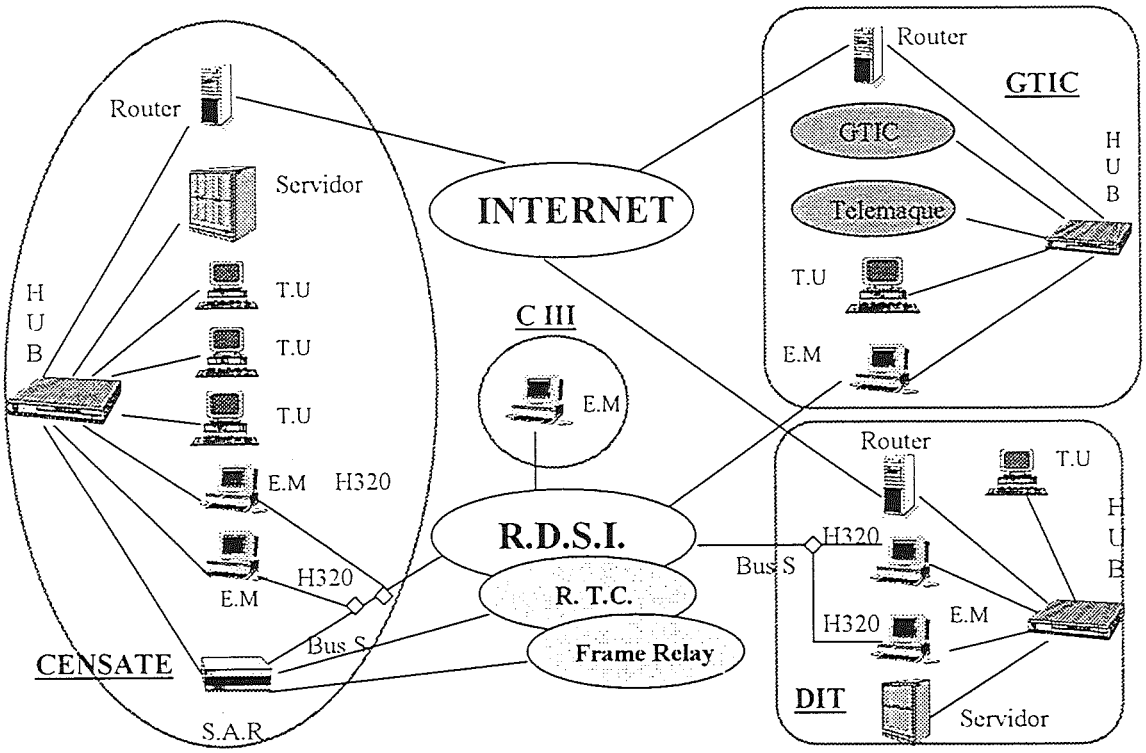
El CENSATE ha completado su primera fase y, por tanto, cubre las necesidades básicas

especificadas. Éstas son (desde el punto de vista funcional):

- Servidor convencional de ficheros y dispositivos: discos de backup, scanner, impresora color, etc.
- Servidor de procesos: proporcionar capacidad de procesamiento masivo para ejecución de procesos remotamente.
- Servidor de correo electrónico: para intercambio de mensajes internos y externos.
- Servidor de acceso a otras redes: Gateway a *Internet*, consultas a bases de datos...
- Servidor Multimedia: aplicaciones sobre HTTP, (servidor WWW, evaluación de prestaciones, etc).
- Trabajo cooperativo
- Servidor de base de datos.
- Aplicaciones de videoconferencia.

A continuación se presenta la arquitectura física del CENSATE como centro de teleservicios. Posteriormente se detallará la configuración de cada elemento integrante.

5.1 Arquitectura física global



5.2 Descripción de elementos integrantes

5.2.1 Terminal de Usuario (T.U.)

- Pentium 120 Mhz
- 16 MB de RAM
- VGA de 64 bits con 2 MB de RAM
- Bus: 2 ISA y 3 PCI
- CD-ROM 8X
- Monitor 14 pulgadas
- Disco duro IDE 1.2 GB
- Disquetera 3''1/2
- Windows 95
- Tarjeta de red Ethernet

5.2.2 Equipos Multimedia (E.M.)

- Pentium 133 Mhz
- 32 MB de RAM
- VGA de 64 bits con 2 MB de RAM
- Bus: 3 ISA y 4 PCI
- CD-ROM 8X
- Monitor 17 pulgadas
- Disco duro IDE 1.6 GB
- Altavoces
- Sound Blaster 16 bits

- Disquetera 3''1/2
- Windows 95
- Tarjeta de red Ethernet

5.2.3 Servidor con UPS

- NetServer LX PRO Monoprocesador
- Ampliable hasta 4 procesadores
- Pentium Pro 166 Mhz
- 128 MB de RAM
- 512 KB de Caché interna
- Bus: 4 EISA y 6 PCI
- CD-ROM SCSI 4X
- Disco duro de 16 GB (12 GB lógicos)
- RAID 5 :Disk Striping with Parity
- Sistema Fast&Wide SCSI2
- SVGA de 14 pulgadas
- UPS HP de 600 KVA para 10 min
- Tarjeta de red Ethernet

5.2.4 Comunicaciones intra-LAN

- HUB Ethernet HP
- Capacidad de 12 puertos
- Gestión SNMP
- Cableado de Categoría 5 : 100 MHz.
- Sistema de gestión de red OpenView

5.2.5 Conexión con Internet

- Conexión con Red Iris e Internet a través de un PC 386 sobre Linux haciendo las funciones de rutador (router) y cortafuegos (firewall).

5.2.6 Comunicaciones acceso remoto: R.T.C./ R.D.S.I/ Frame Relay

- 1 Router CISCO 2503
- Multiprotocolo: IP, FR, X.25, PPP...
- 1 interfaz Ethernet y 2 interfaces serie
- 1 BRI RDSI
- 1 conector V.35 para FrameRelay
- Software IOS
- 1 Router ASCEND MAX 200 plus
- Multiprotocolo: IP, IPX...
- 8 slots PCMCIA (8 tipo RTC, 4 RDSI)
- 1 modem PCMCIA 28.800 bps (V.34)
- 1 modem PCMCIA RDSI BRI

5.2.7 Videoconferencia

- PictureTel LIVE 200P
- Tarjeta Audio-Video RDSI
- Para Windows 95
- Cámara PAL
- Micrófono
- Altavoces
- Compatible H320
- Aplicaciones de trabajo compartido
- Compatible T120 (futura ampliación)
- Posibilidad de conexiones por LAN

5.2.8 Scanner + Impresora

- HP DeskJet 1600 CM
- Calidad laser B/N : 600x600 ppp
- Alta calidad en color : 300x300 ppp

5.2.9 Software

- Microsoft BackOffice 2.0
- Netscape Mail Server (gestión correo)
- SQL Server 6.0
- Internet Information Server
- Windows NT Server 4.0
- Systems Management Server
- Microsoft Windows 95 para EM y TU
- Lotus Notes Domino

6. Desarrollo del Proyecto

El cuadro siguiente muestra las fases originales del proyecto, su duración y su cumplimiento.

FASE	DURACIÓN	RESULTADO	GRADO DE CUMPLIMIENTO
1. Estudio y desarrollos básicos	9 meses (1-7-96 al 1-4-97)	Centro básico para Teletrabajo	**Completada (100%)**
2. Primera experiencia de teletrabajo	7 meses (1-4-97 al 1-10-97)	Generación de cursos mediante teletrabajo	**Iniciada (15%)**
3. Incorporación de desarrollos avanzados y seguridad	9 meses (1-9-97 al 1-6-98)	Centro avanzado de Teletrabajo	No iniciada
4. Segunda experiencia de teletrabajo	7 meses (1-5-98 al 1-12-98)	Demostración de generación de empleo	No iniciada
5. Diseño y gestión de teletrabajo con EPOs	2 meses (1-12-98 al 1-1-99)	Transmisión de conocimientos	No iniciada
6. Tercera experiencia de teletrabajo	7 meses (1-1-99 al 1-8-99)	Demostración en entorno real de trabajo	No iniciada
7. Evaluación global	2 meses (1-5-99 al 1-7-99)	Diseño, desarrollo y gestión de centros de teletrabajo	No iniciada

7. Proyectos sobre el CENSATE

7.1 Teleformación en Internet-Intranet

Se imparten a un grupo de teletrabajadores cursos técnicos completamente interactivos y prácticos enfocados a ampliar los conocimientos de los trabajadores sobre las redes tanto WAN (Internet) como LAN-MAN (Intranet).

7.2 Teleformación avanzada

Se imparten desde el CENSATE unos cursos a teletrabajadores remotos y se supervisa todo el proceso prestando especial atención a la monitorización del proceso de aprendizaje.

7.3 Jornadas de Administraciones Locales y TIC

Se realizan unas jornadas técnicas en colaboración directa con el CENSATE sobre la importancia y la ideoneidad de la integración de las Tecnologías de la Información y las Comunicaciones en el entorno de trabajo de las Administraciones Locales. Se pretende concienciar y formar a los responsables de las Administraciones Locales para que comprendan las nuevas realidades tecnológicas y hagan un uso óptimo de ellas.

7.4 Base de datos TIC

Un teletrabajador especializado lee a primera hora todas las noticias del sector TIC y elabora remotamente una base de datos con las noticias actualizadas y resumidas. Esta base de datos está en formato Web y es fácilmente accesible desde Internet. También se realizan estadísticas y análisis de los patrones de uso.

7.5 Elaboración de cursos bajo Web

Varios teleformadores elaboran a través del CENSATE unos cursos en formato Web que son fácilmente accesibles desde Internet. Se realizan, así mismo, controles de rendimiento y seguimiento de los cursos, ejercicios, estadísticas y análisis de los patrones de uso.

7.6 Teleadministración

Se realizan experiencias piloto de administración pública remota a través del Centro de Teleservicios.

7.7 El Web del teletrabajador

Se elaboran a través del CENSATE toda clase de documentos e informes que sirven de apoyo a todos los teletrabajadores que consulten el servidor Web del Centro de Teleservicios.

Agradecimientos

Este trabajo se ha realizado con la ayuda de la Comunidad Autónoma de Madrid y de la CICYT (TEL96-1399-C02-01)

Grupo III:
Aplicaciones en Educación

Teleeducación

Una Intranet Educativa

M. J. VERDÚ PÉREZ, M. A. PÉREZ JUÁREZ Y R. MOMPO GÓMEZ
DEPARTAMENTO DE TEORIA DE LA SEÑAL, COMUNICACIONES E INGENIERIA TELEMATICA
E.T.S.I. DE TELECOMUNICACIÓN DE VALLADOLID. UNIVERSIDAD DE VALLADOLID
C/ REAL DE BURGOS S/N. 47011 VALLADOLID. ESPAÑA.
TEL.: +34-83-423260. FAX.: +34-83-423261
Correo electrónico: marver@tel.uva.es. marper@tel.uva.es. jyr@dvnet.es

Abstract:

The main purpose of the authors is to incorporate Information Technology to the learning process (primary and secondary levels). As a first objective, we intend to reduce the difficulties that rural area children have due to they atending education in schools far from their living places. We also want to incorporate computers, multimedia materials and Internet tools to the classrooms. This will allow teachers and pupils to generate their own multimedia and Internet materials and to use all the resources Internet offers them. Besides, cooperative learning could be done even between pupils from different schools. In order to gain our objectives, we are developing an Educational Intranet and some multimedia CD-ROMs to support our 'virtual classroom'.

1. Introducción

En el trabajo que aquí se presenta se busca aplicar las Nuevas Tecnologías de la Información y las Comunicaciones (NTICs) en el proceso de formación de la población infantil de Castilla y León. Se trata de la creación de un *instrumento de apoyo* a la formación básica para una población infantil ultradiseminada. El objetivo final consiste en la creación de una **Intranet Educativa** con contenidos de apoyo a la formación y servicios que compensen la inaccesibilidad a esos mismos servicios culturales cuando éstos se encuentran localizados en zonas urbanas.

Los autores del presente artículo formamos parte de un grupo de trabajo denominado "Grupo Canalejas", unos de cuyos campos de trabajo son la multimedia educativa y las redes telemáticas aplicadas en los contextos educativos.

El "Grupo Canalejas" es un grupo interdisciplinario e interuniversitario que investiga las posibilidades y ventajas de la aplicación de las nuevas tecnologías de la información y las comunicaciones al proceso formativo. En concreto, el grupo está formado por dos equipos de investigación, uno de la *Universidad de Salamanca* y otro de la *Universidad de Valladolid*. El primero lo constituyen profesores de la Facultad de Educación de Salamanca y el segundo profesores de la Escuela Técnica Superior de Ingenieros de Telecomunicación de Valladolid y una licenciada en psicología, profesora de la Escuela de Profesorado de E.G.B de Segovia.

La razón fundamental que justifica la participación de los dos equipos en las tareas que el grupo está realizando, es el carácter *interdisciplinar* de todo el proyecto. Por la propia naturaleza de las investigaciones y experimentos, se requiere la presencia de profesionales pertenecientes, al menos, a dos campos de trabajo: de un lado, el campo de la educación, ya que los productos elaborados van a utilizarse en procesos de aprendizaje; y de otro, el campo de las telecomunicaciones, más concretamente el ámbito de la telemática, ya que el grupo confía en la informática, la multimedia y la telemática como elementos que conformen la solución a muchos de los problemas que se presentan en los contextos educativos. Y la presencia de una licenciada en psicología nos permitirá profundizar más y mejor en los tan importantes (y muchas veces olvidados) aspectos psicológicos de la introducción de Internet en los escenarios educativos.

Hay que añadir que el presente grupo de trabajo cuenta con el apoyo de una empresa de nuestro entorno regional llamada Divisa Informática S.A.

2. Internet en las Escuelas

Uno de los proyectos que tenemos entre manos el "Grupo Canalejas" y sobre el que trata este escrito, es el de ***Internet en las Escuelas***. Consiste en realizar un experimento piloto sobre las posibilidades de la red Internet en la enseñanza primaria y secundaria. En el experimento tomarán parte unos 100 colegios de la comunidad de Castilla y León, que utilizarán los servicios típicos de Internet más otros particulares que nosotros les ofrezcamos a través de nuestra Intranet.

Internet en las Escuelas es un proyecto que sintoniza perfectamente con la política regional de telecomunicaciones ya que utiliza una tecnología universal, barata y accesible desde cualquier rincón de la región.

Nuestro objetivo es doble:

- Servir de prueba piloto para evaluar las posibilidades de la utilización de la red Internet en la enseñanza primaria y secundaria, haciendo

gran hincapié en las posibilidades de Internet para reducir considerablemente los desplazamientos de los estudiantes de las zonas rurales a centros educativos de mayor tamaño.
- Formación ocupacional para los estudiantes que participen en esta prueba piloto, pues tendrán la oportunidad de aprender a generar contenidos para la red Internet, algo que ya empieza a ser valorado a la hora de buscar empleo. Con esta experiencia se introducirán en la nueva *Sociedad de la Información* y ayudarán a que dicha nueva sociedad avance.

3. Objetivos del proyecto

A grandes rasgos, lo que queremos es construir una Intranet para canalizar diversos servicios de apoyo a la formación. Los objetivos genéricos que perseguimos con ello son los siguientes:

- Incorporar los conceptos y la utilización de las herramientas de la *Sociedad de la Información*, en la impartición de las materias comunes de la escuela. Es decir, contar con nuevos medios donde poder consultar la información acerca de las materias tradicionales. Dichos medios podrían ser información multimedia contenida en soporte material (CD-ROM) y obtenida a través de las redes telemáticas (Internet/Infovía).
- Utilizar las posibilidades que nos ofrece la telemática como 'elemento de equilibrio socioeconómico y regional': la creación de servicios de apoyo por medio de las redes telemáticas nos va a ayudar a reducir las diferencias existentes en la educación de los niños de Castilla y León, debidas precisamente a las características demográficas de dicha comunidad.
- Incorporar a las escuelas los recursos materiales necesarios y dotar a profesores y alumnos de la formación necesaria para poder utilizar dichos medios. Queremos que estos nuevos recursos *se integren en el aula como un elemento más.*
- Promover la capacidad de publicar información propia (*generación de contenidos*) en Internet así como la posibilidad de implementar diversas aplicaciones telemáticas.
- Fomentar la cooperación entre escuelas, así como el intercambio de experiencias y conocimientos en el ámbito de las nuevas tecnologías.

Una consecuencia indirecta de este proyecto consiste en la experimentación de un instrumento básico para el fomento del uso de las NTICs en la región de Castilla y León, el cual, superada la fase piloto, podría ser generalizado.

4. Estructura demográfica y educación en Castilla y León

La acción a llevar a cabo toma en consideración la estructura de poblaciones diseminadas que caracteriza la región de Castilla y León, con el inconveniente, en muchos casos, de los traslados de los niños en período de educación obligatoria a centros escolares fuera de su lugar de residencia. Esta dispersión trae además consigo la separación de los núcleos de población de los puntos en los que radican los servicios culturales.

Si consideramos la educación como un bien universal y como una necesidad social, entonces ésta ha de llegar a todos los estratos sociales, y al último rincón donde exista un asentamiento de población.

La ubicación de un centro escolar pone al alumnado en condiciones de desigualdad según sea la proximidad al mismo. Esta desigualdad es más fácil de resolver en zonas urbanas, en donde la concentración de la población es muy alta, que en áreas rurales, donde la población se encuentra más dispersa [1].

Queda pues establecida una desigualdad de carácter geográfico entre los núcleos urbanos y las zonas rurales, existiendo entre los primeros una *mayor facilidad física de acceso educativo.*

Hay que tener en cuenta que el acceso educativo no se reduce a la escuela tradicional, sino que abarca también otros entornos de aprendizaje como universidades, bibliotecas, museos, fuentes prácticas de conocimiento tales como centros de investigación y desarrollo.... y otra serie de servicios culturales que suelen ofrecerse en los grandes núcleos urbanos. En definitiva, la mayor accesibilidad a todos estos recursos de información y formación por parte de la población urbana, pone a las comunidades rurales en una situación de *pobreza de información* frente a las urbanas.

Diferencias de este tipo existen prácticamente en cualquier región del mundo. Sin embargo, estos problemas se acentúan en regiones como Castilla y León, donde la población está ultradiseminada. En Castilla y León hay 2248 municipios, de los cuales 2117, según datos del año 1995 [2], tienen menos de 2000 habitantes. Esos 2117 municipios constituyen el 94.17 % del total.

La existencia de un excesivo número de localidades con reducida población dificulta la dotación de servicios públicos mínimos. Centrándonos en la educación, hay que añadir además la importancia de la estructura por edades de la población de Castilla y León, ya que es

precisamente en esos pequeños y numerosos núcleos rurales donde la población está más envejecida, concentrándose de manera más notable la juventud en los núcleos urbanos.

Es necesario introducirse en la realidad de una comunidad para poder señalar las líneas de actuación adecuadas en materia de educación. La realidad de Castilla y León es muy diferente a la de otras comunidades de España. La ruralidad que la caracteriza de una manera tan especial no ha podido ser tenida en cuenta desde los programas generales del MEC [3]. Hay que actuar por tanto, siguiendo los objetivos fijados después del análisis de dicha realidad.

Por tanto, queda clara la importancia del factor espacio como elemento clave en el análisis educativo y la necesidad de aplicar políticas educativas regionales que respondan a las necesidades actuales o futuras de planes de desarrollo integrado de dimensión regional. Y esas políticas, según las necesidades actuales, deben conllevar la introducción de la tecnología, porque de hecho, la tecnología es la solución a muchas de essas necesidades.

La dotación de servicios públicos mínimos puede resultar cara en regiones como Castilla y León, en las que existe un excesivo número de municipios con reducida población. Mediante el empleo de la Tecnología de la Información podemos proporcionar esos servicios mínimos de la forma más barata y más eficiente posible. Con este fin, y en el campo que nos ocupa, se han aplicado y se aplican políticas de concentraciones y comarcalizaciones escolares. Estas políticas son necesarias, porque las arcas de la Comunidad y del Estado tienen un fondo, pero deben ser apoyadas por una política de telecomunicaciones que ayude a minimizar los perjuicios que supone esta política de concentración escolar a los niños de las zonas rurales. Se trata de, al menos, reducir esa desigualdad que existe entre los niños de zonas urbanas y los de zonas rurales, y esto no se puede hacer con medidas estrictamente educativas o desde el sistema de enseñanza. En todo esto juega un papel muy importante la Sociedad de la Información.

5. El papel de la Sociedad de la Información en las Escuelas

Actualmente estamos presenciando la evolución de nuestra Sociedad Industrial actual hacia una nueva *Sociedad de la Información*. Dicha sociedad será una sociedad en la que los ciudadanos sean capaces de hacer uso de diversos servicios de telecomunicación avanzados para mejorar los distintos aspectos de su vida cotidiana [4]. Uno de los aspectos que se puede sin duda mejorar es el de la educación, a todos los niveles: educación obligatoria (primaria y secundaria), enseñanza universitaria, formación permanente...

En la enseñanza tradicional los alumnos asisten a clase a una hora determinada y la calidad del profesorado y, por tanto, de la enseñanza, depende de la oferta docente de la zona en la que se encuentre su centro. Con la Sociedad de la Información llega la teleeducación: los alumnos pueden recibir la mejor formación donde quieran y a la hora que prefieran y siempre pueden comunicarse con el profesor, sin tener que residir éste en la misma zona que los alumnos. El que todos los alumnos tengan acceso a la enseñanza del mejor profesor en un determinado tema, hace realidad el dar a todos las mismas oportunidades de formación.

Por otra parte, las NTICs también van a facilitar el acceso a todo tipo de centros de investigación, Universidades, museos, bibliotecas,... que, junto con la escuela, también deben contribuir al proceso de formación del niño. Sin embargo, hasta ahora, este hecho era otra fuente más de diferencia de oportunidades.

Además, el empleo de materiales multimedia interactivos y de hipertextos, va a promover un método de enseñanza basado en búscar información, adquirir conocimientos y resolver problemas. El profesor toma un nuevo rol de orientador y diseñador de medios y métodos. El alumno pasa a ser el protagonista del proceso de formación, un 'investigador' que activamente busca información, la analiza y es capaz e incorporarla a proyectos colaborativos de trabajo [5]. Con las nuevas tecnologías se potencia la utilización de nuevos métodos de enseñanza más abiertos, en los que se promueve el trabajo personal y colaborativo.

En definitiva, por primera vez en la historia, el acceso a la información y al conocimiento en centros escolares va a convertirse en mundial, más abierto, y más asequible [6].

Pero la relación entre la Sociedad de la Información y las escuelas no termina aquí. Las escuelas también tienen un papel activo en el desarrollo de esta nueva sociedad.

6. El papel de las escuelas en la nueva Sociedad de la Información

Hemos visto que la enseñanza puede beneficiarse mucho de la nueva *Sociedad de la Información*. Pero dicha sociedad no existe todavía en su plenitud. Todos y cada uno desempeñamos un papel fundamental en la construcción de esta nueva

sociedad, ya que la misma será o no realidad en la medida en que los ciudadanos logren dominar el uso de los nuevos servicios de telecomunicación.

También hemos hablado de que la nueva sociedad es el resultado de una nueva revolución. Pues una buena parte de la culpa de que esta revolución esté teniendo lugar podemos echársela a la creación y desarrollo de Internet.

Cualquier tipo de información puede encontrarse hoy ya en Internet mucho antes de que aparezca publicada de forma impresa. Además, los usuarios de Internet pueden copiar dicha información a su ordenador independientemente de la parte del mundo en que se encuentre el ordenador en el que se haya almacenada la información.

Dos usuarios situados en puntos geográficamente distantes pueden comunicarse mediante mensajes electrónicos o, incluso, establecer una conversación en tiempo real. Asimismo, grupos de personas, alejadas geográficamente entre sí pero interesadas en un mismo tema, pueden intercambiar opiniones e información a cerca del mismo.

El enorme crecimiento de las redes de comunicación a partir de 1990 ha acelerado el establecimiento de la biblioteca virtual, una parte esencial de la cual está constituida por la información de dominio público. A estos recursos, almacenados en bases de datos en todo el mundo, se puede acceder de forma casi instantánea mediante Internet : catálogos de bibliotecas, periódicos electrónicos, artículos de todo tipo...

El hecho principal y a tener en cuenta es que recursos de todo tipo -programas, información, imágenes...- que se encuentran almacenados en distintos ordenadores conectados a Internet puedan ser compartidos por el resto de usuarios de Internet.

Hasta hace pocos años, el uso de Internet quedaba reducido al ámbito científico- académico, mientras la mayor parte de la sociedad permanecía al margen. En nuestros días, la situación ha cambiado drásticamente, no sólo ha aumentado el número de usuarios de forma considerable sino que además, Internet ha logrado penetrar en un amplio conjunto de sectores de la sociedad. Este cambio radical se debe fundamentalmente a la disminución del coste del hardware -ordenadores...- y a la creación del World Wide Web.

¿Qué papel deben jugar los centros educativos en esta nueva revolución liderada por el fenómeno de Internet? El paso de la sociedad industrial a la sociedad de la información está suponiendo, por lo que puede deducirse de todo lo comentado anteriormente, un *nuevo modo de conocer* [7]. El modo al que accedemos a la información, el modo en el que trabajamos, ha cambiado, y la escuela debe cambiar. Por tanto, es en la escuela donde los niños, futuros dirigentes y trabajadores de este país, deben adquirir destreza suficiente para poder manejarse en la *Sociedad de la Información*: deben conocer y aprender a utilizar los recursos de lo que ya hoy es una realidad en otros contextos de la vida: Internet y sus servicios.

7. Las Intranets

Una **Intranet** es una red de ámbito privado, a la que sólo acceden los usuarios autorizados. Es decir, la información y los servicios de la red no están a disposición de todo el mundo. Lo que caracteriza a una **Intranet**, que la distingue de otras redes privadas, es que usa los mismos protocolos y programas que Internet. Este es el motivo de que las **Intranets** sean más baratas (tanto a la hora de ponerlas en funcionamiento, como a la hora de mantenerlas) que las redes privadas que usan sistemas propietarios de colaboración, de correo o de trabajo en grupo en general. Otra ventaja de una **Intranet**, es que en la mayoría de los casos todo lo que necesita el usuario es un navegador. El navegador actúa como interfaz común para el acceso a la información y a todos los servicios ofrecidos por la **Intranet**, lo que reduce la necesidad de entrenamiento especializado [8] [9].

Una **Intranet** se podría definir como 'una red corporativa que se lleva a cabo utilizando el soporte de la red **Internet**'.

8. La 'Intranet Educativa'

Dos son las características que hacen que una Intranet sea especialmente idónea para un contexto educativo en la sociedad actual: bajo coste y facilidad de uso. El establecimiento de una **Intranet Educativa**, por otra parte, proporciona nuevas oportunidades para que los estudiantes desarrollen habilidades en lo que se refiere al manejo de la información, tanto en su generación, como en su búsqueda y en su manipulación. Y todo este aprendizaje se realiza de una manera interactiva y participativa, lo que contribuye a mantener la motivación de los alumnos [10].

Por tanto, dado que es una tecnología accesible y barata, podemos incluirla en un plan de telecomunicaciones para la enseñanza primaria y secundaria, con el objetivo de disminuir las desventajas que sufre la población rural diseminada de Castilla y León, acercándole una serie de servicios a los que ahora tienen más difícil acceso

que la población urbana. Aquí se pretende la creación y puesta en uso de una **Intranet** que vincule a estos grupos infantiles a un centro de apoyo a la formación, el cual actúe como compensación de la distancia a otros focos de acción de formación. Intentamos romper su aislamiento respecto a servicios, cultura e información.

También, mediante CD-ROMs y a través de la Intranet, los alumnos tendrán acceso a cursos multimedia interactivos de calidad y contenido adecuado a su nivel educativo. El profesor hará de guía en el seguimiento de los cursos, y, mediante herramientas telemáticas se posibilitará la comunicación del niño con el profesor así como con otros alumnos, para la realización de trabajos, resolución conjunta de problemas...

Al mismo tiempo, el establecimiento de una **Intranet Educativa** con acceso controlado a **Internet**, permitirá que los alumnos participen en proyectos de comunicación e intercambio universal, así como que adquieran destrezas de manejo, gestión y selección de la información, destrezas necesarias para el siglo que se aproxima.

9. Descripción del proyecto *Internet en las Escuelas*

En *Internet para las escuelas* se propone el establecimiento de una Intranet Educativa, que conlleva una serie de fases, como se expone a continuación.

9.1 Creación de un Centro Servidor Internet

En este centro servidor residirían aplicaciones multimedia de carácter didáctico, lúdico o cultural. Se utilizaría el servidor del Centro para el Desarrollo de las telecomunicaciones de Castilla y León (CEDETEL), situado en el Parque Tecnológico de Boecillo. Las escuelas accederán a este centro servidor a través de Infovía. Los profesores y alumnos depositarán su información en dicho servidor.

9.2 Creación de la Intranet Educativa

Para ello son necesarias dos etapas.

En la primera etapa se procede a la identificación de las poblaciones rurales para la experimentación de la **Intranet**. Asimismo, identificación de los centros escolares implicados en la formación de esas poblaciones de alumnos, grupos experimental y de control.

En la segunda etapa se lleva a cabo el diseño de la **Intranet**, estipulaciones para hacerla operativa y primeros ensayos. Se diseña la Intranet según los puntos siguientes:

- Habitualmente los alumnos sólo tienen acceso a la **Intranet Educativa**. Con ello se controlan a los contenidos a los que acceden los alumnos.
- El acceso libre a **Internet** sólo será posible cuando alguien autorizado (por ejemplo el profesor) envíe un comando de desbloqueo del acceso a **Internet**, desde su ordenador.
- También es posible permitir el acceso sólo a determinados ordenadores de **Internet**. Por ejemplo, a aquellos de otros Centros Educativos que formasen una red europea de centros docentes.

También se lleva a cabo una identificación de los servicios a ofrecer: campos de conocimiento, tipos de actividad y características de los diseños. Entre los servicios a ofrecer tenemos un servicio de intercambio documental y de conocimiento entre profesores y alumnos de diferentes centros, que permitirá a los alumnos consultar dudas y comentar temas con otros profesores distintos a los de su escuela. Y también hay un servicio de orientación para los alumnos de último curso de enseñanza secundaria en materia de posibles carreras universitarias y salidas al mercado laboral. Mediante este servicio los alumnos recibirán la atención vía correo electrónico de profesores de distintos centros universitarios que se presten a colaborar en el servicio.

9.3 Formación del profesorado y alumnos

Tanto a los profesores como a los alumnos de las escuelas se les formará en la utilización de Internet, así como en la generación de contenidos para Internet (elaboración de páginas Web).

Además a los profesores se les formará en la generación de contenidos multimedia, para que puedan elaborar CD-ROMs con algunos temas de sus asignaturas.

La formación del profesorado es especialmente importante por dos motivos:

Primero, porque uno de los factores que retrasa la introducción de las nuevas tecnologías en el sector de la educación es el de la necesidad de formar a los profesores para que puedan utilizar las herramientas multimedia en su trabajo diario [6].

Y segundo, porque un problema que ha existido desde que aparecieron los primeros materiales multimedia educativos es su falta de calidad en lo que se refiere al aspecto pedagógico. Formando a los profesores en la generación de contenidos multimedia se podrá conseguir material educativo multimedia de más alta calidad pedagógica.

9.4 Atención permanente a las escuelas

Se atenderán dudas tanto acerca de la generación de contenidos como sobre los equipos instalados. Además se establecerá una línea permanente de atención telefónica y por correo electrónico.

9.5 Gabinete de producción de contenidos educativos

Se creará un gabinete de producción de contenidos educativos que trabajará en equipo con los profesores de las escuelas. El objetivo de este gabinete es ayudar a que los profesores generen algún contenido multimedia en CD-ROM. El trabajo en equipo entre formadores y técnicos ayudará a que los productos obtenidos sean de calidad tanto en su aspecto pedagógico como en el técnico.

10. Equipamiento necesario en los centros

Con el objetivo de soportar todo lo especificado, y a modo de orientación, se propone el siguiente equipamiento en los centros:

- Cinco PCs multimedia conectados en red local
- Una impresora
- Un scanner
- Una cámara de vídeo, para uno de los PCs
- Un Infolink, que ofrece a cada puesto de la red local la posibilidad de acceder a Internet y a InfoVía.

La configuración descrita se completaría con software de tipo educativo, ofreciendo, en conjunto, las siguientes posibilidades:

- Conexión Internet/Infovía, con capacidad para utilizar todas las herramientas disponibles en dichas redes: Correo electrónico, acceso a Grupos de Noticias (News), transferencia de ficheros, acceso a servidores Web,....
- Posibilidad de realización de páginas Web, que podrán ser cargadas en un servidor Web local
- Edición de gráficos
- Capacidad de edición multimedia, con posibilidad de incluir imágenes procedentes de distintas fuentes de vídeo
- Servidor de transferencia de ficheros (ftp) local
- VídeoConferencia, a través de Internet, en un puesto de cada red local
- Audioconferencia, a través de Internet, en todos los puestos de la red local

El software educativo cubrirá una serie de áreas (enciclopedia de propósito general, literatura, matemáticas, atlas geográfico, etc.). Sin embargo, no se pretende cubrir todas las áreas educativas ni todos los ciclos. Lo que se pretende es ofrecer una selección de dicho software orientado a mostrar las posibilidades de dichas aplicaciones multimedia.

Por otra parte, se incluiría el software Microsoft Office, que incluye las herramientas Word, Excel, Power Point,..., lo que permite utilizar cada PC como un puesto de trabajo completo.

Si bien todo el equipamiento descrito anteriormente es el que consideramos que puede soportar de forma más eficaz y diversa la Intranet Educativa, somos conscientes de los costes que supone. Pero en un futuro, esperemos no muy lejano, con el abaratamiento de los equipos y algún tipo de ayuda económica lo haremos posible. Hasta que llegue ese momento, y para lograr nuestros objetivos básicos, es suficiente con que los colegios tengan acceso a Internet y una infraestructura de equipos mínima para poder experimentar con los servicios que nosotros ofrecemos.

11. Conclusiones y Trabajos Futuros

En esta comunicación hemos expuesto nuestro proyecto de una Intranet Educativa en colegios de primaria y secundaria de Castilla y León, con el doble objetivo de disminuir las desventajas que tienen los niños de zonas rurales respecto a los de zonas urbanas, y de introducir a profesores y alumnos en los nuevos métodos de trabajo de la Sociedad de la Información.

Internet en las Escuelas puede ser una realidad, y una realidad beneficiosa, como hemos intentado demostrar en este informe. Lo que pretendemos nosotros es abrir una puerta hacia una infraestructura de enseñanza única, global y unificadora para Castilla y León; una infraestructura basada en las telecomunicaciones que nos permita hacer realidad la idea de una ***Escuela Abierta***, que llegue a todos los rincones de la comunidad.

Para ello, en este momento contamos con el ATF (Asistente Telemático de la Formación) que la empresa Divisa Informática S.A. está desarrollando. Las herramientas que permiten el contacto de los alumnos entre sí y con el profesor, así como las herramientas de gestión escolar, van integradas en el ATF. También hemos seleccionado los colegios y hemos diseñado contenidos para la Intranet, así como manuales para enseñar a maestros y alumnos a elaborar páginas Web.

En breve tendremos en marcha la intranet con 'usuarios reales' y analizaremos las consecuencias de la introducción de Internet en los colegios, buscando nuevos servicios que ofrecer que cubran las necesidades existentes. También

estudiaremos los aspectos psicológicos y pedagógicos, teniendo en cuenta los problemas que surjan de este tipo, analizarlos, conocer cual es su fuente, y ver cómo podemos solucionarlos. Siempre con el objetivo último de proporcionar una enseñanza mejor para todos.

Referencias

[1] Monreal, J (Ed.). *Población y Estructura Educativa*. Secretariado de Publicaciones. Universidad de Murcia. Murcia (1982).

[2] Junta de Castilla y León. *Datos Estadísticos de los Municipios de Castilla y León. 1997*. Junta de Castilla y León. Consejería de Economía y Hacienda. Servicio de Estudios (1997).

[3] López Hernández, V., González Hernández, C., Arias Franco, M. F., y Gutiérrez Turrión, L. "La Educación de Personas Adultas en Castilla y León: Ruralidad, Formación y Desarrollo". *Actas del Congreso 'La Educación de Personas Adultas en Castilla y León'*, Junta de Castilla y León, Consejería de Educación y Cultura, Valladolid (1994).

[4] Grupo de Análisis de la Sociedad de la Información, *España en la Sociedad de la Información*, Colegio Oficial de Ingenieros de Telecomunicación, Madrid (1996).

[5] Cabero Almenara, J. "Navegando, construyendo: la utilización de los hipertextos en la enseñanza", Universidad de Sevilla. *<http://www.doe.d5.ub.es/te/any96/cabero_hipertext/* > (22 de Abril, 1997).

[6] Information Society Project Office, *La Sociedad de la Información... y el Ciudadano*, Informe sobre la disponibilidad y la utilización de los sistemas de información y comunicaciones, ISPO, Bruselas (1996).

[7] Bartolomé Pina, A. R. "Preparando para un nuevo modo de conocer", Departamento de Didáctica y Organización Educativa, Universidad de Barcelona. *<http://www.doe.d5.ub.es/te/any96/bartolom_pineda/* > (22 de Abril, 1997).

[8] Holtz, S. *The Intranet Advantage*, PCWEEK, USA (1996).

[9] "Intranet Corporativa. La Web interna". Informe *PC Magazine*. Madrid, Año 9 / No. 93, pp. 124-170 (1996).

[10] Alonso Tapia, J. *Motivación y aprendizaje en el aula. Cómo enseñar a pensar*. Ed. Santillana, Madrid (1994).

DISTANCE LEARNING WITH WWW

Manuel JUAN ESCRIVÁ
Fátima MARTÍ ADSUAR
Daniel PALACIOS MARQUÉS
Departamento de Sistemas Informáticos y Computación
Universidad Politécnica de Valencia
e-mail: palacios&dsic.upv.es

ABSTRACT

This paper presents our current work on a distance learning system using Internet as communication media and world wide web servers as centre of information supply. Wide availability of Internet access on a campus and e-mail complement the proposed system. The paper also presents a practical case using free software tools and different hardware platforms. In fact, the described system is designed to be platform independent.

1.- INTRODUCTION

Our starting point is the usual way of teaching at this university and covers different learning items and a usual way of spreading these items (e.g. photocopied lessons or slides). With this in mind, we have oriented our search to available tools for simplifying publishing (teachers) and accessing (students) to the different kinds of media (not only text : also video, audio and multimedia presentations).
Another point we have taken into account is the interaction issues between students and teachers that usually takes place in the classroom (speeches) and in the form of individual questions in the teacher's office.

WEB-RELATED TECHNOLOGIES

We have found that WWW servers [1] and all the different plug-ins available for web browsers enable them to be a proper platform for our needs.

Our work is based on the next key elements and technologies :

1. A WWW server that contains the whole information of different subjects. (We are using Linux 3.0 and NCSA httpd free software).
2. Students use web browsers to access to the services of our group of subjects, we are currently using NetScape 2.0 (which also includes Mail and News reader).
3. NetScape plug-ins enable browser to access all media types (e.g. StreamWorks [2] lets real-time video and audio, Shockwave [3] lets multimedia presentations). This kind of enhancements suits the client-server model. While client software is usually free, the server one must be bought (we have tried beta versions of them or servers out of our campus).
4. VRML extensions enable non-inmersive virtual reality web pages that show a more friendly approach to the traditional classroom model.
5. Other tools, in the form of separate programs, allow us to complete the scene to enable chat sessions (e.g. NetScape Chat)
6. Another interesting tool could be "Talk" that lets users establish communication without the need of an IRC server.

All the proposed tools and environment can run over different

All the proposed tools and environment can run over different hardware platforms, either PCs as Macs (available in practically every laboratory). We are currently using Microsoft Windows 95 in our PCs but this is not a requirement (most of the software we are using is free and available for MacOS and Linux).

A PROPOSAL OF A DISTANCE LEARNING SYSTEM

What we are presenting is a complex system that integrates all the above software to supply students with a distance learning equipment. This lets them to do almost the same work as a regular student can do in our classes.

Environment description

A common infrastructure in many universities is a campus wide LAN. In our case, each building has a 10BASE-T Ethernet connected to each other by means of a 100 Mbps FDDI network.

Our first step consists on a server machine with Linux OS powered by a Pentium with 32 Mb of RAM and a 2GB hard disk. This computer has an anonymous FTP account that holds all the free software needed to access the services, and lets students downloading it onto their computers.

The above server also has a web server that students can browse and it contains a huge amount of information about several subjects.

Services description

There are two types of services which has been deployed to connect users to each other or to access to information sources: asynchronous and synchronous. Traditionally, communication services in distance learning belong to the asynchronous type. Examples are electronic mail (e-mail), file transfer (ftp), searching service (archie) or any other service offered by a traditional WWW server. These services are very useful for assisting students questions or bibliographic searches in any time. But the on-line connection to the teacher may be also interesting in some situations. Even other students may simultaneously connect and interchange their opinions or views about a topic. The integration of both kind of services can be represented by the concept of "virtual classroom".

This concept represents a virtual view of the entities which are in a classroom and the operations which are defined to manage such entities. For example, an entity "blackboard" can be defined in order to allow the teacher to display information about a subject. The information stored in the "blackboard" can also be protected with only read or read-write permissions or it can be assigned with other kind of attributes. Other type of entities can be defined, depending on the information which can be showed in the classroom: "slides" to show a presentation, "blackboard" to display a lesson, "audio" and "video player" to reproduce sounds or images.

One of the main advantages of these "virtual" entities is its availability in temporal and spatial terms. One student can connect from his home with a remote teacher and he can also send a question at any time. There can be a number of instances of connected "teachers" in a given time and the student can also connect with other students.

The services provided by the entity "teacher" are not limited to show information. It can also answer to questions which are sent by the students in voice, text or graphic format. There can be interactive sessions where student and teacher can set up a dialogue based on multiple media. These on-line services are complemented with asynchronous services which allow to collect requests independently of the time when they were produced.

While the asynchronous services are implemented in a easy manner using the current communication technology, the

synchronous services impose strong requirements. These requirements are related with the multimedia characteristics of the information sources managed in distance learning [4] (e.g. voice and video). Another feature consists of the multiple sources of information which can be activated simultaneously and the synchronisation needed between these sources.

When you connect to these services, the first page you view is a road map of the different subjects on the server.
If you click on the different elements you go to the proper page:

- The blackboard lets you access to the available lessons. You can read them on screen or print them for a later use. (Our photocopy shop is not very glad with this).
- The teacher icon lets you post a message to the teachers of this subject.
- The slide icon lets you access to the available presentations of this subject.
- The student icons, which appear labelled with a name, let you post a message to this student.
- If you click on the table you open a chat application that lets you talk in real time to the other students that are currently in this virtual classroom.

After explaining how our system works, we are going to show the information of the involved subjects, students can access:

- The Web server holds the lessons (as a book, but with the added benefit of live modifications and corrections. You always access an up to date version of each lesson). We are developing a virtual-reality interface trying to ease the use of the system, in order to make it gentle even to people not related to computer science matters. We are planning to include new possibilities like audio and video inside a lesson, integrating into it other material usually not available to the students outside the classroom (e.g. a video that covers all the aspects of a network card installation).
  You can also find the slides and multimedia presentations used in the classroom.
- The Mail system lets communication between students and teachers and allows to exchange all kind of information and also to gather homework. We are currently evaluating on the issue of an E-mail exam system (e.g. you can send a different test each time).
- The IRC server enables a way of on-line talk between students, but we think that on-line services are less interesting because they force a synchronisation in time we cannot guarantee.
- Our News server is, in a certain way, the off-line version of the chat system referred above. However it is also the place where you can find the FAQs and, in general, the place where public technical discussions take place.
- The Talk tool lets students establish a person-to-person private on-line talk, useful to a rapid interchange of short information (e.g. to meet at the coffee machine).

We have been trying to establish a common environment to let computer aided distance learning a realistic option, however the full power of the system needs a big pipe greater than current phone modems can offer. We hope that wide availability of narrowband ISDN (and perhaps, the future broadband ISDN) will raise the possibilities of these new ways of using computers to get a course of any degree.

A PRACTICAL CASE

In this section we are going to explain the different services provided for a WWW server implemented at the Computer Science Faculty of the Polytechnic University of Valencia. At the beginning of the 95-96 course, this server started supporting only to Telematics, which is one of the subjects of the Faculty, but at the moment the rest of the subjects related to Computer Networks are offering services on this server. These are mainly:

- Publication of marks. Students can know the result of their exams even from their homes. This option is really interesting for students who live out Valencia.
- Storing of academic information. Usually, at this University students only can buy practices bulletins in the photocopy shop, however Telematics students have another possibility: get them through our web. This option is free and they do not need to queue up.

  In addition, exams solutions are also stored in the server. Students can consult them after finishing their exams and they can get an idea about the results of them. Solutions are also useful to check student's knowledge when they are preparing an exam.
- Procedures of registration and access control to the practical sessions of a course. This utility provides the tools for access control of each students group to the laboratory. Each group has to register at the beginning of the session, when they come in the laboratory and start using a computer, and also when they finish their turn. Registering allows the teachers to check easily the presence of each group, as well as the use of the laboratory along each session, that help us to choose the suitable duration of a session. For example, during the course 95-96 the duration of each telematics session of laboratory was three hours, but we have detected that many students were not at the laboratory more than two hours, so next year we are going to do two sessions of ninety minutes.

  Another point related to access control is how to book a session. Here, different subjects have different needs. Usually subjects with many students, as Telematics (about four hundred students), have a static timetable. At the beginning of the course each group reserves his turn each week and they are always the same turn during all the course. There are another subjects (generally not obligatory) that prefer dynamic reservation. In this case each group can reserve several sessions by week until the maximum tolerate, changing the sessions each week.

  In addition to obligatory sessions, there are also free access sessions. In this time students can complete and improve their practices. Access control program lets them to reserve this kind of sessions and allows us to check the presence of each group in real-time.

  Moreover, the control of results can also be included in this service, so the student can transfer electronically his work or fill-in a questionnaire displayed by the web browser.
- Possibility to connect with the teacher to ask some topic about the course. This is an example of integration of a typical Internet application like the "electronic mail" into the WWW. Each teacher has a page web with information about him, mainly interesting data for the

students : timetable for attention to them, e-mail address, etc.

Using electronic mail, students have an important tool to do questions about the subject, they do not need to go physically to the teacher's office. It is also useful because e-mail lets students practise with this tool and, on the other hand, writing their doubts is helpful to them to organise and to clarify their ideas.

Including teacher's timetable in the web allows the students to have always the timetable updated, so they do not have problems to talk with their teacher if they want to do it face to face.

- Next course, the WWW server will include a list of the most frequently asked questions (FAQ service) about the subject Telematics. This list will be updated using the queries which will have been obtained by the previous service. This one is useful both teachers and students. As the student's doubts are very often the same, students can sometimes find the solution to their questions in this FAQ, so they do not need to ask to the teacher. Reading the FAQ's can also help them to pose themselves new questions.

  Teachers can use this information to emphasise in the classroom the aspects of the subject more frequently asked, because they are more complex or the students are more interested in them. These FAQ's can also be useful to evaluate the subject and trying to answer them, students will check his knowledge about the subject and teachers can take out ideas to make tests.

CONCLUSION

The proposed system has several important features we consider the key of success :

- Used software is free (at least for educational use) and available for each known platform.
- Different free tools have eased publishing efforts because you can choose HTML converters for all the principal word processors and presentation programs.
- Available computers in the campus meet the minimum hardware configuration needed to run this software.

After two years of experimentation the proposed system has shown its usefulness, and we have found that students access rate is raising day by day. However, we also want to point that the not so wide availability Internet access offer has the effect of limiting out of campus operation. Moreover, this university does not let phone remote access to students, which means that students must sign an agreement with an Internet Service Provider (not free). In any case, we want to reinforce that extra-campus access is limited by the low speed of these connections, although we hope this later problem will be solved in a near future.

REFERENCES

[1]T.Berners-Lee et al. "The World-Wide-Web", Comm. ACM, vol. 37, No. 8, Aug. 1994 pp. 76-82

[2]StreamWorks software, Xing Technologies Inc. *http://www.xing-tech.com*

[3]Shockwave software, Macromedia Inc. *http://www.macromedia.com*

[4]N. J. Muller "Multimedia over the network", BYTE magazine, March 1996, pp. 73-82

Servicios de Telepresencia en la Facultad Virtual

Guillermo Gil, Carmen Pastor, Roberto Uriarte
ROBOTIKER, ÁREA DE TECNOLOGÍAS DE TELECOMUNICACIÓN
PARQUE TECNOLÓGICO DE ZAMUDIO, EDIFICIO 202
48170 ZAMUDIO, BIZKAIA
Correo electrónico: guille@robotiker.es

Abstract:

With the continuos availability of more powerful networks and more powerful systems at lower prices, and with the development of phenomenas like the INTERNET, new telecommunication services have appeared in the domain of distance and open learning. From the technology point of view, these services are based on electronic mail, *web*, chat, docu-conferencing or interactive audio-visual services. These basic services are used to provide tele-presence, tele-teaching and tele-training. Among them, the new high quality, audio-visual services are having increased impact as they allow for the distance presence of a lecturer given by a teacher, avoiding unnecessary displacements of students from one site to the physical scenario where the lesson is actually given.

1. Introducción

Los servicios audiovisuales, que son ampliamente utilizados en el mundo de la Radiodifusión, tienen, sin embargo, una utilización muy limitada en aplicaciones interactivas. Esta limitación viene dada por las capacidades de las redes, los costes de los sistemas y la calidad de los servicios, en general muy inferiores a los equivalentes de televisión.

En los últimos años han aparecido un gran número de sistemas y aplicaciones basados en la comunicación audiovisual interactiva. Sin embargo, la calidad de estos sistemas en comparación con los servicios audiovisuales tradicionales es realmente pobre, y muy limitada para su utilización como soporte a, por ejemplo, la tele-formación.

Por otro lado, el desarrollo de la tecnología de vídeo digital de alta calidad empieza, aunque con todavía grandes limitaciones, a extenderse. Empiezan a aparecer codificadores y de-codificadores MPEG2 de bajo retardo que posibilitan la utilización de este método de compresión no sólo para aplicaciones de radiodifusión sino también para aplicaciones interactivas de video-comunicación. En paralelo, las redes de mayor capacidad se están desplegando en ámbitos restringidos como empresas de alta tecnología, parques tecnológicos y campus universitarios.

Este artículo analiza un conjunto de alternativas que se han contemplado desde ROBOTIKER para la provisión de servicios de tele-presencia. Se ha seleccionado esta aplicación por tratarse de una necesidad real existente en estos ámbitos previamente mencionados, que los sistemas y tecnologías tradicionales no pueden resolver con satisfacción. Para ello, se considera la combinación de diferentes tecnologías audiovisuales de alta calidad y diferentes infraestructuras de red. Finalmente se presenta una plataforma para la demostración de la formación a distancia. Esta plataforma se está poniendo en marcha dentro del proyecto ADTT2 - *Advanced Digital Television Technologies 2 - VIDEOCOM*, proyecto EUREKA EU1711.

2. Necesidades y soluciones tecnológicas para la Tele-formación

La principal necesidad que pretende cubrir la tele-formación en el ámbito universitario es la de evitar el desplazamiento de personas de unas ubicaciones a otras, tanto de alumnos como del profesorado. Los alumnos deben ser capaces de realizar el seguimiento de las clases, trabajar en equipo, ejecutar experimentos, formar preguntas, entregar trabajos, etc., evitando al máximo los innecesarios desplazamientos. Los profesores deben ser capaces no sólo de impartir la lección, sino de gestionar el desarrollo de la misma: recibir el *feedback*, contestar preguntas, establecer los programas y los métodos didácticos adecuados, etc.

Sobre estas ideas, a las que se han añadido otras que completan la tele-enseñanza efectiva, se ha construido el término "Facultad Virtual".

Las soluciones tecnológicas a la "Facultad Virtual" son variadas, e incluyen, además de la tecnología en la infraestructura de red, cuestiones tales como:

1. correo electrónico y grupos de noticias;
2. acceso a WWW;
3. vídeo-conferencia y docu-conferencia, y otras aplicaciones de vídeo (servidores de vídeo, p.e.);
4. aplicaciones de presentación remota (*whiteboard* o pizarra electrónica);
5. aplicaciones de gestión y control de puestos remotos.

Dentro de estas posibles soluciones que se pueden establecer para el concepto de "Facultad Virtual", la Tele-presencia aborda las necesidades relativas a un profesor cuando imparte su clase en tiempo real a un conjunto de alumnos, aislados o agrupados, físicamente localizados en ubicaciones remotas.

2.1 Tele-presencia

Tele-presencia es un término genérico que se emplea para la coordinación, cooperación y control remoto en diferentes aplicaciones: tele-operación, tele-compra, tele-banca, tele-enseñanza, etc. y que pretende sustituir operaciones que hasta la fecha se desempeñan en ámbitos locales. La Tele-presencia implica recibir información precisa para actuar, bien físicamente (tele-operación médica) o "virtualmente" (tele-edición CAD). Se trata de una comunicación interactiva, bidireccional y multimedia.

3. Alternativas para la prestación de servicios de Telepresencia en la formación

La utilización de soluciones de Tele-presencia en la formación requiere menos opciones que en la mayor parte de aplicaciones posibles, ya que el contenido de información es fundamentalmente audio-visual y no es preciso controlar u operar dispositivos remotos. Las aplicaciones de Tele-presencia para formación deben garantizar por tanto una alta calidad audiovisual. Las tecnologías disponibles en el mundo de la radiodifusión, como la Televisión de Alta Definición o, incluso, la Televisión de Definición estándar, digitalizadas con las técnicas MPEG2 pueden representar un nivel aceptable, tras las limitaciones de los sistemas audiovisuales del mundo de los ordenadores (H.261, MPEG1, etc.)

El soporte empleado para ello debe ser compatible con los equipos e infraestructuras disponibles. En el ámbito universitario se dispone de infraestructuras avanzadas, tanto de equipamiento como de red y se están empezando a desplegar redes de banda ancha capaces de soportar el tráfico de los servicios audiovisuales de alta calidad. Las redes lógicas se configuran en torno a Internet, con el protocolo IP como el corazón de las mismas. Las capacidades multi-punto se soportan con conceptos como MBONE (*Virtual Multicast Backbone On the interNEt*). Para las redes físicas existen diferentes opciones desde el *Frame Relay* hasta el ATM (*Asynchronous Transfer Mode*), en función de la capacidad requerida y la disponibilidad de recursos.

Aunque para la provisión de servicios de Tele-presencia no debiera ser un requisito su soporte sobre IP, éste se considera interesante de cara a favorecer la prestación de otros servicios simultáneamente, como los de *Chat*, los de correo electrónico o pizarra electrónica. Es por ello por lo que se ha considerado esta cualidad en las tres alternativas analizadas para la prestación de la Tele-presencia del proyecto ADTT2:

1. Televisión de Alta Definición (HDTV), con compresión MPEG2, sobre Internet multi-punto (MBONE);
2. Televisión de resolución estándar sobre MBONE;
3. canal independiente de HDTV sobre redes de banda ancha (ATM).

A continuación se describen en detalle estas alternativas.

3.1. HDTV sobre MBONE

El empleo de tecnologías de alta definición dentro del entorno de los equipos informáticos aún se encuentra en fase de desarrollo, por lo que existe una gran dificultad para obtener tarjetas digitalizadoras/codificadoras de MPEG2 con entrada de vídeo de alta definición.

Una solución a considerar consiste en utilizar un dispositivo *standalone* con un interface ATM para codificar la señal MPEG2. Este equipo enviaría la señal a una estación de trabajo con interface ATM que recogería la señal MPEG2, y la introduciría en el canal IP multicast, y se enviaría a la red MBONE.

En recepción, sería necesario tanto decodificar como visualizar la señal proveniente de la red. Debido a las especiales características de una señal de alta definición, este proceso posee varios aspectos a considerar. Sin embargo, el aspecto más crítico es la disponibilidad de equipamiento compatible que permitiera extraer la señal audiovisual del canal IP.

La única posibilidad de tratar imágenes en alta definición en recepción sería equivalente a la del emisor: la aplicación recibiría los paquetes MPEG2/IP de la red ATM, y reenviaría los paquetes ATM al decodificador *standalone* que visualizaría la imagen de HDTV.

Otro problema adicional de esta alternativa es la capacidad de los dispositivos de red para gestionar tráfico IP de alta velocidad (en torno a los 40 Mb.p.s.). Entre estos dispositivos hay que mencionar los concentradores de red, los *Mrouters* (*routers multicast*), los protocolos de adaptación al *backbone* ATM, etc.

3.2. SDTV y MBONE

Dadas las dificultades en el tratamiento de HDTV sobre IP, cabría la posibilidad de emplear imágenes de alta calidad, pero en resolución estándar. Este hecho exigiría capacidades de 2 a 16 Mb.p.s. para la transmisión del contenido audiovisual en la red IP. Hay que tener en cuenta que la calidad de televisión de radiodifusión se considera aceptable para la mayor parte de las personas. En este caso existe en el mercado un mayor abanico de equipos (cámaras, codificadores, dispositivos de red, etc.) por lo cual en emisión es posible instalar una tarjeta codificadora MPEG2. La señal codificada es introducida en MBONE a través de aplicaciones que realicen el encapsulamiento.

En recepción, el proceso es inverso, una vez decodificada la señal, es necesario visualizarla en pantalla. Este proceso es más sencillo, ya que es posible usar tarjetas de vídeo y monitores estándar usando tarjetas de superposición de vídeo y decodificadores hardware o software.

La aplicación software en recepción debe recoger los paquetes MPEG2 en IP, decodificarlos y luego enviarlos al proceso de visualización o, en el caso de un decodificador que entregue una señal de vídeo, enviar esta señal a la tarjeta de superposición para que la visualice en pantalla.

Como ventaja de esta alternativa, está, además de su viabilidad tecnológica, la capacidad multicast y su flexibilidad para adaptarse en el futuro a estándares de imagen de mayor resolución.

3.3. HDTV y ATM

La tercera alternativa es mantener la calidad del canal audiovisual en Alta Definición pero sin utilizar las capacidades IP multicast, lo que evitaría los problemas de introducción de la información audiovisual en el canal IP y su tratamiento posterior en los dispositivos de red. Para ello, se construiría un canal audiovisual específico que uniría las sedes remotas con la sede de emisión. En paralelo a este canal, sería preciso otro canal de datos para intercambiar la información complementaria (correo electrónico, *webs* referentes a la lección, la pizarra electrónica, etc.)

Debido a los costes de esta alternativa, se minimizarían los puestos desde donde se pudiera seguir una sesión de tele-formación. Estos puestos podrían consistir en "aulas virtuales" dotadas del equipamiento necesario ubicadas en puntos estratégicos de la organización.

4. El proyecto ADTT2 *Advanced Digital Television Technologies* (EU1711)

El proyecto ADTT2 pretende demostrar las posibilidades de las nuevas aplicaciones audiovisuales, en concreto las de televisión digital de alta definición. Parte de los resultados de proyectos anteriores en el ámbito de la Televisión digital, pero con aplicaciones específicamente interactivas, frente a las de difusión del proyecto precedente. Otro de los aspectos innovadores es la utilización del soporte ATM como infraestructura de red.

4.1. Plataforma de demostración

La plataforma de demostración plantea la utilización de redes de comunicaciones de alta velocidad (red de banda ancha ATM, redes CATV, redes *switched Ethernet*) de forma que se permita la transmisión de la información multimedia (especialmente vídeo) de forma óptima, garantizando la calidad del servicio. Para los servicios sobre soporte IP se utilizarán protocolos de emulación de redes locales sobre ATM (VLAN y LANE).

La solución técnica está basada en la siguiente arquitectura (véase Fig. 1):

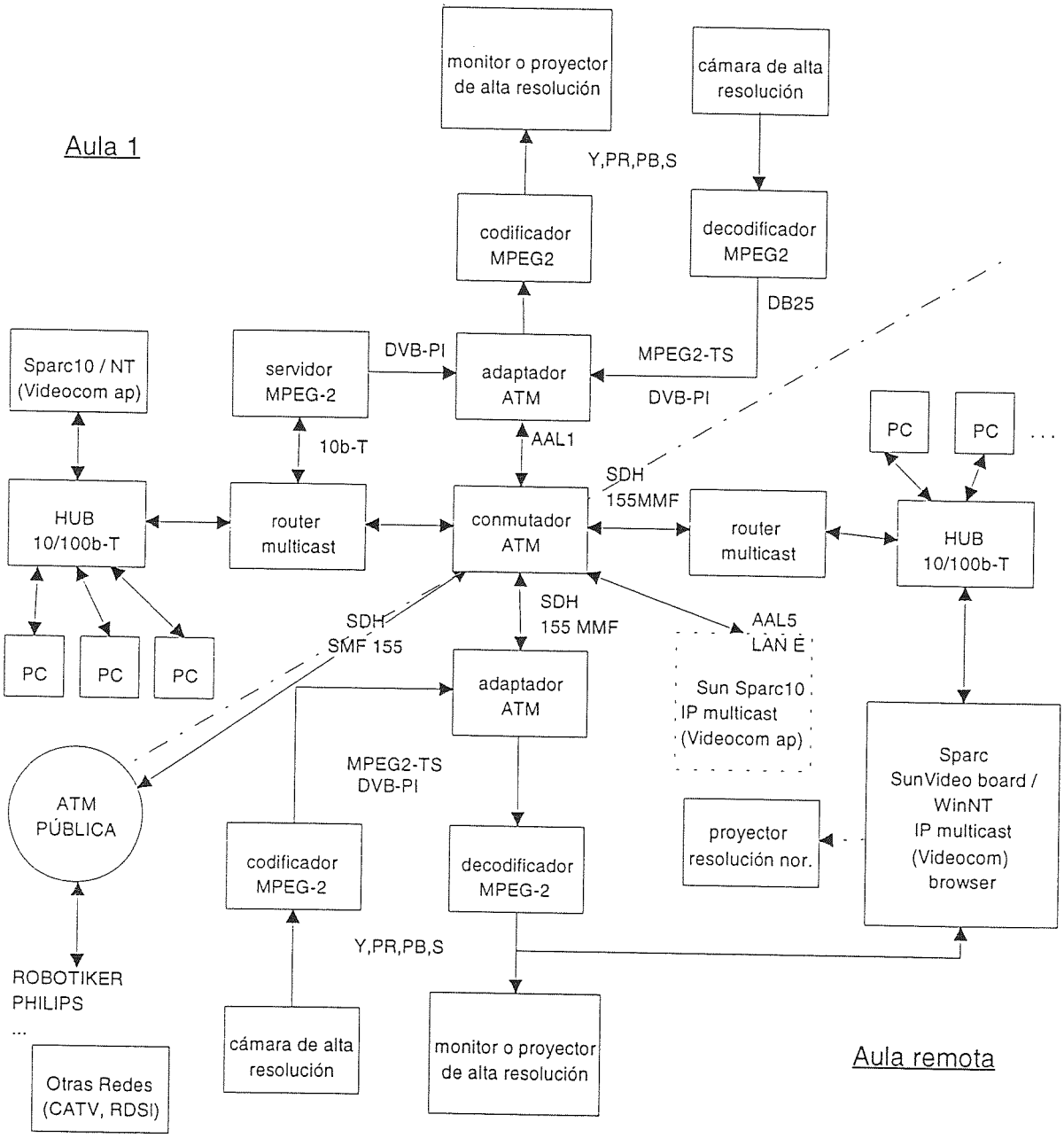


Fig. 1: Arquitectura Global del Demostrador

Esta solución se basa en las alternativas propuestas anteriormente. El sistema descrito proporciona un canal paralelo de datos (IP multicast sobre ATM) añadido al servicio de videoconferencia sobre ATM, basado en compresión MPEG2 para obtener una mayor calidad. Este canal de datos posibilita la interconexión de estaciones de trabajo o PCs remotos sobre la red ATM, de forma que puedan trabajar sobre un espacio compartido común a todos ellos. Con ello, se consigue compartir aplicaciones típicas del mundo Internet (docu-conferencia, correo electrónico, *web*, etc.).

La red IP multicast puede ser implementada mediante la conexión de las redes locales Ethernet existentes sobre el backbone de ATM, utilizando para ello routers multicast Ethernet-ATM.

Finalmente, mencionar que la existencia de adaptadores ATM es sólo precisa si los equipos a conectar a la red ATM (codificadores MPEG2, servidores de vídeo MPEG2) no incorporan este interface.

5. Conclusiones

De las evaluaciones realizadas en diferentes experiencias, resulta evidente que la Tele-formación, en cuanto a la prestación de enseñanza presencial a distancia, requiere de unas

capacidades audiovisuales superiores a las que los servicios de vídeo-comunicación proporcionan en la actualidad. La limitación de estas capacidades es una consecuencia directa de las infraestructuras de comunicaciones empleadas. La Tele-formación con estas herramientas se centra en experiencias de baja duración con impacto muy limitado: en estas experiencias existe el seguimiento por parte de los alumnos, que se hace complejo por las calidades del vídeo y del audio, pero se carece por completo de los mecanismos y medios para que el profesor pueda supervisar y controlar el desarrollo de la acción formativa.

Con el despliegue de nuevas redes avanzadas en el ámbito universitario unido a la progresiva disponibilidad de tecnología audiovisual digital (MPEG2, por ejemplo) se hace posible la Tele-formación, resolviendo algunos de los problemas anteriores y siempre en condiciones acotadas. Con ello se permite la Tele-presencia del profesor en determinas localizaciones dotadas del equipamiento y medios necesarios. Sin embargo, las experiencias existentes con estas nuevas tecnologías carecen de flexibilidad y resultan de cierta complejidad de operación, lo que las hacen solamente adecuadas por el momento a experiencias de demostración poco operativas.

Referencias

[1] *MBone: A Primer.* Datapro Information Services Group, a division of The McGraw-Hill Companies, Inc.

[2] *Video on the PC.* Datapro Information Services Group, a division of The McGraw-Hill Companies, Inc.

[3] *Video Technology.* Datapro Information Services Group, a division of The McGraw-Hill Companies, Inc.

[4] López Angulo, T. "Análisis del Estado del Arte. Plataforma Avanzada Multimedia. MPEG2". *ATT/13297/AEA/MPG.* ROBOTIKER (1997).

[5] Pastor, C. y Uriarte, R. "Proposal for the IBC demonstration". *SER/20297/INF/PLA.* ROBOTIKER (1997).

[6] San Millán, A. "Análisis del Estado del Arte. Plataforma Avanzada Multimedia. ATM". *ATT/13297/AEA/ATM.* ROBOTIKER (1997).

Un proyecto de teleformación: la Facultad Virtual

J. ARAMBERRI, F. ABAL, M. GAMBOA, J. LASA Y J. MIGUEL
DEPARTAMENTO DE ARQUITECTURA Y TECNOLOGÍA DE COMPUTADORES
FACULTAD DE INFORMÁTICA, UNIVERSIDAD DEL PAÍS VASCO UPV/EHU
APDO. 649 20080-SAN SEBASTIÁN
Tel: (943) 218000 / Fax: (943) 219306 / E-mail: acparmij@si.ehu.es

Abstract

Terms such as "Virtual Campus" or "Virtual Faculty" are currently being used to refer to the development of educational activities using telecommunication resources that provide powerful mechanisms for distance learning. Traditional universities follow a "synchronous" teaching model, where instructor and students join together in the same space (the classroom) and at the same time (that given by the timetable). In order to translate this scenario to a distance learning environment, telepresence systems are required; examples of those are videoconferencing systems, and many classes of groupware. The implementation of such systems is of special interest for traditional universities, because it is a way of broadening their area of action by means of removing the distance barrier.

1. Introducción: Concepto de "Facultad Virtual" o "Campus Virtual"

El término de "Facultad Virtual" o "Campus Virtual" comienza a utilizarse para referirse al desarrollo de la actividad docente utilizando recursos de telecomunicación que proporcionan mecanismos para la formación a distancia.

La Universidad tradicional sigue un modelo "síncrono", coincidiendo profesor y alumnos en el tiempo (horario) y en el espacio (clase, seminario, laboratorio). Para trasladar este escenario a la formación a distancia se necesitan sistemas de "telepresencia", que permitan la colaboración de alumnos y profesores situados en lugares alejados: aplicaciones de videoconferencia y de trabajo cooperativo.

También existen Universidades dedicadas a la "Formación a Distancia" basadas en sistemas de "autoestudio" (UNED, UOC). Están dirigidas a aquellas personas que, por razones de ocupación profesional u otras, establecen su propia planificación para el estudio. Se trata de un modelo "asíncrono", en el que no coinciden profesor y alumnos ni en el tiempo ni en el espacio, salvo para contadas actividades (tutorías, evaluaciones). El soporte telemático de estas organizaciones requiere sistemas de comunicación asíncrona, como el correo electrónico, acceso a materiales docentes de autoestudio almacenados de forma electrónica, y otros mecanismos de búsqueda y consulta de información (bases de datos de Biblioteca, acceso a Internet, etc.)

La "Facultad Virtual" es de especial interés para la Universidad tradicional, ya que amplía considerablemente su radio de acción al eliminar la barrera de la distancia. Hoy día existen experiencias con los sistemas de "telepresencia" antes mencionados, y también se están utilizando sistemas de autoestudio en universidades como la UOC. Una utilización óptima de los recursos de comunicación pondría a disposición de la actividad docente los dos sistemas, que son claramente complementarios.

Es también al mismo tiempo un reto, pues va a permitir una competencia "real" entre Universidades. En pocos años encontraremos ofertas educativas, incluso de otros países, accesibles mediante sistemas telemáticos en nuestro propio entorno geográfico. La competencia no vendrá de las Universidades basadas en el modelo "asíncrono", sino de otras Universidades tradicionales. El mercado casi "cautivo" de la única Universidad Pública de la Comunidad Autónoma del País Vasco será accesible a otras muchas Universidades, públicas o privadas.

2. La UPV/EHU y el "Campus Virtual"

Los grandes números dicen que en la Universidad del País Vasco / Euskal Herriko Unibertsitatea hay unos 3.300 profesores, 65.000 alumnos, y tres campus principales, además de una serie de centros fuera de los campus. Se trata de una universidad tradicional, orientada a la enseñanza "presencial".

Una de las directrices organizativas de la UPV/EHU es evitar, en la medida de lo posible, la duplicación de su oferta docente en los diferentes campus. El resultado de esta decisión es que la oferta de titulaciones se dispersa por la geografía de la CAPV, y que los alumnos tienen que optar por:

1. Elegir una universidad privada que funcione en su territorio.
2. Trasladar su lugar de residencia hasta el lugar de estudio.
3. Elegir una titulación diferente a la deseada.

4. Realizar todos los días el trayecto domicilio-centro de estudios.

Las dos primeras opciones plantean dificultades serias, fundamentalmente desde un punto de vista económico. Algunos alumnos no acceden a las universidades privadas porque cobran tasas más elevadas que las de la pública, por no haber sido admitidos en ellas, o por motivos ideológicos. En lo referido al traslado de domicilio, la UPV/EHU no dispone de una oferta amplia de residencias o colegios mayores, y el acceso a viviendas en alquiler es especialmente costoso en la CAPV. La tercera opción resulta bastante habitual, aunque su motivación suele venir de las limitaciones de matrícula en los centros con más demanda.

El resultado de todo esto es que un elevado número de alumnos recorren todos los días distancias de cientos de Km. En este contexto, la incorporación de un "Campus Virtual" a la UPV/EHU puede ser especialmente útil.

Por otra parte, los nuevos planes de estudio que se han implantado (o se están implantando) en todas las universidades españolas permiten a los alumnos la realización de asignaturas optativas de libre elección. Disponer de una "Facultad Virtual" permitirá a los alumnos de la UPV/EHU acceder a asignaturas de otros centros.

Otra razón más para la implementación de estas tecnologías es la optimización de recursos humanos. En ocasiones resulta difícil alcanzar la "masa crítica" para la realización de ciertas actividades, como por ejemplo la apertura de una asignatura optativa o la impartición de un curso de doctorado. La existencia del "Campus Virtual" facilita sobremanera la consecución de un número apropiado de participantes en una actividad. Incluso es posible llegar a acuerdos con otras Universidades para realización de actividades conjuntas.

3. La red de servicios telemáticos de la UPV/EHU

Desde el punto de vista de la tecnología, nuestra universidad abordó hace ya tiempo, y con recursos abundantes, la instalación de servicios telemáticos. En los años 1990-1993, y por iniciativa de la SPRI, una sociedad pública dependiente del Gobierno de nuestra Comunidad Autónoma, se instaló una red de comunicaciones avanzadas. Esta red corporativa, basada en Ethernet y con una red troncal de tecnologías FDDI y Frame Relay, ha servido también para familiarizar a un elevado colectivo de docentes e investigadores con las aplicaciones telemáticas.

En estos momentos la UPV/EHU se encuentra inmersa en un proceso de actualización de su red corporativa. Mientras que los equipos de los usuarios seguirán estando basados en Ethernet a 10 Mb/s, la red troncal va a cambiar completamente, pasando a tecnología ATM. Los enlaces entre campus pasarán de los 2 Mb/s de las conexiones Frame Relay, a los 34 Mb/s.

Por otra parte, los sistemas de Videoconferencia sobre red IP han sido objeto de análisis y desarrollo en la UPV/EHU. En una reunión de la Comunidad de Trabajo de los Pirineos (Jaca, 1994) este grupo ya presentó una ponencia denominada "Herramientas para Trabajo Cooperativo", describiendo las tecnologías y las primeras aplicaciones existentes. Se han realizado múltiples experiencias piloto, algunas de las cuales están descritas en la sección 5 de este artículo.

4. Sistemas de teleenseñanza

Una solución de telepresencia orientada hacia la Facultad Virtual tiene varias componentes principales, que podemos simplificar en la siguiente relación:

- Infraestructuras y servicios de comunicaciones.
- Terminales de usuario.
- Contenidos.
- Metodología.

Los dos primeros elementos se pueden asociar con las "tecnologías". Existen diversos "modelos", unos con origen en los operadores de telecomunicaciones convencionales (del estilo de las PTTs europeas), y otros derivados de la experiencia de Internet.

Los operadores tradicionales contemplan estos servicios con la misma filosofía que los de voz. Son comunicaciones punto a punto, soportadas inicialmente por la red de voz (RDSI), con una relación coste/prestaciones relativamente alto.

Un ejemplo clásico es el terminal de Videoconferencia telefónica tipo H.320, utilizado por la UNED para mantener reuniones con sus centros asociados (Figura 1).

Utilizando tecnologías más avanzadas, como ATM, hay ejemplos de uso en áreas muy concretas, que económicamente justifican soluciones de este tipo aunque su coste sea elevado. Un ejemplo de ello lo tenemos en el proyecto INSURRECT, para enseñanza de técnicas quirúrgicas. Seis universidades británicas se han asociado para compartir profesores y pacientes. Además encuentran beneficios adicionales: al no estar presentes los alumnos en el quirófano, disminuyen los riesgos de infección para los pacientes.

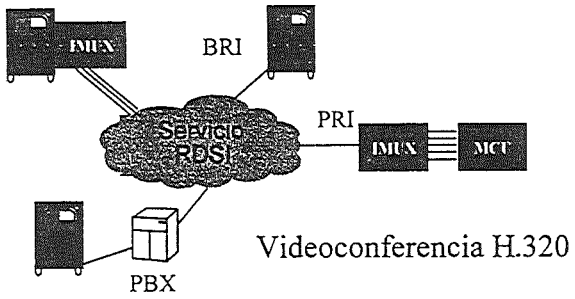


Figura 1. Sistema de videoconferencia basado en la RDSI.

Varias ETS de Ingeniería de Telecomunicación españolas ya están integradas en un sistema de videoconferencia basado en el uso de satélites. El sistema se configura como un aula virtual que integra varias aulas en distintas localizaciones geográficas, donde se reúnen los alumnos con un profesor que se encuentra en el aula presencial desde la que se imparte la clase. Los alumnos, tanto los localizados en la misma aula presencial como los que asisten desde las aulas distantes, pueden realizar preguntas de forma interactiva a través de un micrófono. La arquitectura del sistema consiste básicamente en una red VSAT bidireccional que utiliza el segmento espacial (banda Ku) del satélite Hispasat. Se trata de una arquitectura en estrella desde una estación central (situada en la ETSIT de Madrid) a las estaciones de usuario, con un enlace de salida de 2 Mb/s (desde el aula presencial hacia las remotas) y un número seleccionable de hasta 10 enlaces de entrada de 64 Kb/s (desde las aulas distantes hacia la presencial y entre distantes), que configuran una capacidad de varios cientos de terminales de usuario.

El "modelo Internet", materializado en la red MBone, contempla una filosofía más abierta, similar a los procedimientos de "difusión" utilizados en las emisiones de radio o TV. Está enfocado a una participación en "grupo", utilizando como terminales de usuario equipos informáticos. Es adecuado para redes corporativas, generalmente presentes en universidades. La red telemática de I+D financiada por el Gobierno Español, denominada RedIRIS, propicia este modelo como sistema piloto para las actividades de videoconferencia entre universidades.

También la iniciativa denominada Internet 2 en la que participaban inicialmente 34 universidades americanas, y ahora más de 100, utiliza este modelo para sus proyectos de Campus Virtual, Biblioteca Virtual, Trabajo Cooperativo entre investigadores, etc.

5. Proyecto IKASTEL Facultad Virtual

El Grupo de Redes de Computadores del Dpto. de Arquitectura y Tecnología de Computadores participa desde Septiembre de 1996 en un proyecto PGTI, financiado por el Departamento de Industria, Agricultura y Pesca del Gobierno Vasco, denominado "IKASTEL: Facultad Virtual" (Ref. C148T314). El consorcio está formado por varias empresas (AVA Multimedia, Softec, Tea Cegos, Euskalnet), un centro del EITE (Robotiker), y liderado por la UPV/EHU.

Este proyecto pretende desarrollar un sistema operativo de teleenseñanza, proporcionando aplicaciones y sistemas para teleasistencia conjunta entre grupos de alumnos con distintas ubicaciones. En un paso más, se contempla también la teleasistencia desde el domicilio, utilizando red telefónica conmutada o RDSI, y terminales de usuario basados en computadores personales (ver Figura 2).

La red corporativa de la UPV/EHU proporciona ya la infraestructura de comunicaciones necesaria para realizar experiencias de teleenseñanza. Ahora se trata de sustituir en la medida de lo posible las situaciones presenciales por otras telepresenciales.

Un análisis de las actividades incluidas en una acción docente nos permite fijar los requisitos para los puestos de los usuarios. Tomando como ejemplo una actividad tradicional, como es la lección magistral, podemos establecer los recursos tecnológicos necesarios.

Ya se ha efectuado una primera fase de evaluación, instalación y prueba. Con los equipamientos informáticos disponibles, y las comunicaciones de datos existentes en la UPV/EHU, se han efectuado varias experiencias, participando también en actividades similares de otras organizaciones. Destacan entre ellas:

- Cursos de Verano de la UPV/EHU (edición 1996), incluyendo el curso denominado "Telecomunicaciones y Sociedad: el futuro inmediato".
- "Jornada de Difusión" del proyecto IKASTEL el 18 de Diciembre de 1996, entre Lejona (Biblioteca), Bilbao (ETSII y IT) y San Sebastián (Palacio de Miramar y Facultad de Informática).
- Conferencia de Larry Landweber sobre "Internet II" desde la sede de Fundesco (Madrid, Febrero 1997).
- Curso de Posgrado de la Universidad Rovira i Virgili sobre Educación y Tecnologías (Castellón, 1997).

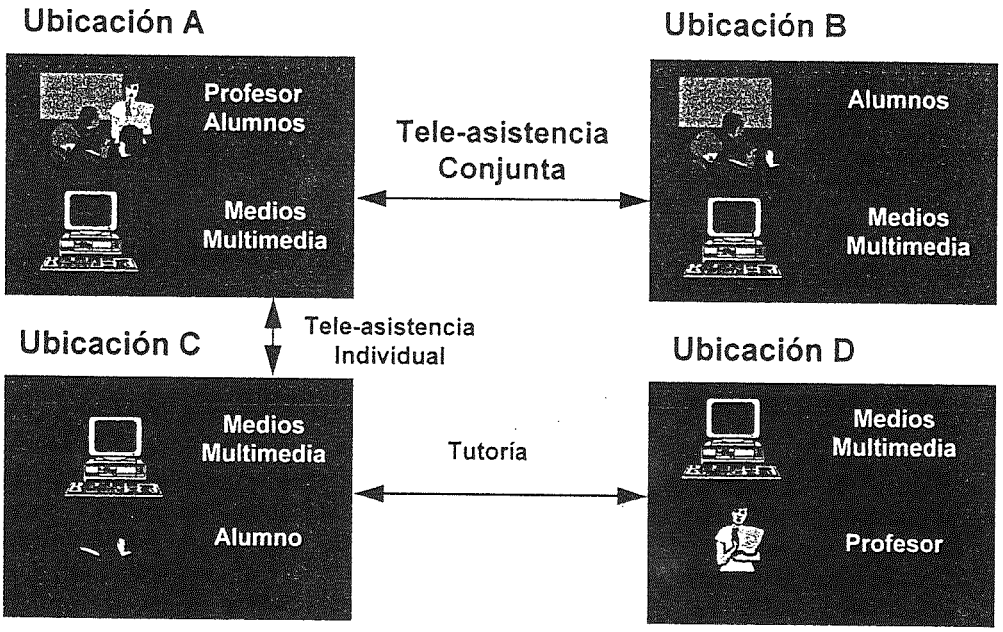


Figura 2. Modelo de teleformación para IKASTEL.

- "I Semana de Psicología e Internet", Facultad de Psicología de la UPV/EHU (San Sebastián, abril 1997). Se realizaron presentaciones con videoconferencia desde Castellón (U. Rovira i Virgili) y Tarragona (U. Jaime I).

Los resultados demuestran las posibilidades de la tecnología, que permite en su estado actual el desarrollo de actividades de teleformación con una calidad suficiente.

6. Aplicaciones de MBONE

Para apreciar las posibilidades de las aplicaciones utilizadas en las experiencias de teleformación, se presentan a continuación unas cuantas figuras obtenidas de sesiones reales. Corresponden a las actividades mencionadas en el apartado anterior.

La Figura 3 corresponde a la aplicación Session Directory (SDR), que realiza funciones de guía de programación para las actividades desarrolladas sobre MBone. Los usuarios emplean SDR para comprobar qué reuniones están en marcha, y para anunciar futuros eventos.

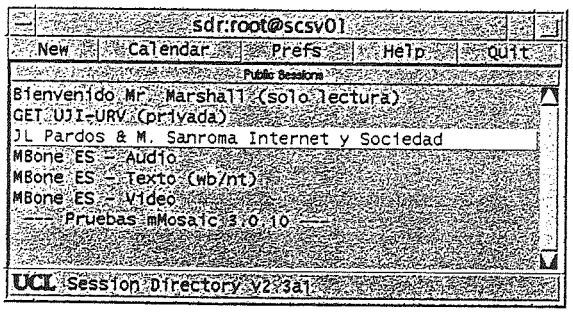


Figura 3. Ventana de la aplicación SDR.

La Figura 4 muestra la aplicación Visual Audio Tool (VAT), que se emplea para emitir/recibir sonido dentro de una sesión MBone. Nótese cómo en la parte izquierda aparece una lista de todos los participantes en la sesión, y en la parte derecha tenemos los niveles del sonido recibido y emitido.

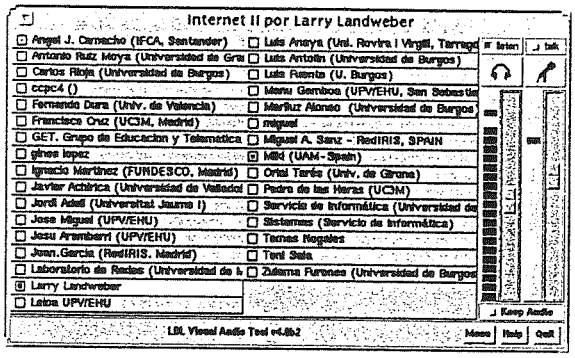


Figura 4. Ventana de la aplicación VAT.

La Figura 5 muestra una de las ventanas de Videoconference Tool (VIC), la aplicación para la emisión y recepción de vídeo. Un usuario podría disponer en su terminal de tantas ventanas como participantes haya en la reunión, o bien conmutar una única ventana entre diferentes fuentes de vídeo.

Otra aplicación especialmente útil es una pizarra electrónica, que tiene una doble finalidad. Por una parte, puede utilizarse de forma análoga al tradicional proyector de transparencias. Por otra, ofrece un espacio de trabajo compartido en el que todos los participantes de una sesión pueden aportar información y comentarios.

Figura 5. Ventana de la aplicación VIC.

Otras aplicaciones en desarrollo son los editores de textos en red, orientados al trabajo cooperativo entre los participantes en el grupo, y los sistemas de votación en tiempo real, para que los participantes en una sesión puedan expresar sus opiniones sobre diferentes aspectos de la misma.

La calidad del servicio de audio, vídeo y docuconferencia en MBone depende básicamente de dos aspectos tecnológicos. Las comunicaciones por una parte son el cuello de botella en la transferencia de información entre los participantes. Las características de la Internet actual, congestionada en muchas ocasiones, impide la viabilidad de sesiones con otros países. Pero también otro recurso que alcanza la saturación es la capacidad de procesamiento en los terminales de usuario, encargados de comprimir y descomprimir sobre la marcha las imágenes intercambiadas. La utilización de sistemas hardware de ayuda a la compresión y descompresión alivia bastante este problema.

7. Equipamiento necesario

El equipamiento necesario para un sistema de teleeducación depende del puesto de trabajo a dotar. Podemos distinguir varios grupos diferenciados:

Un puesto de un alumno a distancia necesita de un computador personal, ampliado con elementos de conectividad (conexión a la RTB, RDSI, etc.) y con una tarjeta de sonido (más altavoces y micrófono). Con esto se puede participar en la recepción de datos, sonido y vídeo. Además, se pueden enviar datos y sonido. Si se considera oportuno, se puede ampliar el equipo con una cámara de vídeo y una tarjeta digitalizadora, para así poder generar vídeo.

Un puesto de un educador necesitará todos los sistemas anteriores, más equipamiento (hardware y software) para la preparación del material multimedia educativo.

Si se trata de dotar una sala para que un grupo de alumnos asista a una clase impartida de forma remota, se necesita un computador (PC o estación de trabajo) con equipamiento de sonido y vídeo de calidad. La imagen generada habrá que enviarla hacia una pantalla de gran tamaño, usando preferiblemente un cañón de vídeo. El sonido de la sala tiene que cuidarse especialmente, conectando el computador a un sistema de megafonía y ubicando los micrófonos de tal forma que se eviten acoplamientos. En lo referido a la captación de imágenes, interesa tener más de una cámara para dar diferentes tomas del auditorio. Por último, en cuanto a la red, dadas las exigencias de calidad de este sistema, es necesario que las salas participantes en una actividad estén interconectadas a través de una red de altas prestaciones.

Las experiencias realizadas por el grupo nos han llevado a concluir que, cuando se realiza una clase a distancia, es importante tener en cuenta estos puntos:

- El alumno debe disponer, con la suficiente antelación, de un "manual" con información sobre la clase que va a recibir. De esta forma se puede realizar un mejor seguimiento de la clase.
- El material utilizado a modo de transparencias, con herramientas como WB, tiene que estar particularmente bien cuidado. Es importante elegir unos tipos de letras apropiados, en forma y tamaño, huyendo de letras pequeñas. Hay que tener en cuenta que las características de la pantalla del emisor pueden ser distintas de las del receptor, y que un gráfico que se visualiza bien en una estación de trabajo puede ser irreconocible en un PC con una pantalla de 640 x 480 puntos.
- El sonido tiene que estar particularmente bien cuidado. La clase puede convertirse en un fracaso si el sonido llega entrecortado, o se producen acoples. Si hay problemas de ancho de banda, es preferible sacrificar el vídeo a cambio de tener un buen sonido.
- El vídeo resulta vistoso, pero no es particularmente útil, en la mayoría de los casos. Esto, claro está, si el vídeo se limita a transmitir el busto del conferenciante. El vídeo no resulta apropiado para mostrar gráficos o texto; aunque se puede usar con este propósito, la pizarra electrónica resulta mucho más apropiada.
- No conviene realizar sesiones de larga duración. Conviene que el conferenciante realice paradas, quizá cada cuarto de hora, para dar a los asistentes la oportunidad de participar. Al mismo tiempo, estas paradas pueden ser útiles

para hacer ajustes técnicos que mejoren la calidad de la clase.

8. Tareas actuales

En esta fase del proyecto, continuamos un estrecho seguimiento de los avances de la tecnología, tanto en el aspecto de las comunicaciones como de los equipos y aplicaciones de usuario.

La red corporativa recientemente contratada por la UPV/EHU a Euskalnet elimina de nuestras preocupaciones cualquier problema relacionado con el ancho de banda de las comunicaciones. Una infraestructura de este tipo tiene capacidad suficiente para transportar numerosas sesiones en paralelo. Queda por concretar la solución para proporcionar acceso doméstico, para el que se conocen algunas alternativas utilizando productos comerciales.

Estamos realizando pruebas exhaustivas de combinaciones de hardware y software para este tipo de aplicaciones. Aunque el objetivo básico es que cualquier ordenador pueda participar en estas actividades (estaciones de trabajo Unix, PCs con Windows o Unix, Macintosh), lo cierto es que el equipamiento básico de estos sistemas no suele incorporar hardware para sonido y vídeo, y es necesario realizar una lista de equipamientos "compatibles".

También se está trabajando en mejorar las aplicaciones, desde dos frentes. Por una parte, hacerlas multiplataforma (operativas sobre varias combinaciones de computadores y sistemas operativos) y, por otra, mejorar y ampliar su funcionalidad.

Pero conviene siempre tener en cuenta que la tecnología es sólo el soporte para las actividades de teleformación, el "continente" o el "medio" que las soporta. Sin dejar de vigilar su continua evolución es conveniente conceder un protagonismo especial a la actividad fundamental y sus resultados: la docencia.

En este momento se ha impulsado desde el Vicerrectorado de Investigación la puesta en marcha de una experiencia piloto más ambiciosa que las realizadas hasta la fecha. Una docena de centros de los diferentes campus de la UPV/EHU ya han mostrado su interés en participar en la experiencia.

Con este fin se ha contemplado la realización de una serie de "cursos piloto" y "pruebas abiertas" que nos proporcionen un mejor conocimiento del nuevo medio de comunicación, y nos permitan establecer los procedimientos y las metodologías para una utilización sencilla y eficaz de la "Facultad Virtual".

Esta tarea no es precisamente tecnológica, sino que requiere la colaboración de otros colectivos. Se precisa la participación de:

- Profesores que preparen su material docente en este nuevo soporte.
- Alumnos que reciban la docencia y evalúen su eficiencia y "usabilidad".
- Pedagogos y expertos en Ciencias de la Educación que analicen el desarrollo de las actividades

9. Conclusiones

Las actividades de teleformación parece que van a jugar en el futuro próximo un importante papel en el ámbito de la educación superior.

Pueden tener un considerable efecto económico, tanto en aspectos de plantilla docente como en los relacionados con los desplazamientos de profesores y alumnos.

También pueden ayudar a deslocalizar y reagrupar colectivos y recursos, haciendo abordables determinadas actividades que necesitan una masa crítica para justificar económicamente su existencia. En esa línea siempre se ha contemplado su influencia en la Gestión del Territorio, por la facilidad para proporcionar servicios en áreas con baja densidad de población.

Aunque la tecnología haya alcanzado ya un cierto grado de madurez, es preciso experimentar con el sistema para establecer la metodología que lo convierta en un servicio eficaz. Es preciso desarrollar metáforas más ricas de las actividades de transmisión de conocimiento, e incorporarlas de manera natural al soporte tecnológico que se está desarrollando.

Referencias

Abusando de la tecnología, indicamos una sóla referencia en WWW, que actualizamos con cierta frecuencia. En ella se encuentra información del proyecto "IKASTEL: Facultad Virtual", mencionado en el texto, junto con numerosos apuntadores a información de otras instituciones.

"http://www.cd.sc.ehu.es/MultiMed"

SimulNet: un entorno de Tele-laboratorios

MARTÍN LLAMAS NISTAL, LUIS ANIDO RIFÓN, MANUEL J. FERNÁNDEZ IGLESIAS
ÁREA DE INGENIERÍA TELEMÁTICA. DEPTO. DE TECNOLOGÍAS DE LAS COMUNICACIONES
E.T.S.I. TELECOMUNICACIÓN. UNIVERSIDADE DE VIGO
CAMPUS UNIVERSITARIO S/N. 36200 VIGO
Correo electrónico: {martin,lanido,manolo}@ait.uvigo.es

Abstract:

In this paper we present SimulNet, a distributed and remote access computer based training system. Simulnet provides a tele-laboratory environment based on Internet and WWW. This is achieved by delivering software through the Internet which can be run on any computer. These distributed applications are simulators of those tools which can be found in a conventional laboratory.

1. Introducción

SimulNet es una plataforma de acceso a tele-laboratorios virtuales basados en el uso de simuladores. Como sistema de teleaprendizaje práctico ha de permitir el acceso remoto a los simuladores, la supervisión de la ejecución de éstos por parte de los profesores, y la comunicación entre los diferentes usuarios que forman el sistema, en especial la comunicación entre profesores y alumnos. El objetivo último es la consecución de un laboratorio virtual en el que sus usuarios no echen de menos ninguna de las características propias de un laboratorio tradicional.

El resto de este artículo se organiza de la siguiente manera: en el apartado 2 se ofrece una perspectiva histórica de los trabajos realizados dentro del Área de Ingeniería Telemática durante estos últimos años, que han servido de referencia y motivación para el presente trabajo. En el apartado 3 se exponen los objetivos perseguidos para a continuación, en el apartado 4, describir la arquitectura del sistema. Finalmente, en el apartado 5 presentamos algunas conclusiones.

2. Antecedentes

El Área de Ingeniería Telemática de la Universidad de Vigo[1] desarrolla sus actividades docentes en la Escuela Técnica Superior de Ingenieros de Telecomunicación[2]. Durante los últimos años se han llevado a cabo en el Área una serie de trabajos relacionados con la creación de entornos pedagógicos basados en computadores. Estos entornos pretendían servir primero como complemento de la asignatura de *Fundamentos de Ordenadores* del plan 1985, y luego, con la implantación del nuevo plan de estudios o plan 1994[3], como soporte para determinadas asignaturas de carácter práctico pertenecientes a dicho plan.

Estas asignaturas, *Fundamentos de Ordenadores I*[4], de 3 créditos teóricos y 1,5 prácticos, y *Laboratorio de Arquitectura de Ordenadores*[5], de 3 créditos prácticos, están relacionadas con la arquitectura de los ordenadores y basan sus programas primordialmente en el texto *Conceptos básicos de arquitectura y sistemas operativos*[1], del profesor Gregorio Fernández Fernández. En este texto se definen y describen ordenadores ficticios con fines didácticos para que el alumno pueda comprender los conceptos fundamentales de la arquitectura de un ordenador, sin perderse en los detalles de implementaciones concretas.

Naturalmente, estos conceptos han de ser explicados en las correspondientes asignaturas teóricas (o partes teóricas de las asignaturas), pero una comprensión plena por parte del alumno requiere ponerlos en práctica, y experimentar con los conocimientos adquiridos.

Esta última faceta ha sido el objetivo de los trabajos anteriormente mencionados, objetivo que ha sido cumplido con plena satisfacción, ya que actualmente los laboratorios de Fundamentos de los Ordenadores I y Laboratorio de Arquitectura de Ordenadores, están basados plenamente en herramientas desarrolladas por el Área.

Estas herramientas permiten simular en ordenadores reales el comportamiento de los ordenadores pedagógicos descritos en [1], y por lo tanto realizar prácticas sobre todos los aspectos de su funcionamiento de forma directa.

Algorítmez [1] es una máquina simplificada que emula a las reales y cuyo objetivo es ayudar a profundizar en la compresión de la función de los ordenadores, en el nivel de máquina convencional, y en su uso para programar. Algorítmez presenta varios modos de direccionamiento de la memoria principal, tiene formatos de instrucciones de longitud variable, y utiliza una pila simulada en la memoria principal para anidar las llamadas a subprogramas y los servicios de interrupciones.

[1] http://www.uvigo.es/
[2] http://www.teleco.uvigo.es/
[3] http://www.teleco.uvigo.es/planes/nuevo.html
[4] http://www.teleco.uvigo.es/programas/p1/pfo1.html
[5] http://www.teleco.uvigo.es/programas/p2/plao.html

IDEA [2] es un entorno de programación para este ordenador ficticio en el que se simulan sus características, presentando una interfaz con el usuario basada en sistemas de menús y ventanas, utilizando la plataforma suministrada por Turbo Vision[3], lo cual facilita la utilización por parte de usuarios no experimentados. Este entorno supuso el primer intento de realización de sistemas pedagógicos de simulación.

En una segunda fase, se han desarrollado entornos de simulación para todos los ordenadores pedagógicos descritos en [1]. El primero de ellos ha sido el de Símplez. Símplez es un ordenador pedagógico que contiene las unidades funcionales básicas en cualquier ordenador y que es utilizado en [1] como soporte para la explicación de los conceptos básicos de la arquitectura de los ordenadores. Este simulador contiene un entorno integrado en el que se ha incluido editor y ensamblador y actualmente es utilizado por los alumnos de primer curso de la E.T.S.I.T. de la Universidad de Vigo en el Laboratorio de Fundamentos de los Ordenadores I.

Como siguiente paso se desarrolló un simulador para Símplez+i^4 [1]. Este ordenador es una modificación de Símplez al que se le añade tres modos de direccionamiento y un esquema de interrupciones sencillo. Su objetivo es introducir los conceptos de direccionamiento de la memoria principal, en sus diferentes modos, e interrupciones. Este entorno de simulación presenta como principal novedad respecto del de Símplez la inclusión de una base de tiempos para su utilización en el manejo de las interrupciones de monitor y teclado.

Finalmente se ha desarrollado un entorno de simulación de Algorítmez (ESAL)[6] que presenta como características añadidas a las del entorno IDEA las siguientes:

1. Posibilidad de configuración de los diferentes modos de direccionamiento, y mnemónicos del ensamblador.

2. Simulación del nivel de micromáquina, con posibilidad de modificar las microinstrucciones y el contenido de la memoria de control, es decir, con posibilidad de microprogramar.. Estas características permiten la emulación de otros ordenadores sobre el propio Algorítmez simulado.

3. Seguimiento de los cronogramas de las diferentes instrucciones y de la evolución de las microórdenes generadas con cada instrucción sobre la propia ruta de datos de Algorítmez.

4. Simulación de entrada/salida. Se ha implementado un módulo de simulación del sistema de interrupciones y de acceso directo a memoria utilizado por Algorítmez. Con ello es posible la simulación de la comunicación con periféricos, como el teclado, el monitor, la impresora, las unidades de disco o el puerto serie.

Este entorno de simulación es utilizado para prácticas de microprogramación, entrada/salida y máquina operativa en el Laboratorio de Arquitectura de Ordenadores, y junto con los de Símplez y Símplez +i^4 en las prácticas de Fundamentos de Ordenadores I.

Los sistemas que acabamos de describir se encuadran dentro de lo que se conoce como sistemas de enseñanza basados en ordenador. Estos sistemas están pensados para ejecutarse en un ordenador sin ninguna conexión con el mundo exterior, excepto obviamente la natural de teclado, pantalla e impresora. Esto significaría que si algún alumno quisiera realizar las prácticas fuera del laboratorio, como por ejemplo en su casa, lo que tendría que hacer es llevar copia del sistema en disquete e instalarlo en su ordenador personal (PC).

El desarrollo y popularización en los últimos tiempos de la telemática, y concretamente de sistemas que utilizan Internet, como puede ser el WWW, además de la difusión de la propia Internet, han hecho posible su aplicación a los sistemas de enseñanza basados en ordenador.

Es evidente que los profesionales de la docencia en estos campos, son los primeros que pueden apreciar las ventajas que suponen los avances tecnológicos aplicados al campo de la enseñanza. Es por ello por lo que suelen colocarse en la vanguardia de su utilización en aplicaciones educativas.

Ejemplo de aplicación de estas tecnologías es el sistema ASTRO[7], que ha sido desarrollado en el Área de Ingeniería Telemática para la Agencia Espacial Europea, ESA[8].

ASTRO [5][6][7] es un sistema de cursos sobre WWW estructurados jerárquicamente como un conjunto de documentos hipermedia. Como valor añadido proporciona funcionalidades de navegación por el curso basada en el estado del alumno. Esta información de estado depende de la interacción previa del estudiante, del modo de navegación elegido (entrenamiento, repaso) y de los resultados

[6] ftp://ftp.ait.uvigo.es/pub/pc/algoritmez

[7] Advanced Software Tools for TRaining Operators

[8] Número de contrato C/11212/94/NL/FM (SC).

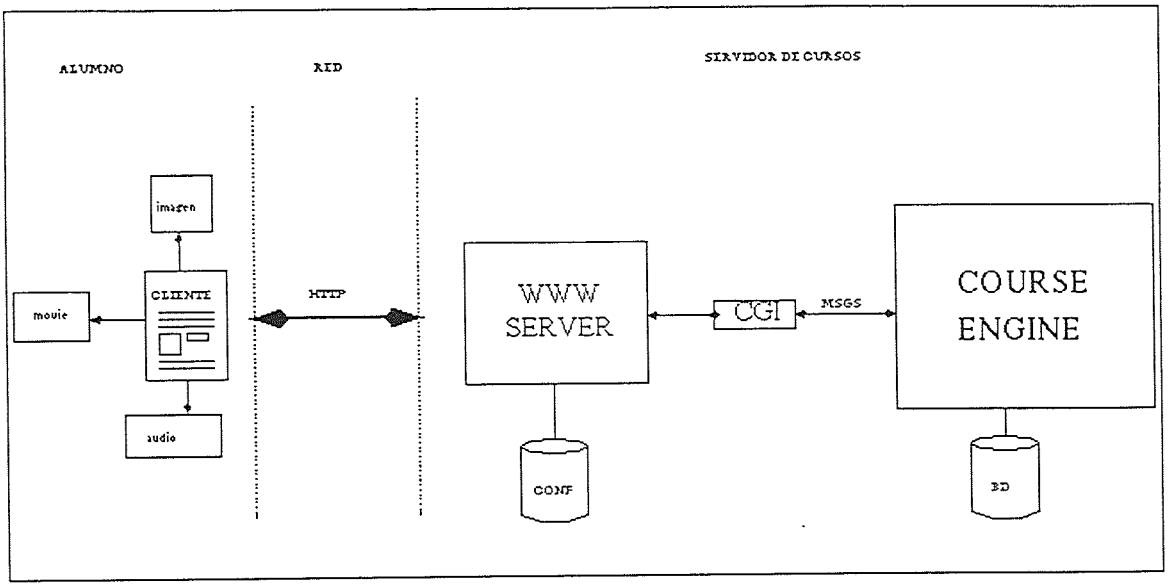


Fig. 1. Arquitectura del Sistema ASTRO.

obtenidos durante la evaluación de sus conocimentos. Además ASTRO permite la autenticación de los usuarios, y proporciona un canal de comunicación entre estudiantes y tutores.

Podemos decir que el concepto fundamental que maneja ASTRO es el mantenimiento del estado de los alumnos entre diferentes sesiones de aprendizaje. Para poder llevar a cabo este objetivo es necesario proporcionar funcionalidades adicionales a las del servidor HTTP empleado como servidor de documentos HTML. Para ello se ha desarrollado una aplicación servidora (Course Engine) que proporciona toda la funcionalidad propia del sistema. En la fig. 1 podemos apreciar una visión general del Sistema ASTRO.

Dentro de ASTRO se reconocen tres agentes diferentes:

1. Creadores de cursos. Encargados de generar todos los elementos utilizados durante el proceso de aprendizaje por parte de los alumnos: generación de gráficos, ficheros de sonido, vídeo, texto, etc., y cualquier otro material necesario. No desempeñan ningún papel pedagógico.

2. Tutores. Desempeñan el papel pedagógico dentro del sistema. Proporcionan la estructura del curso organizándolo a partir del material que le proporciona el creador de los cursos.

3. Alumnos. Utilizan ASTRO para acceder a los cursos a los que tengan acceso. Reciben la ayuda de los tutores.

ASTRO está orientado básicamente a lo que llamamos cursos teóricos organizados en un conjunto de lecciones, y no tanto a servir de soporte a un laboratorio clásico. Es por ello que surgió la idea de poder desarrollar un sistema de tele-laboratios: laboratorios basados en herramientas de simulación como las descritas anteriormente (IDEA, ESAL) pero con las facilidades de teleenseñanza presentes en sistemas como ASTRO. A este sistema de tele-laboratorios lo llamamos SimulNet.

Como paso previo al diseño y desarrollo de SimulNet, hemos realizado un estudio de un conjunto de sistemas de teleenseñanza, con especial atención a aquellos que utilizan Internet y WWW. El 90% de los sistemas a los que tuvimos acceso no contemplan la realización de prácticas de los conocimientos adquiridos a través de los cursos teóricos.

Por otra parte, en un estudio previo a la implantación de un sistema de vídeo interactivo con finalidad docente, Hansen [4] establece que los estudiantes retienen el 25% de lo que leen, pero que son capaces de retener el 70% de aquello en lo que participan de forma directa. Por lo tanto, aparece como imprescindible la incorporación en los sistemas de teleenseñanza de elementos que emulen los laboratorios del sistema docente convencional.

Éstas han sido nuestras principales motivaciones para la puesta en marcha del sistema de tele-laboratorios SimulNet.

3. Objetivos

SimulNet utilizará Internet como elemento soporte. La elección de este entorno se ha basado simplemente en dos requisitos que imponen los potenciales usuarios de este sistema, los estudiantes. Estos dos requisitos son: que sea de fácil acceso y

que sea económico. Internet reúne, en estos momentos, estas dos características.

Para poder ofrecer las funcionalidades docentes típicas de un laboratorio, es necesario incorporar una serie de servicios que minimicen los posibles efectos negativos de la separación física entre profesores y alumnos.

En primer lugar se proporcionará un seguimiento de las acciones que los alumnos realizan en su interacción con los simuladores del laboratorio. Los profesores recibirán información relativa a las acciones realizadas por sus alumnos. De esta forma se puede realizar un seguimiento exhaustivo a aquellos alumnos que se encuentren en dificultades en un momento dado, o recibir, simplemente, información genérica de control de su actividad.

Otro aspecto fundamental es potenciar el aprendizaje cooperativo. La teleenseñanza no debe ser sinónimo en ningún momento de aprendizaje en solitario. Este aspecto es recogido en SimulNet a través de un canal de comunicación entre los usuarios del sistema de laboratorios. Se incorporarán cuatro herramientas diferentes sobre este canal:

- Sistema de correo interno dentro de SimulNet. Completamente análogo al correo electrónico convencional.

- Conversaciones directas entre dos usuarios. Esta herramienta permitirá el intercambio de información en forma de texto.

- Multi-conversaciones. Similar a la herramienta anterior pero con la participación de un grupo de usuarios en lugar de únicamente dos.

- Pizarra virtual. Al igual que la herramienta anterior, ésta permitirá la participación simultánea de varios agentes en una sesión de intercambio de información gráfica. Se dispondrá de un panel compartido por los usuarios participantes. La información introducida por cualquiera de ellos se verá reflejada en el resto de pizarras.

Tanto en la multi-conversación como en la pizarra virtual, existirá un profesor que realice las funciones de moderador. Éste tendrá la posibilidad de establecer el modo de funcionamiento de la comunicación: libre, en la que todos pueden introducir información libremente, o tutor, en la que los alumnos actúan como espectadores ante las explicaciones realizadas por el profesor. También se ofrecerá la funcionalidad necesaria para que los profesores impidan comunicaciones entre usuarios en determinados momentos. Esta característica puede resultar especialmente interesante en aquellas ocasiones en las que los alumnos realizan prácticas objeto de evaluación. El resultado de estas prácticas, al igual que los exámenes en el sistema académico tradicional, puede ser almacenado en un sistema de cuadernos de notas que proporcionará SimulNet.

Además se ofrecerá la posibilidad de definir turnos de laboratorio. Por supuesto, SimulNet no impondrá restricciones de espacio u horarios, como sucede en el sistema docente convencional. Sin embargo, permitirá el establecimiento de un conjunto de turnos de laboratorio, para agrupar a alumnos en torno a profesores. Los alumnos pertenecientes a un turno de laboratorio dirigirán sus preguntas al profesor responsable del turno. De la misma forma, las acciones generadas por ellos en su interacción con los simuladores se enviarán a su tutor responsable. Por otro lado, la existencia de turnos de laboratorio se podrá utilizar para imponer, excepcionalmente, restricciones de acceso al laboratorio.

La participación interactiva de los alumnos en el proceso de aprendizaje se conseguirá gracias a la ejecución de los simuladores en los propios ordenadores de los estudiantes. Esto será posible gracias a la tecnología Java que permite distribuir programas ejecutables a través de Internet.

Finalmente, se proporcionará un módulo de gestión del sistema de laboratorios. Hemos definido la figura del gestor de SimulNet, el cual dispondrá de una aplicación para la realización de tareas de instalación y mantenimiento de todas las estructuras de datos necesarias para el funcionamiento del sistema. A través de una herramienta con interfaz gráfica, podrá dar de alta y/o baja a alumnos y profesores, crear turnos de laboratorio, configurar el comportamiento de los simuladores, etc.

Por medio de este módulo de gestión, será posible agrupar los simuladores que constituyen los laboratorios en grupos. La agrupación de laboratorios puede obedecer a razones de índole organizativa: departamentos universitarios, grupos de investigación, etc. No existirá interferencia entre los simuladores, profesores o alumnos pertenecientes a diferentes grupos de laboratorio, los cuales pueden ser gestionados directamente desde la misma aplicación de gestión. El gestor recibirá información sobre las entradas y salidas de los diferentes usuarios en cada uno de los laboratorios.

4. Arquitectura

A partir de los objetivos definidos en el apartado anterior se está implementando la siguiente arquitectura, para cumplir con dichos objetivos.

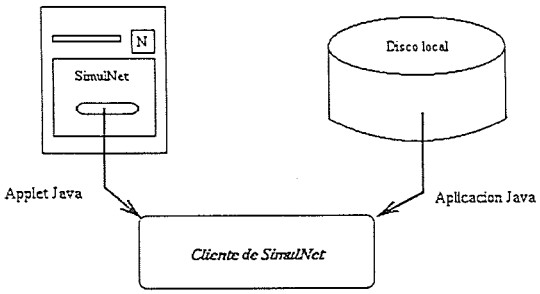


Fig. 2. Modalidades de acceso al sistema

SimulNet utiliza la tecnología empleada actualmente en Internet. El sistema está compuesto por varios módulos que cooperan para implementar todas las funcionalidades definidas. Todos ellos se desarrollan utilizando Java con lo que se asegura la independencia de la plataforma, un alto grado de interactividad y una fácil utilización de los recursos de Internet.

Existen dos subsistemas principales:

El cliente de SimulNet.

Es el elemento que se ejecuta en la máquina cliente utilizada por alumnos o profesores para acceder al sistema. Debido a las restricciones de seguridad[9] impuestas por los desarrolladores de Java es necesario utilizar dos modalidades de acceso para proporcionar toda la funcionalidad definida anteriormente

Estos dos modos de acceso son:

1. Por medio de un navegador WWW comercial en el cliente para presentar el documento HTML, entregado desde el servidor de SimulNet, que contendrá el applet Java de entrada al sistema. Este applet lanza cualquier otra aplicación necesaria que sería entregada por el servidor a través de Internet. El navegador WWW sólo se utiliza como cliente HTTP para acceder al documento base de entrada y el único requisito que debe cumplir es la compatibilidad con Java en su versión 1.0.2[10] o superior.

2. A través de una aplicación Java completamente independiente almacenada en la máquina cliente de profesor o alumno. En este caso los simuladores pueden utilizarse en funcionamiento autónomo. Al mismo tiempo, el tutor o alumno podrá conectarse cuando lo estime oportuno al servidor para beneficiarse de las ventajas del laboratorio virtual: canal de comunicación, seguimiento de alumnos, etc.

Los dos modos de conexión anteriores tienen diferentes ventajas. El primero de ellos, basado en el acceso a través de WWW y ejecución sobre applets, no necesita de la instalación de software adicional en el cliente. Por lo tanto, cualquier conexión comercial proporcionada por las compañías proveedoras de acceso a Internet será suficiente. Sin embargo, debido a restricciones de seguridad en la implementación de Java, el alumno o tutor sólo podría utilizar el almacenamiento de datos de los simuladores en el servidor, no en la máquina cliente.

El segundo modo de acceso supera esta inconveniente gracias a que sí se permite que las aplicaciones Java puedan acceder al disco local, proporcionándose los dos modos de almacenamiento mencionados. El único requisito de esta segunda modalidad de acceso es que el ordenador cliente soporte el kit de desarrollo de Java en su versión 1.0.2 o superior.

Con independencia de la modalidad de acceso y sus implicaciones en el almacenamiento de datos, el funcionamiento es exactamente el mismo en ambos casos y compartirán los mismos objetos Java. Por ello nos referiremos al conjunto de estos módulos de forma genérica como el cliente de SimulNet, tal como se muestra en la figura 2.

Las funciones que realizará el cliente de SimulNet son las siguientes:

- Proporcionar una interfaz gráfica al usuario.
- Permitir la interacción con el simulador.
- Gestionar la comunicación con el servidor.
- Proporcionar acceso a las herramientas de docencia.
- Permitir el seguimiento de las acciones de los alumnos.
- Gestionar el acceso y almacenamiento de datos.

El servidor de SimulNet

Se ejecuta en el computador servidor. Esta compuesto por tres elementos principales según se muestra en la figura 3. El SimulNet Engine, el servidor HTTP y la aplicación de gestión.

El servidor HTTP proporciona al cliente el documento HTML que sirve de puerta de entrada en el caso de acceso a través de WWW, y una conexión de red a nivel de aplicación con el cliente de SimulNet para entregarle todos los objetos Java necesarios en el software que se ejecutará en el lado del alumno o del profesor. El servidor HTTP no desempeñará ningún papel en el caso de los clientes

[9] Los applets ejecutados sobre un navegador WWW no pueden acceder al sistema de ficheros de la máquina cliente.

[10] La incorporación de las sucesivas versiones del DK ha de venir acompañada de la compatibilidad de ellas en los navegadores clientes utilizados.

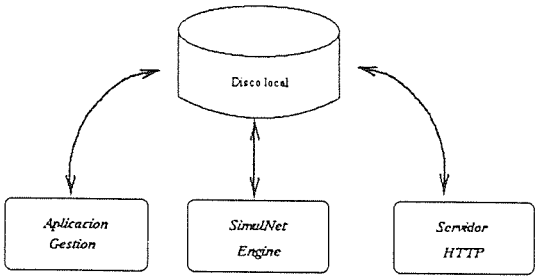


Fig. 3. Elementos en el servidor

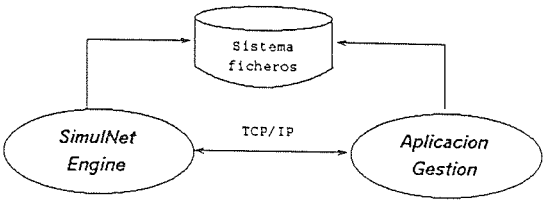


Fig. 4. Comunicación SimulNet-Engine y Aplicación de Gestión

que utilicen como modo de acceso el basado en aplicaciones Java.

El SimulNet Engine es una aplicación Java que soporta la funcionalidad del sistema. Lleva a cabo las órdenes proporcionadas por la aplicación de gestión para implementar el comportamiento indicado. Permite añadir alumnos, profesores, modificar horarios, etc. sin necesidad de parar el sistema y que se vean afectados los usuarios conectados. Esta comunicación entre la aplicación de gestión y el SimulNet Engine se lleva a cabo a través de un doble canal: utilización de sockets TCP/IP y compartición del sistema de ficheros, como se muestra en la figura 4. Gestiona la comunicación entre agentes y se encarga de entregar a los clientes la información almacenada en el servidor y accesible a alumnos y a profesores. Es necesario que sea el SimulNet Engine el que actúe de puente en la comunicación entre clientes ya que no es posible el establecimiento de conexiones de red desde un applet con un host diferente al que ha entregado al cliente el propio applet. Esta característica marca las comunicaciones desde los clientes con acceso a través de WWW y hace necesario su mantenimiento para permitir la comunicación entre éstos y cualquier otro.

Las funciones principales que se llevan a cabo en el servidor son:

- Entrega del software necesario a los clientes con acceso a través de WWW.
- Control de acceso.
- Almacenamiento y entrega a los tutores de la información de seguimiento de los alumnos.
- Gestión de las comunicaciones entre clientes a través de las herramientas de docencia.
- Gestión del almacenamiento de datos de los agentes en el servidor.
- Acceso a los cuadernos de notas por parte de los tutores.

4.1 Comunicación entre subsistemas.

La comunicación entre los dos elementos principales presentados en el apartado anterior: cliente y servidor, se realiza a través de sockets sobre el protocolo TCP/IP. Esa es la base del sistema distribuido que conforma la plataforma: varias aplicaciones Java cooperando por medio de comunicación sobre TCP/IP. Además de ello se utiliza el protocolo HTTP para la entrega a través de la Red del software Java necesario en el cliente con acceso a través de WWW. La comunicación entre los diferentes elementos se ilustra en la figura 5.

El SimulNet Engine proporciona un canal de comunicación virtual entre tutores y alumnos, y en algunos casos entre alumnos únicamente, por medio de un doble canal físico:

- Tutor/Alumno - SimulNet Engine
- SimulNet Engine - Alumno/Tutor

Por lo tanto, como se pretende mostrar en la figura 6, el canal de comunicación sobre TCP/IP entre el SimulNet Engine y los alumnos o tutores proporciona un canal de comunicación virtual directo entre ellos con el propósito de conseguir el objetivo último del laboratorio virtual como se refleja en la parte superior de la figura.

5. Conclusiones

El propósito inicial y más importante de la plataforma SimulNet es la consecución del laboratorio virtual. Sin embargo, es lo suficientemente abierta y flexible como para permitir su utilización con unos objetivos diferentes a la teleenseñanza.

De igual forma que se distribuyen aplicaciones que simulan el comportamiento de una herramienta de laboratorio, se podrían distribuir aplicaciones de una naturaleza muy distinta, con el valor añadido que proporcionan las funcionalidades de seguimiento y comunicación de los agentes. Así, podría crearse un sistema de distribución de software controlado por una entidad, que podría ser el gestor de la plataforma o bien el equivalente a un tutor cuando se utiliza la plataforma con fines de teleenseñanza. La ventaja que se obtendría con este sistema es que los usuarios no tendrían que disponer del software localmente, con el consecuente problema de consumo de disco duro, sino que podrían ejecutarlo directamente a través de un

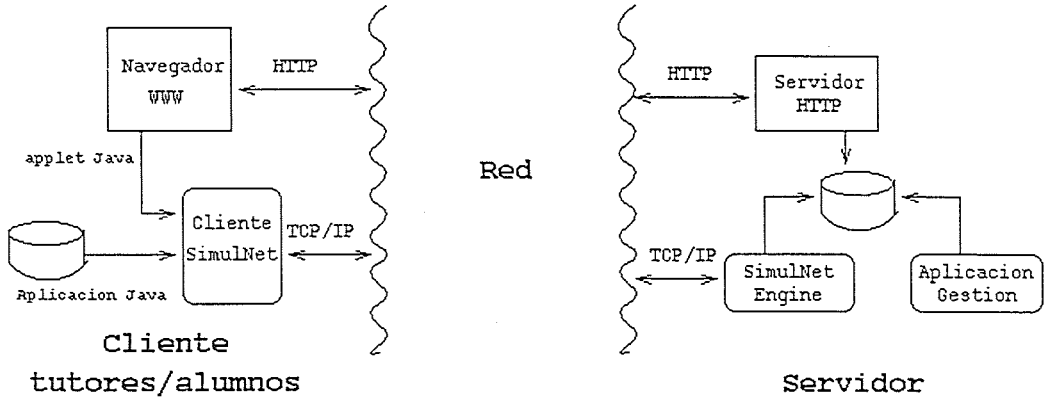


Fig. 5. Comunicación entre los módulos.

navegador de WWW con independencia del tipo de máquina que posean.

Además, el proveedor de este servicio podría efectuar operaciones de control sobre la distribución, como pueden ser: la restricción de acceso de ciertos usuarios a algunos programas o a determinados módulos de programas; el cálculo de estadísticas sobre la utilización de los programas que pone a disposición de los usuarios, con el fin de incluir nuevo software del tipo que más se utiliza eliminar el que no se usa; la monitorización del modo en que los usuarios manejan los programas para poder ayudar a los que se encuentren con problemas e incluso el establecimiento de tarifas en función del software que realmente se esté ejecutando. Las funcionalidades de comunicación implementadas en SimulNet permitirían el intercambio de información entre los distintos usuarios del sistema y el gestor.

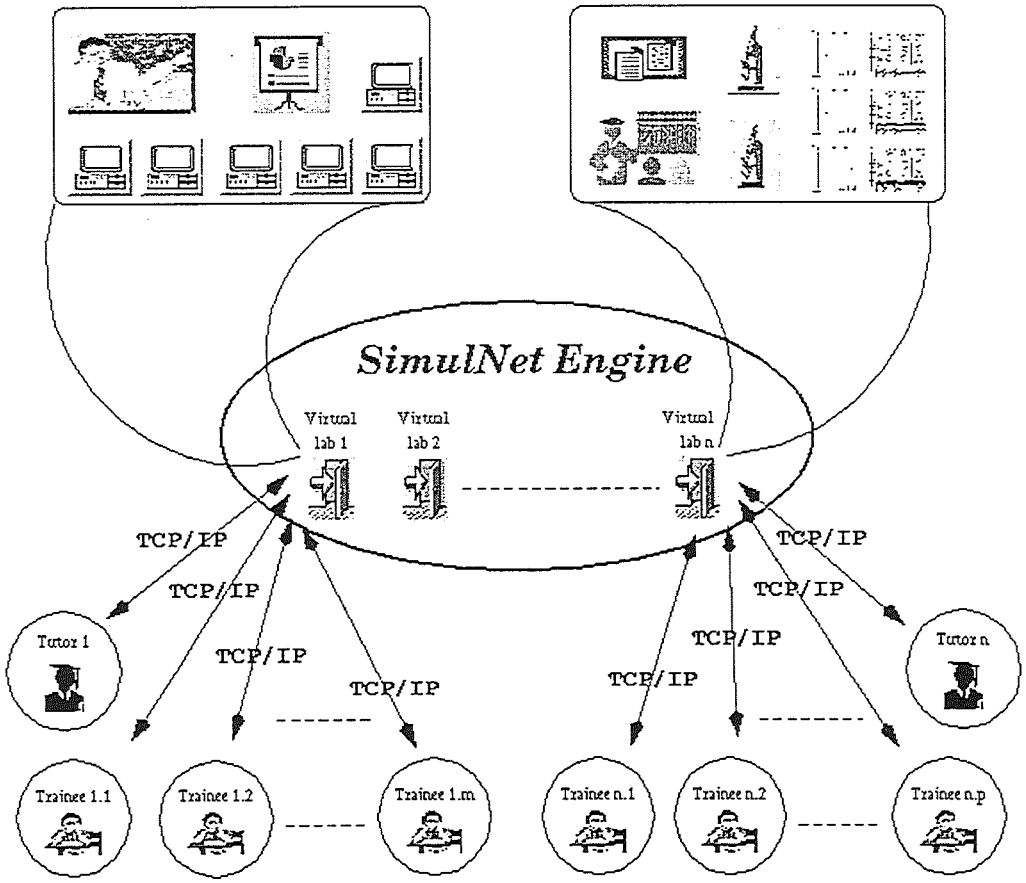


Fig. 6. Arquitectura general de SimulNet.

Otra de las posibles mejoras de las herramientas de docencia existentes en esta versión de SimulNet puede ser la integración de la pizarra virtual y el multi-talk. Es decir, pretendería integrar en una nueva herramienta las funcionalidades ofrecidas por la pizarra y el multi-talk, permitiendo introducir gráficos y texto por parte de los participantes en la sesión. Esta nueva herramienta incrementaría las posibilidades de comunicación entre agentes.

Una línea de trabajo especialmente interesante consiste en la incorporación de la ejecución sincronizada entre tutor y alumnos. Con esto podría seguirse en el cliente de los alumnos la ejecución del simulador por parte del tutor. Esto significaría que las acciones que ejecutara el tutor en su simulador se verían reflejadas en el simulador del alumno.

Esto permitiría un mejor seguimiento de las prácticas por parte de los alumnos y una mejor explicación por parte de los tutores. Por ejemplo, en el entorno ESAL citado anteriormente, podría seguirse la ejecución paso a paso en el simulador del tutor por parte de los alumnos, reflejándose el efecto de las instrucciones sobre la ruta de datos.

También resultaría interesante la incorporación de un sistema de seguimiento de cursos. Debido a la facilidad que ofrece el lenguaje Java para tratar enlaces Web, el sistema SimulNet, además de la aplicación de laboratorios virtuales, podría ser utilizado para el seguimiento de cursos en formato HTML incorporando las funcionalidades necesarias para la realización de exámenes.

Agradecimientos

Este trabajo se inscribe dentro del proyecto "ETLAO: Entorno de TeleLaboratorio para Arquitectura de Ordenadores", el cuál está parcialmente subvencionado por el "Programa de Apoio á Innovación Educativa" de la Universidade de Vigo

Referencias

[1] Fernández Fernández, Gregorio. *Conceptos básicos de arquitecturas y sistemas operativos. Curso de ordenadores.* Madrid: Sistemas y Servicios de Comunicación, S.L. (1994).

[2] Burguillo Rial, J. C.. *Desarrollo de un entorno de programación para Algorítmez.* P/238. Servicio de Biblioteca, E.T.S.I. de Telecomunicación. Universidade de Vigo. (1995).

[3] *Turbo-Vision for C++. Guía del usuario.* Borland International. (1995).

[4] Hansen, E. "The role of interactive video techonlogy in higher education: case study and proposed framework". Education Technology, 13-21. September (1990)

[5] Llamas, M., Fernández M. J., Gil, A., Rodríguez, R. y Suárez, A. "A cost-effective approach to WWW Education and Training". *Protocols for Multimedia Systems.* 191-206. (1996).

[6] Llamas, M., Fernández M. J., Gil, A., Rodríguez, R. y Suárez, A. "Cómo convertir el Web en un entorno educativo" *Novática: Internet Avanzado-II.* 39-44. (1997).

[7] Llamas, M., Fernández M. J., Gil, A., Rodríguez, R. y Suárez, A. "The World Wide Web as a distributed educational environment" *Telematics for future education and training.* EAEEIE. 153-156. (1996)

EL AULA VIRTUAL Y LOS NUEVOS SERVICIOS TELEMÁTICOS: PROYECTO PARA EL DESARROLLO DE UN SISTEMA DE EDUCACIÓN A DISTANCIA

Félix Hernández de Rojas
Rafael Mompó Gómez
Universidad de Valladolid. ETSI Telecomunicación.
C/ real de Burgos s/n
47011 Valladolid
felher@dvnet.es
rafa@tel.uva.es
Alvaro de Miguel Bernáldez
Divisa Informática S.A,
Parque Tecnológico de Boecillo,
47151 Boecillo
alvaro@dvnet.es

Abstract:

ATF or "Asistente Telemático a la Formación" provides a powerful virtual classroom environment where pupils and teacher interact . Flexible learnig is possible using telematic services (WWW, e-mail, news, ftp) and networks in an asyncronous way. Too, teachers may manage their courses, adding pupils, evaluations and multimedia contents. This paper presents a global vision of ATF, their benefits and architecture. Emphasis in placed in telematic services, funcionality and ATF client interface to distributed educational server.

1. Introducción

Las modernas redes telemáticas y herramientas multimedia permiten crear nuevas experiencias en el ámbito de la educación a distancia.

El acceso electrónico remoto y las comunicaciones por red, están convirtiendo el aprendizaje global en un modelo de vida para muchos estudiantes, permitiendo el acceso a múltiples recursos de conocimiento y a otras comunidades de usuarios. Redes internacionales (Internet) amplían los tradicionales campus universitarios, antes constreñidos a los locales de los centros educativos.

Existen múltiples experiencias parciales sobre servicios telemáticos en Internet [1]: redes de formación a través de correo electrónico [2] o espacios (*sites*) Web [3] [4] [5], donde se pueden acceder a los cursos en formato hipertexto. Los contenidos existentes son principalmente tutoriales sobre telecomunicaciones o informática. Sin embargo, en fechas muy próximas, podremos acceder a nuevos cursos orientados a estudiantes en otras materias.

Nuestro interés no es el desarrollo de los contenidos de los cursos. Pedagogos y Psicológos tienen por misión guiar a los tutores en la correcta elaboración de estos. Tampoco es estudiar o diseñar la infraestructura de red necesaria para desarrollar los cursos a distancia. Concretamente nuestra investigación se centra en el diseño y desarrollo de sistemas integradores de los servicios de red, en concreto de aquellos inspirados en las redes TCP/IP actuales.

El aula virtual, como concepto educativo [6] [7] (también el nombre asignado al cliente informático de nuestro proyecto) ha sido diseñado con fin de crear un entorno ficticio, que sin suplantar al aula presencial tradicional, busca superar las desventajas tradicionales de la educación a distancia. No pretende ser un sistema autosuficiente. La experiencia actual demuestra que los entornos de educación a distancia deben ser dirigidos a grupos específicos (discapacitados) o como complemento a cursos tradicionales (apoyo a tutorias o consultas y documentación en línea).

2. Presentación del proyecto ATF

El sistema telemático que aquí se presenta (ATF o Asistente Telemático a la Formación) es un

desarrollo de Divisa Informática S.A., empresa proveedora de servicios telemáticos ubicada en Castilla y León, en el que colabora la Universidad de Valladolid a través de la Escuela Técnica Superior de Ingenieros de Telecomunicación de Valladolid.

Para encontrar una descripción detallada, dirigirse a la página oficial del proyecto: http://atf.dvnet.es o en su defecto a su correo electrónico oficial: sys-atf@dvnet.es.

3. Objetivos fundamentales del proyecto ATF. Principales aportaciones

El objetivo último será superar las clásicas dificultades asociadas a la falta de motivación del alumno y la baja interactividad de los sistemas de educación a distancia tradicionales. El alumno ya no se siente aislado, y cuenta con la colaboración de sus compañeros y tutor.

De esta forma, la interacción entre el usuario y el sistema, se realiza de forma *asíncrona* [8]: El alumno accede al curso en cualquier momento del día. Esto permite mantener una actitud más reflexiva. Debe pensar antes de responder algún examen o cursar alguna petición al profesor.

ATF, básicamente busca integrar los nuevos servicios telemáticos [9] [10], y proporcionar una interfaz común (y racional) de su uso. Esta integración genera otros servicios, más próximos al ámbito educativo y desprovistos de su tradicional dificultad de compresión.

ATF permite realizar formación a distancia en Internet. Pero también en cualquier otra red de ordenadores. Grandes organizaciones con redes privadas (Intranet) o multinacionales necesitan de entornos flexibles y comunes que les permitan mantener una política de formación continua.

Los sistemas tradicionales de educación a distancia, enfocaban principalmente sus esfuerzos en la figura del alumno. ATF, además ofrece al profesor la posibilidad de gestionar el entorno asociado al aula virtual y del curso:

- ✓ Incorporando nuevos alumnos y su información personal.
- ✓ Añadiendo y modificando nuevas unidades de contenidos al curso y las características fundamentales de este.
- ✓ Proponiendo exámenes a realizar por los alumnos.
- ✓ Examinando los progresos de sus alumnos.
- ✓ Publicando eventos de interés general a todos los alumnos.

El sistema está especialmente orientado a entornos donde los contenidos se generan en formato multimedia y se organizan racionalmente en unidades independientes.

El sistema definitivo consta de una interfaz de acceso al curso (seguimiento de los contenidos y evaluaciones) para el caso del alumno. El profesor gestiona el curso, sus alumnos, propone exámenes e incorpora eventos asociados al curso.

La herramienta puede ser incorporada en cursos multimedia, y acceder a servidores remotos o a datos locales, almacenados en CD-ROM.

4. Servicios telemáticos involucrados

Dentro del actual diseño e implementación del sistema se contemplan los siguientes servicios telemáticos:

- World Wide Web
- Correo electrónico
- News (tablón de noticias)
- FTP anónimo

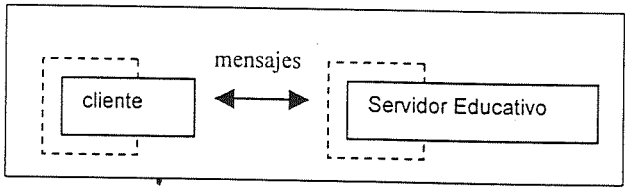
Aunque no se descarta la incorporación futura de otros como herramientas de conferencia, servidores de audio, etc.

Fundamentalmente se desea ofrecer un núcleo básico de servicios telemáticos, sobre los cuales en posteriores desarrollos incorporar otros, más especializados. Un diseño modular del software, permitirá disgregar los servicios en la red y en diferentes plataformas hardware y sistemas operativos.

5. Arquitectura del sistema ATF. Servidor educativo

La arquitectura definida, sigue la estructura tradicional cliente servidor (Fig. 1), aunque esta se encuentra modificada de forma parcial:

Fig. 1. Arquitectura básica del sistema ATF



donde el *Servidor Educativo* es el compendio de todos los servidores asociados, localizados en el proveedor de servicios telemáticos. Un requisito necesario a la hora de diseñar la arquitectura, fue el considerar una

distribución de estos en diferentes máquinas, tal que se permita su futura escalabilidad.

Una relación de servidores involucrados bajo ATF y que componen el Servidor Distribuidp Educativo puede ser esta:

- Servidor Web
- Base de Datos Relacional
- Servidor SMTP
- Servidor POP3
- Servidor News
- Servidor FTP anónimo

La función del servidor Web, es actuar a modo de *Servidor de Aplicación.* El cliente, accede a través del servidor Web a los contenidos del curso y otros módulos adicionales bajo descarga selectiva. También actúa de interfaz de acceso a la base de datos. El cliente consulta directamente la base de datos con el fin de obtener cierta información administrativa.

6. El cliente del sistema ATF

El cliente, denominado aula virtual y desarrollado para sistemas operativos Windows95 y NT, realiza las funciones fundamentales de un navegador tradicional, siendo la interfaz personal de trabajo del alumno y profesor.

Funcionalmente, se descompone en dos áreas. Una sobre la cual se localizan las herramientas del cliente (la parte izquierda de la Fig. 2). El otro área, permite navegar y visualizar los contenidos de los cursos. (la parte derecha de la Fig. 3)

Como aspecto favorable a la interfaz, el alumno nunca asocia los contenidos a direcciones de formato URL (o cualquier otro formato no entendible). De forma más pedagógica, los contenidos se pueden asociar a frases breves (herramienta de favoritos) o a textos personales más largos (herramienta de notas) creados por el alumno.

La interfaz también dispone de una herramienta de correo. Otra herramienta interesante desarrollada fueron las news, aquí bajo la denominación de tablón de anuncios. Estos dos clientes permiten la comunicación directa entre los participantes del curso y su tutor. Al implementar estas herramientas, se buscó la facilidad de uso. No son herramientas avanzadas o complejas, por contra, son altamente intuitivas en su manejo. El usuario muchas veces utiliza un porcentaje muy limitado de la funcionalidad de su correo, ya que el esfuerzo asociado al aprendizaje siempre es alto. El objetivo es no desmotivar al usuario, debiendo aprender a utilizar un entorno complejo, más aún si no se encuentra familiarizado con las redes telemáticas. También, como se advierte en la figura, el atractivo visual se consideró como un elemento importante, que incita al usuario, y lo introduce en el entorno ficticio del aula.

La interfaz de cliente incorpora otras herramientas importantes: Aquellas asociadas al acceso al curso y su gestión. Junto a estas, se dispone de la posibilidad de realizar los exámenes propuestos por el tutor o acceder a los eventos definidos por este (por ejemplo, avisos de clases presenciales)

Sin embargo, el diseño de ATF no busca que su interfaz cliente se restrinja al aula virtual desarrollado. No es pues un desarrollo cerrado. La interfaz puede estar constituida por cualquier navegador comercial. De esta forma, los servicios ofrecidos por el sistema, pueden ser disfrutados por usuarios (al menos en su mayor parte) que no dispongan del cliente del aula virtual. El aula virtual, no obstante, ofrece una normalización de acceso al

Fig. 2. Aspecto de la interfaz de aula virtual

sistema y a los cursos educativos.

7. Experiencias de cursos basadas sobre ATF

Las primeras versiones del sistema (más concretamente, su parte cliente) fueron examinadas por el *Grupo Canalejas*. Para ello, utilizaron sus cursos generados en TOOLBOOK. De esta manera se mostraba la posibilidad adicional de presentar contenidos en otros formatos diferentes a HTML.

El profesor no necesita aprender una herramienta especial predeterminada de autor o utilizar alguna otra, que permitan exportar a formato final HTML. El objetivo final es incorporar el cliente de aula virtual en CD-ROM [11] junto con sus cursos multimedia, y en última instancia todas las funcionalidades del sistema ATF.

Por otro lado, también el sistema ATF pretende ser la herramienta utilizada por el Grupo Canalejas en su intento de acercar la red Internet [12] a las escuelas de Castilla y León. Para ello, en un primer momento se usará como cliente cualquier navegador comercial.

Ya existen programados durante el Otoño de 1997, cursos por instituciones u organizaciones de Castilla y León, que pretenden impartir su formación a distancia, a través del sistema ATF. La experiencia extraída de estos, será fundamental para futuros desarrollos y mejoras del sistema.

8. Conclusiones

ATF aporta las siguientes ventajas a los sistemas tradicionales de educación a distancia:

- Acceso remoto a cursos multimedia y a documentos o referencias
- Control de los progresos del alumno
- Integración de un entorno distribuido de enseñanza

pero también al profesor :

- Nuevas herramientas que permiten mejorar la productividad de la enseñanza

y al alumno:

- Flexibilidad y acomodación a su disponibilidad personal

a pesar de las dificultades encontradas para:

- Desarrollar servicios de alta calidad
- Ofrecer una Interfaz simple, dirigida a usuarios que desconocen los servicios telemáticos.
- Diseñar una Arquitectura del sistema: actuación sobre los servicios (entorno de red) y elección de la política de distribución de las aplicaciones.
- El estado tecnológico actual de las redes de telecomunicación (RTC o RDSI). Los usuarios que acceden a través de modem, ven limitada la velocidad de acceso.

Agradecimientos

ATF ha contado desde el primer momento con el apoyo de la Escuela Técnica Superior de Ingenieros de Telecomunicación de la Universidad de Valladolid y AVADECO (Asociación Vallisoletana de Comerciantes) También hay que resaltar la ayuda prestada por el *Grupo Canalejas* de la Universidad de Salamanca y CEDETEL (Centro Para el Desarrollo de las Telecomunicaciones de Castilla y León). El proyecto ATF es una acción financiada por el Fondo Social Europeo y FORCEM.

Referencias

[1] Juan Alvarez, "Educación a Distancia: Recopilación de Experiencias", DIT-UPM, Julio 1995

[2] Red Tándem Internacional de Correo Electrónico

[3] Espacio Web de la UNED, http://www.uned.es/

[4] Universidad Abierta de Cataluña, http://www.uoc.es/

[5] Open University of UK, http://www.open.ac.uk/

[6] Murray Turoff, " Designing a Virtual Classroom", 1995 International Conference on Computer Assisted Instruction ICCA'95

[7] Starr Roxanne Hiltz, "Teaching in a virtual classroom", 1995 International Conference on Computer Assisted Instruction ICCA'95

[8] Starr Roxanne Hiltz, "Impacts of collegue-level courses via Asyncronous Learning Networks", http://www.njit.edu/njIT, NJIT, Philadelphia, October 1995

[9] Ron Owston,"The Teaching Web: A Guide to de World Wide Web for all Teachers", http://www.edu.yorku.ca/~rowston/chapter.htm

[10] F. Buendía, M. Sanchez, E. Baydal, "Distance learning with the WWW", Departamento de Ingeniería de Sistemas, Computadoras y Automática. Universidad Politécnica de Valencia, Escuela Universitaria de Informática.

[11] Grupo Canalejas, "Proyecto de Cursos multimedia interactivos sobre telecomunicaciones combinando un CD-ROM multimedia e Internet", 1997.

[12] Grupo Canalejas, "Proyecto Internet en las Escuelas para Castilla y Leon", 1997.

Nuevas Metodologías Docentes

Desarrollo de herramientas para gestión de redes basadas en el protocolo SNMP.

Raúl Mata Campos, Ildefonso Ruano Ruano.
Dpto. Electrónica. E.U.P Linares. Universidad de Jaén.
Javier Almendro Sagristá.
E.T.S.I. Telecomunicación. Universidad de Málaga.
Correo electrónico: raul@ait.ujaen.es

Abstract:

This paper describes the development of some tools which permit the implementation of a network management software. Furthermore, this software avoids programming repetitive routines that are common for all management aplications. These tools allow the user to concentrate all his effort on solving management problems, instead of wasting time in message encoding and network communications.

1. Introducción

La gestión de redes se ha convertido hoy en un importante campo de desarrollo de aplicaciones telemáticas, apareciendo diferentes arquitecturas de gestión que tratan de resolver el problema de controlar una red de forma remota. La más extendida de todas ellas está basada en el llamado Simple Network Management Protocol (SNMP) [1], en el cual aparecen elementos gestionables y elementos que realizan la gestión de red. En cada uno de estos elementos reside una entidad, llamada agente para las estaciones que son gestionadas y administrador (manager) para las que efectúan la gestión.

La comunicación entre las entidades se realiza intercambiando mensajes, los cuales se describen utilizando una sintaxis abstracta llamada Abstract Syntax Notation One (ASN.1) [2] y se codifican utilizando la sintaxis de transferencia Basic Encoding Rules (BER) [3]. Una vez construidos los mensajes, éstos se intercambian por la red utilizando el protocolo UDP (User Data Protocol) [4]. La implementación de estas tareas implica un trabajo tedioso y repetitivo, que alarga el tiempo de desarrollo de cualquier aplicación de gestión. El objetivo de este documento es presentar unas herramientas software que realizan todos los procesos comentados, independizando de esta forma la aplicación de gestión de los trabajos relativos a la codificación y descodificación de mensajes, y de su envío y recepción a través de la red.

Estas herramientas se presentan en forma de una librería, que contiene una serie de funciones que pueden ser utilizadas por diferentes programas para permitir el desarrollo de aplicaciones de gestión, enfocadas especialmente para el desarrollo de agentes. Estas funciones, que se describirán más adelante, son el único punto de comunicación entre las herramientas que se describen en este documento y la nueva aplicación de gestión que se esté desarrollando.

En la sección 2 se presenta en detalle la estructura interna que se utiliza para representar los mensajes que intercambian las entidades de gestión, mientras que los procesos de codificación y descodificación se analizan en la sección 3. La descripción de las funciones accesibles de estas herramientas se realiza en la sección 4. Finalmente, un ejemplo de implementación de un agente se presenta en la sección 5, para terminar presentando una serie de conclusiones en la sección 6.

2. Descripción de estructuras internas.

Como ya se ha indicado, en la especificación del protocolo SNMP se especifica que la sintaxis abstracta utilizada para la descripción del protocolo y definición de los mensajes que se intercambian las entidades será la ASN.1 y la sintaxis de transferencia el conjunto de reglas BER.

Sin embargo, desde el punto de vista de la programación, la definición de un mensaje en el lenguaje ASN.1 no es la más adecuada para ser almacenada, ya que habría que hacerlo como si de un texto se tratara. Por ello se utilizará una estructura de datos para la representación interna de los mensajes SNMP. Dicha estructura, que será descrita con posterioridad, almacenará todos los datos de interés que hagan falta para describir un mensaje. Por tanto la codificación de un mensaje partirá de una estructura interna y como resultado se obtendrá la codificación del mensaje según las reglas BER. En el caso de la descodificación, a partir de una serie de

octetos codificados, se obtendrá una estructura de datos del tipo anterior, más fácil de manejar.

2.1 Definición de la estructura mensaje.

Se analizarán en primer lugar las estructuras utilizadas para la representación interna de los mensajes get-request, getnext-request, set-request y get-response, dejando para después el análisis de los traps. Se puede observar en [1] que estas cuatro primeras PDUs (Protocol Data Unit) tienen una estructura común, en la que aparecen la versión, la comunidad, el identificador de petición, indicador e índice de error y una lista de variables. En esta lista de variables, cada elemento está compuesto por el nombre de la variable y el valor.

La estructura en C utilizada para almacenar todos estos valores se define de la siguiente manera:

```
struct mensaje
    {
    un_char          *comunidad;
    int16            lon_com;
    un_char          pdu_cmd;
    int32            req_id;
    int32            error_status;
    int32            error_index;
    un_char          *ber;
    int16            lon_ber;
    struct var_snmp  primero;
    };
```

Puede observarse que los campos de la estructura llamada mensaje se corresponden con los de la PDU de forma intuitiva. Aparecen además una serie de campos auxiliares utilizados para almacenar la longitud de otros campos, además de datos internos necesarios para las tareas de codificación y descodificación.

Para almacenar la lista de variables que contiene cada mensaje SNMP se utiliza una lista enlazada de elementos, cada uno de los cuales representa una variable y su valor. De esta forma, se permite que el número de variables contenidas en cada mensaje no sea una restricción, sino que permite la existencia de tantas variables como sea preciso y utilizando tan sólo la memoria que su almacenamiento exija.

2.2 Definición de la estructura var_snmp.

En este apartado se analiza la estructura interna utilizada para almacenar toda la información necesaria para tener definida una variable SNMP. Hará falta tener almacenado el nombre de la variable y su valor, de forma que éstos serán los dos campos significativos de la estructura. Para la representación de estas dos informaciones, se utilizarán sendas estructuras. Una de ellas contendrá el identificador de objeto (Object Identifier) y la otra su valor.

El identificador de objeto se almacena en formato de números y puntos, es decir, como una secuencia de subidentificadores, de forma que recorriendo el árbol del MIB (Management Information Base) [5] se puede conocer la variable a la que nos estamos refiriendo. Para almacenar el valor de la variable, se utilizará una estructura flexible de forma que permita almacenar diferentes tipos de valores en función del tipo de la variable.

A continuación se presenta la estructura llamada var_snmp, utilizada para almacenar cada una de las variables que componen la lista de variables que intercambian las entidades. La relación entre la estructura var_snmp y la definición de VarBindList en la definición del protocolo [1] es nuevamente inmediata:

```
struct var_snmp
    {
    struct oid       var_oid;
    struct sn_val    var_val;
    struct var_snmp  *var_ant;
    struct var_snmp  *var_sig;
    };
```

Los campos 'var_sig' y 'var_ant' se utilizarán para enlazar los elementos de la lista, apuntando a los elementos adyacentes.

Los campos var_oid y var_val son los utilizados para representar los identificadores de objeto y su valor, ambos de tipos también definidos en estas herramientas. No se analizarán en detalle, por no aportar especial información en cuanto al uso de estas herramientas para la implementación de aplicaciones de gestión.

2.3 Otras consideraciones de interés.

En cuanto a la definición de la estructura interna destinada a almacenar los traps, es bastante parecida a las ya descritas, con la única diferencia de la aparición de campos destinados al almacenamiento de la rama private (enterprise), el time stamp, la dirección del agente que produce el trap, así como su tipo genérico y específico. Las variables se almacenan, al igual que para los otros tipos de PDUs, en una lista de variables.

Definidas de esta forma las estructuras, la misión del módulo de codificación de las

herramientas es la traducción de dichas estructuras a una serie de octetos que contengan el mensaje SNMP codificado según las reglas BER, y de igual forma, la descodificación de una cadena de octetos en una estructura de las ya descritas, que será más fácil de procesar por el resto de las funciones de la aplicación de gestión.

Por tanto, las herramientas presentarán al usuario un interfaz de representación de datos como el descrito hasta el momento, de forma que para enviar un mensaje a la entidad remota, la nueva aplicación de gestión suministrará una estructura como la analizada, y las herramientas procederán a su codificación según las reglas BER. De igual forma, cualquier mensaje cuyo destino sea la aplicación bajo desarrollo, será descodificado y traducido a una estructura de representación interna como la analizada, la cual será ofrecida a la aplicación para que proceda a su tratamiento.

La forma en que se realiza el paso de octetos codificados a estructura interna de las herramientas BER y viceversa se describe en la sección siguiente.

3. Procesos de codificación y descodificación.

Se analizará en primer lugar el proceso de codificación, para describir a continuación la descodificación y su traducción a una estructura interna.

Para el proceso de codificación se parte de una estructura ya creada y con todos los valores de los campos correctamente instanciados. Dicha estructura es recorrida dos veces: la primera para realizar cálculos sobre el número de octetos necesarios para realizar la codificación, y la segunda para realizar el proceso de codificación de la estructura en una secuencia de octetos, la cual será almacenada en el campo reservado a tal efecto. El motivo de este doble recorrido por la estructura se debe a la preferencia de utilizar asignación dinámica de memoria, para no limitar las longitudes de mensajes por el hecho de haber definido longitudes máximas de almacenamiento.

La estructura de codificación para cada elemento del paquete SNMP, según las reglas BER, contiene tres campos: identificador, longitud y contenido. El identificador informa del tipo de la variable que se está almacenando, la longitud muestra los octetos que ocupa el siguiente campo, así como la propia longitud, y por último, el campo contenido almacena la información en sí (nombre la de variable, valor, ...).

El codificador tiene que saber, por tanto, la longitud del campo contenido y del propio campo longitud. Además, debido al anidamiento de estructuras, el número de octetos utilizados para almacenar los tres campos de un elemento codificado, se convertirán en campo contenido de otra estructura que lo englobe. Debido a esto, cuando estas longitudes van siendo calculadas, el proceso se realiza desde las estructuras más internas hacia el exterior. Finalmente se obtiene la longitud que tendrá la cadena codificada. Dicha longitud es el punto de partida de la fase siguiente.

En esta segunda fase el proceso de codificación continúa situándose en las estructuras más internas y procediendo a su codificación. Como ejemplo, se describirá el proceso que sufre una variable SNMP durante su codificación, en concreto, la variable sysDescr (1.3.6.1.2.1.1.1.0) conteniendo el valor "SNMPD v2.1 (9.3.1) IBM-PC MS-DOS FTP Software".

Para proceder a su codificación, primero se codificará el valor de la variable, a continuación su identificador y por último la estructura SEQUENCE que define la estructura variable SNMP. La codificación del valor de la variable se realiza en tres pasos: codificación del contenido de la variable, es decir, la cadena de caracteres que contiene el texto indicado, a continuación la codificación de la longitud del campo anterior (almacenará la longitud de 47) y por último el identificador del tipo de dato, en este caso un DisplayString. El mismo proceso se seguirá para almacenar el identificador de objeto: primero se codificará el contenido, que será la lista de subidentificadores que lo componen, a continuación se procederá a la codificación de la longitud y por último se indicará que se trata de un elemento del tipo OBJECT IDENTIFIER. Tan sólo queda codificar el SEQUENCE, donde el contenido será toda la cadena codificada hasta ahora y el número de octetos que ocupe será almacenado en el campo longitud. La codificación de la etiqueta del SEQUENCE finalizará el proceso de codificación de la variable.

Como puede comprobarse, la tarea de codificación es compleja y repetitiva desde el punto de vista de programación, aspecto que justifica la necesidad de las herramientas que se describen en este documento. El resultado final de todo este proceso será una serie de octetos, los cuales ya están listos para ser enviados a la entidad de gestión remota.

Para realizar la descodificación, se toma como entrada un paquete codificado según las reglas BER, es decir, una secuencia de octetos, los interpreta y obtiene como salida una estructura

interna de representación de paquetes del tipo mensaje o trap, ya descritas con anterioridad.

En este caso, la forma de operar es desde el principio de la secuencia de octetos hasta el final, comprobando que cada uno de los campos que van apareciendo son correctos, y traduciéndolos a la estructura interna que corresponda según sea el elemento bajo estudio. De nuevo se realiza una asignación dinámica de memoria conforme se van analizando campos que requieren un almacenamiento.

El resultado final que se obtiene de esta fase de descodificación es una estructura interna, la cual es devuelta por el descodificador para que sea procesada por las rutinas de gestión adecuadas.

Este módulo de codificador-descodificador es el más complicado y extenso desde el punto de vista de programación, ya que además de realizar la traducción de gran cantidad de tipos diferentes, realiza un exhaustivo control de errores que se hubieran podido producir en la transmisión o codificación de la entidad remota. Este control asegura una gran robustez en el código, que detecta los mensajes erróneos, descartándolos, y sólo procesa los correctos. Además se ocupa de liberar la memoria dinámica que ya no sea necesario mantener reservada.

Una característica importante de este codificador es la no utilización de la sintaxis ASN.1 como paso intermedio entre las secuencias de octetos y las estructuras internas, tal y como ocurre en otras implementaciones de codificadores BER. De esta forma se obtiene una mayor velocidad de proceso, ya que se evita un paso intermedio, que además es bastante costoso en tiempo de computación.

4. Funciones de utilidades de gestión.

A continuación se presentan las funciones que pueden ser utilizadas para la implementación de una aplicación de gestión. Como ya se ha comentado, estas funciones, además de realizar las tareas de codificación y descodificación, permiten independizar a la aplicación de las tareas de envío y recepción de paquetes a través de la red. A continuación se analizan esta funciones, mostrando especial interés en su utilización, aunque también se analizará brevemente su implementación. Primero se describirán unas funciones de inicialización, posteriormente se analizarán las funciones que permiten enviar los mensajes a la entidad de gestión remota y recibir los provenientes de éstas.

4.1 Funciones de inicialización.

Se trata de dos funciones que tienen como objetivo inicializar las comunicaciones de red. Cada una de las funciones inicializará uno de los puertos reservados: el puerto 161 para los intercambios de comandos get, set y getnext, y el puerto 162 para los traps.

El objetivo de estas funciones es definir todas las estructuras necesarias para el intercambio de mensajería a través de sockets, de forma que su utilización queda restringida al inicio del programa de aplicación de gestión. Son pues funciones bastantes simples, aunque presentan una serie de características que resultan interesantes de analizar.

Una opción importante de estas funciones es la posibilidad de configurar el socket que se utilice para la comunicaciones como bloqueante o no bloqueante. En la opción de bloqueante, las operaciones de lectura o escritura que se realicen en el socket dejarán bloqueado el proceso hasta que esta operación se complete con éxito (o se produzca un error). El caso contrario se presenta si el socket se configura como no bloquante, devolviéndose el control al proceso que realice el acceso al socket si la operación de lectura o escritura no logra realizarse correctamente.

Esta doble posibilidad de configuración del socket permite que estas funciones de utilización puedan ser utilizadas por el agente o por el manager. En general, el agente seleccionará la opción de socket bloqueante, pudiendo quedar el proceso en espera de la llegada de un comando SNMP, mientras que el manager usualmente comprueba periódicamente los puertos de entrada para comprobar la llegada de mensajes, pero no quedando bloqueado hasta que se produzca una recepción. En cualquier caso, será decisión del programador la elección de un tipo u otro de modalidad, dependiendo de las necesidades concretas de la aplicación.

Una vez inicializados los puertos, estas funciones devuelven un descriptor de socket, el cual será utilizado cada vez que se realice una lectura o escritura en estos puertos.

4.2 Funciones de recepción de mensajes SNMP.

Las tareas de recepción de un paquete por uno de los puertos reservados y su posterior descodificación es el objetivo de las funciones que se describen a continuación. Desde un punto de vista conceptual, las tareas que desempeñan estas dos funciones son muy similares: lectura de un puerto, a partir del descriptor de socket, descodificación del

mensaje SNMP que contiene y comprobación de errores. La descodificación dependerá del tipo de mensaje que se reciba, o lo que es lo mismo, depende del puerto del que provenga el paquete. Así pues, se devolverán diferentes estructuras de representación interna, motivo por el que se diferencian las tareas de recepción de mensajes en dos funciones distintas.

Desde el punto de vista de la programación, la función realiza dos tareas completamente diferenciadas: la lectura de un puerto, a través de un descriptor de socket y la descodificación del paquete utilizando el módulo de descodificación analizado en apartados anteriores.

La utilización de esta función por parte de una aplicación de gestión es muy simple, aunque las tareas que realice internamente sean complicadas. Habrá que indicar a la función simplemente el descriptor de socket que se obtuvo con anterioridad, y que lo ligaba al puerto correspondiente. De esta forma la función capturará del puerto adecuado el posible paquete que se haya recibido, procederá a su descodificación y traducción a una estructura interna, la cual será devuelta para ser analizada por la aplicación de gestión.

Puede notarse que con una sola llamada a una de las funciones descritas se logra tener una estructura lista para ser procesada por la rutina de gestión adecuada, de forma que todo el trabajo de programación que se realice estará orientado a la consecución de una buena aplicación de gestión, y no a tareas de implementación de especificaciones de representación de un protocolo.

Estas funciones de recepción de mensajes, además de devolver la ya comentada estructura interna, ofrecen la posibilidad de conocer la dirección IP de la entidad de gestión que envía el paquete. Esta facilidad, además de ser interesante desde el punto de vista de comprobaciones de seguridad, es especialmente útil cuando se programan agentes que permiten ser gestionados por diferentes managers o cualquier otro tipo de aplicaciones en las que sea interesante conocer la identidad de la entidad de gestión remota. Esta información suplementaria que ofrecen las funciones, puede ser almacenada para su posterior utilización, o desechada si se estima conveniente por no proceder su posterior análisis.

Una última consideración de interés es el análisis del comportamiento de la función dependiendo del tipo de socket declarado, es decir, que si la inicialización del socket lo definió como bloqueante, el proceso que realice una llamada a esta función, quedará suspendido hasta que se realice la recepción de un mensaje por el puerto bajo consideración. Este es el caso típico de un agente, o un proceso que realice exclusivamente la recepción de traps en la aplicación del manager. Si por el contrario la inicialización del socket lo definió como no bloqueante, el proceso no quedará suspendido hasta la recepción de un mensaje, sino que procederá a finalizar la ejecución de la función indicando la imposibilidad de realizar una lectura en el puerto señalado. Este es el caso típico de un manager, que no debe quedar suspendido esperando la respuesta de un agente, que por otro lado puede llegar a no producirse por pérdida del paquete o cualquier otra causa.

4.3 Funciones de envío de mensajes SNMP.

Las funciones que se analizan a continuación son las complementarias de las analizadas en el apartado 4.2. Se trata de funciones cuyo objetivo es la codificación y envío de un mensaje SNMP a la entidad remota. Las tareas que tiene que realizar en este caso son las mismas que las analizadas para las funciones de recepción de mensajes, pero en orden inverso, es decir, que en primer lugar se procederá a la codificación de la estructura de representación interna y posteriormente se enviará el resultado de la codificación al destino correspondiente, utilizando para ello el socket inicializado a tal efecto.

Al igual que para el caso anterior, se define una función para cada uno de los puertos reservados en la definición del protocolo. Para ambas funciones, la información suministrada será la estructura interna que posee la información que se desea enviar a la entidad de gestión remota, el descriptor de socket a través del cual se va a realizar el envío del paquete hacia la red, y la dirección destino de este paquete.

El origen de la estructura de representación interna que va a ser objeto de codificación puede ser de dos tipos: creada por la propia aplicación de gestión, o la misma devuelta por una función de recepción de mensajes SNMP con modificaciones en alguno de sus campos. El primer caso es típico de una implementación de un manager, en que la estructura es creada por la aplicación de gestión que se está desarrollando. También es el caso de un agente en lo relativo a los traps, ya que es el propio agente el que deberá crear la estructura interna que desea enviar. El segundo caso corresponde a la implementación de un agente que devuelve las peticiones que le realiza el manager. La estructura interna podrá ser la misma que se obtuvo en la llamada a la función de recepción de mensajes SNMP, después de haber modificado los campos adecuados que transformen a la estructura en un

mensaje válido para ser codificado y enviado a la entidad remota.

Con la utilización de estas funciones, también se consigue que la simple llamada a una función realice los procesos de codificación y envío de un mensaje a la entidad remota, de forma que la tarea de la aplicación de gestión es determinar cual debe ser el contenido de la estructura interna, sin preocuparse de las tareas rutinarias de codificación y envío a la red.

5. Ejemplo de implementación de un agente.

Una posible utilización de estas herramientas es la implementación de un agente SNMP. En concreto, cuando se desarrolla este tipo de aplicación, lo que resulta de interés es poder hacer la propia definición del MIB del agente, que estará ligado ciertos a parámetros que se deseen controlar de un dispositivo, por ejemplo una placa. Para poder realizar dicha aplicación, el programador solo tendrá que desarrollar una función que para cada uno de las variables contenidas en el mensaje SNMP, realice la operación de lectura o escritura adecuada. Esta operación quedará oculta desde el punto de vista de la aplicación que se propone como ejemplo, ya que la forma en que almacene las variables del MIB y el modo en que se realice el acceso a ellas es dependiente de la aplicación de gestión.

Lo que se presenta a continuación es un fragmento de la típica estructura del código que tendría una implementación de un agente que utilizara las herramientas descritas:

```
        ...
struct mensaje *mens;
struct var_snmp *actual;
        ...

 /* INICIALIZACIÓN DE RED */
dso=iniciar_puerto161(BLOQUEANTE);
        ...

/* BUCLE INFINITO */
 for(;;)
  {
  /* RECEPCIÓN DE MENSAJE */
  mens=rec_comando(dso,direccion);
  if(mens!=NULL)
    {
   obtener_comunidad(mens->comunidad);
        ...

/* PROCESAMIENTO DE VARIABLES */
   actual=mens->primero;
   while(actual!=NULL)
     {
      procesar_variable(mens->actual,mens->pdu_cmd);
      actual=actual->var_sig;
        ...
    }
        ...

/* MODIFICACIÓN DE INDICADORES DE ERROR: ERROR
INDEX, ERROR STATUS ...*/
modificar_indicadores_error(mens);
        ...

/* MODIFICACIÓN DE TIPO DE PDU */
 mens->pdu_cmd=RESPONSE_CMD;
        ...

/* ENVÍO DE LA RESPUESTA */
 if(enviar_respuesta(dso,mens,direccion)<0)
   {
   fprintf(stderr,"\nError al enviar respuesta.\n");
   }
        ....

 }

/*FIN BUCLE INFINITO*/

        ...
```

En el ejemplo de código propuesto, en primer lugar habrá de definir una serie de variables de acuerdo a los tipos de datos definidos en estas herramientas. Independientemente de las inicializaciones particulares que sean necesarias desarrollar en el agente, habrá que inicializar el puerto 161 que es por donde va a realizarse la recepción de mensajes SNMP provenientes del manager. La llamada a la función iniciar_puerto161() realiza dicha operación, indicando además la opción de socket bloqueante.

En el caso general de entrar el agente en un bucle que realice infinitos procesos de recepción y devolución de mensajes, en cada una de las iteraciones se realizarán los siguientes pasos:

- Recepción de un mensaje SNMP por el puerto 161 mediante la función rec_comando(). Esta función devolverá la estructura ya creada, conteniendo el mensaje descodificado en formato de estructura interna, y además suministrará la dirección del manager que ha realizado el envío del paquete.

- Obtención de datos generales que pudieran ser de interés, tales como la comunidad, dirección del manager, etc.

- Procesamiento de cada variable que aparezca en la lista. Se tratará de una función a la que podrían pasársele como argumentos la propia variable y el tipo de comando SNMP (get, set o getnext). En esta función se modificarían los campos valor de cada variable si resultara procedente. Es la

función que englobaría el núcleo de gestión.

- Modificación de los campos indicativos de error si resultara necesario y cambio del tipo de mensaje para el envío de la repuesta.

- Envío del mensaje respuesta al manager, indicándole la estructura que debe codificar y la dirección. En este caso, como ya se comentó, será la misma estructura interna que devolvió la función rec_comando(), aunque deberá presentar algunos campos modificados.

Todo el proceso de recepción de red, descodificación, codificación y envío a la red ha quedado reducido a llamadas a unas funciones, que en total representan unas pocas líneas de código. Lo único que queda por programar serán las rutinas típicas de comprobación de existencia de variables en el MIB que tenga definido el agente, así como la obtención o instanciación de sus valores. Estos aspectos serán los que distingan una aplicación de gestión de otra.

6. Conclusiones

A lo largo del documento se han descrito las herramientas que facilitan la implementación software de gestión de redes, utilizando el protocolo SNMP. El objetivo de independizar al programador de la codificación y descodificación de mensajes, así como de los problemas de acceso a la red se consigue gracias a la librería de funciones propuesta.

Esta librería conforma un conjunto compacto de funciones, de fácil utilización, y que concentra en una sola función todos los procesos de codificación y acceso a la red.

Se puede conseguir desarrollar una nueva aplicación de gestión en poco tiempo, ya que tan sólo deberá preocuparse el programador de las tareas propias de la aplicación de gestión. Utilizando estas herramientas, se ha desarrollado una estación de gestión (manager) con monitorización programable de dispositivos y diferentes tipos agentes, con MIBs específicos definidos para cada uno de ellos.

Por último señalar que el módulo del codificador-descodificador es portable a diferentes máquinas, y de hecho ha sido utilizado bajo sistemas operativos distintos. No ocurre los mismo con el módulo de comunicaciones con la red, que sí es dependiente de la máquina, aunque la adaptación puede realizarse sin excesiva dificultad.

Referencias

[1] Case, J., Fedor, M., Schoffstall, M., y J. Davin,. "Simple Network Management Protocol (SNMP)", RFC 1157. SNMP Research, Performance Systems International, Performance Systems International y MIT Laboratory for Computer Science. Mayo 1990.

[2] Information processing systems - Open Systems Interconnection - "Specification of Abstract Syntax Notation One (ASN.1)". International Organization for Standardization, International Standard 8824, Diciembre 1987.

[3] Information processing systems - Open Systems Interconnection - "Specification of Basic Encoding Rules for Abstract Notation One (ASN.1)". International Organization for Standardization, International Standard 8825, Diciembre 1987.

[4] Postel, J., "User Datagram Protocol", RFC 768, USC/Information Sciences Institute, Noviembre 1980.

[5] McCloghrie, K., y M. Rose, "Management Information Base for Network Management of TCP/IP-based Internets", RFC 1213, Hughes LAN Systems, Inc. y Performance Systems International, Marzo 1991.

Desarrollo de un entorno práctico para el aprendizaje de gestión de redes mediante SNMP

ROBERT MASCARELL CATALÀ, JULIO MIRÓ BORRÁS
DEPARTAMENTO DE COMUNICACIONES
ESCUELA POLITÉCNICA SUPERIOR DE ALCOY
UNIVERSIDAD POLITÉCNICA DE VALENCIA
C/ Capellá Belloch nº 6, 03801 Alcoy (Alicante)

Abstract:

This paper describes the implementation of a PC based managemet station intended for controlling and monitorig a TCP/IP network, using the Simple Network Management Protocol. Also the article presents a simple simulated agent. It will be necessary in case of a real SNMP agent is not present in the network. This system has be recently developed and will be used only for training purposes.

1. Introducción

En la actualidad, las redes de ordenadores han crecido de tal forma que es imposible que nunca falle nada. Además del tamaño, hay que tener en cuenta que la red puede estar geográficamente dispersa y tener componentes de distintos fabricantes. Esto plantea unas dificultades que deben ser abordadas por los administradores de las mismas. Aparece entonces el concepto de gestión de red.

Para facilitar las tareas de gestión de red existen en el mercado una serie de productos para descubrir y solucionar la mayor parte de los posibles problemas de la red e incluso anticiparse a los mismos. Los principales requerimientos que se le deben exigir a estos programas son los siguientes:

- Mínimo tiempo de parada de la red: máxima productividad.
- Costos reducidos de la red asociados al uso de ancho de banda en las redes de área amplia.
- Fácil manejo tanto por parte de ingenieros de la red como administradores de negocio no técnicos.
- Máxima seguridad de la red hasta el nivel de usuario.
- Contabilidad del uso de la red para evitar una mala utilización del ancho de banda.
- Posibilidad de ampliación para mantenerse al día con las nuevas tecnologías y con una infraestructura creciente.
- Flexibilidad para reconfigurar la red cuando y como sea necesario.
- Compatibilidad con los componentes de red de diferentes fabricantes que están ya instalados y con futuros dispositivos que cumplan con las normas.
- Administración de sistemas y aplicaciones altamente integradas.

Conviene observar que la caída de una red informática, o parte de la misma, va a producir una gran pérdida de beneficios por improductividad de sus empleados.

Teniendo esto en cuenta, han aparecido diversos estándares para la gestión de redes. Los más importantes son CMIP para las redes que sigan las recomendaciones OSI y SNMP para las que utilicen TCP/IP, siendo este último el utilizado en el presente artículo.

Aunque, como se indica en su acrónimo, SNMP es un protocolo "sencillo", su aprendizaje teórico resulta bastante complejo debido a la abstracción introducida por el lenguaje ASN.1 en la definición de las bases de información de gestión (MIB). También resulta difícil de imaginar como se puede realizar una gestión efectiva con tan pocas operaciones como define el protocolo. Resulta obvia la conveniencia de disponer de una serie de herramientas que permitan entender el funcionamiento del protocolo, de las MIBs y de los actores que intervienen en la gestión: el gestor y el agente.

En este artículo se describe el desarrollo de una estación de gestión SNMP sobre entorno Windows, con fines didácticos, que permite monitorizar y controlar cualquier nodo de red equipado con SNMP. También se ha desarrollado un agente SNMP con una mib estática, que podrá ser empleado para el estudio del citado protocolo, y un conjunto de programas para la generación de operaciones simples.

Se trata de una herramienta muy útil para la realización de prácticas en las ocasiones en las que no se disponga de ninguna estación de gestión comercial. El sistema permite descubrir automáticamente y representar en la ventana de trabajo los agentes SNMP situados en la red, interrogarlos para averiguar su estado empleando operaciones *Get* y *GetNext*, modificar el valor de algún objeto de su base de datos de gestión mediante

la operación *Set* y recibir notificaciones (*traps*) cuando suceda algún evento extraordinario en el agente.

Gracias al diseño personalizado de plantillas, es posible seleccionar los objetos que se desean consultar y presentarlos en una misma pantalla. Esto resulta muy útil para la visualización de tablas, pues permite ver de forma tabular todas las instancias de los objetos que se seleccionen de una misma tabla. El uso de la plantilla para personalizar los objetos a monitorizar resulta idóneo para la exploración intuitiva de tablas en los agentes, facilitando el entendimiento de cualquier mib estándar (mib-2, rmon,...) o propietaria.

Algunas características del sistema son la detección del tipo de nodo (host, router, bridge, ...) y su representación con diferentes iconos, posibilidad de incorporar nuevas definiciones de mib, ajuste del periodo de sondeo y particularización del nombre de comunidad para cada agente.

En la próxima sección se introducen los conceptos básicos de SNMP. En la sección tercera se verán los diferentes elementos del entorno docente, los objetivos perseguidos y el entorno de desarrollo utilizado. En la sección cuarta se describen las herramientas implementadas, y se abordarán una a una las diferentes aplicaciones, mostrándose algunos ejemplos de su utilización. Para terminar, se expondrán una serie de conclusiones finales.

2. Breve descripción de SNMP

2.1. Conceptos generales

El término PROTOCOLO SENCILLO DE GESTIÓN DE REDES (SNMP) hace referencia a un conjunto de especificaciones para la gestión de redes, como son el protocolo en sí, la definición de una base de datos y conceptos asociados.

2.2. Arquitectura de gestión de redes. SNMP

El modelo de gestión de redes utilizado por la gestión de redes TCP/IP contiene los siguientes elementos clave:

- Estación de Gestión.
- Agente de Gestión.
- Base de Gestión de Información (MIB).
- Protocolo de Gestión de redes.

La estación de gestión sirve como interfaz entre el gestor de redes (humano) y el sistema de gestión de red. En la estación de gestión reside un software denominado SNMP Manager. Se trata de una entidad que puede preguntar a los agentes utilizando operaciones SNMP. Además, proporciona un interfaz -generalmente gráfico- que permite a los usuarios pedir datos o visualizar alarmas. También permite almacenar los datos, posibilitando el análisis de tendencias.

El otro elemento activo en el sistema de gestión de redes, tal como se puede observar en la Figura 1, es el agente de gestión. Las plataformas clave tales como hosts, bridges, routers y hubs, se pueden equipar con SNMP para ser gestionadas desde la estación de gestión. El agente de gestión es un software que proporciona acceso a los datos de gestión de un dispositivo de red particular, responde a peticiones de información y acciones por parte la estación de gestión y puede enviar a la estación de gestión cierta información importante no solicitada de un modo asíncrono.

Los dispositivos situados en la red son gestionados mediante transacciones entre el gestor y el agente. SNMP proporciona dos clases de transacciones de gestión:

- Petición por parte del gestor y respuesta por parte del agente.
- Notificaciones no solicitadas (*traps*) desde el agente al gestor.

Para poder gestionar los recursos, éstos son representados como objetos. Cada objeto es, esencialmente, una variable que representa un aspecto del agente gestionado. A la colección de objetos se le denomina MIB (base de información de gestión). La MIB funcionan como un conjunto de puntos de acceso al agente desde la estación de gestión. Una estación de gestión realiza la función de monitorización leyendo el valor de los objetos de la MIB y puede forzar a que tenga lugar una acción en un agente o cambiar la configuración del mismo modificando el valor de algunas de estas variables.

La estación de gestión y los agentes se comunican utilizando el protocolo de gestión de red. El protocolo utilizado para la gestión de redes TCP/IP es el SNMP, el cual incluye las siguientes capacidades:

- GET: permite a la estación de gestión recuperar el valor de los objetos en el agente.
- SET: permite a la estación de gestión modificar el valor de los objetos en el agente.
- TRAP: permita al agente notificar a la

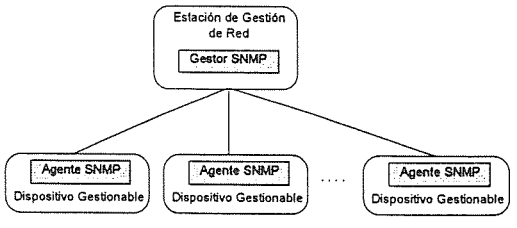


Figura 1

estación de gestión la ocurrencia de eventos significativos.

SNMP forma parte de la familia de protocolos TCP/IP y es un protocolo de nivel de aplicación que opera por encima de UDP. SNMP es un protocolo no orientado a conexión, pues utiliza UDP, por lo tanto, cada intercambio es una transacción separada entre la estación de gestión y un agente.

La Figura 2 proporciona una vista detallada del contexto del protocolo. Desde una estación de gestión se pueden enviar tres tipos de mensajes: *GetRequest*, *GetNextRequest* y *SetRequest*. Los dos primeros son variaciones de GET. Los tres mensajes son reconocidos por el agente mediante un mensaje *GetResponse*, el cual es pasado a la aplicación de gestión. Además, un agente puede emitir un mensaje *Trap* en respuesta a un evento que afecte a la MIB y a los recursos gestionados.

El sistema de seguridad de SNMP, está basado en el concepto de comunidad. Una comunidad es una relación entre un agente SNMP y un conjunto de estaciones de gestión SNMP que define unas características de autentificación y control de acceso. El agente le da un nombre de comunidad y las estaciones de gestión deberán emplear ese nombre de comunidad en todas las operaciones Get y Set. La petición realizada por la estación de gestión únicamente es atendida si el nombre de comunidad es correcto.

2.3 La estructura de información de gestión (SMI) La estructura de la información de gestión (RFC 1155), define el marco de trabajo general dentro del cual una MIB puede ser definida y construida. SMI identifica los tipos de datos que pueden ser usados en la MIB y cómo se representan y nombran a los recursos dentro de la MIB. Así, la MIB puede almacenar sólo tipos de datos simples: escalares y arrays bidimensionales de escalares (tablas).

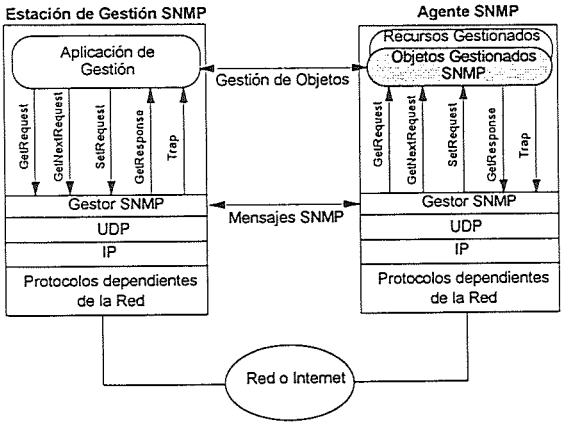


Figura 2

SNMP utiliza el esquema jerárquico de nombrado desarrollado por ISO. En este esquema, el espacio de nombres forma un árbol. Este consiste en una raíz conectada a un conjunto de nodos etiquetados. Cada nodo, a su vez, puede estar conectado a otros sub-nodos también etiquetados. El nombre de un nodo, denominado Identificador de Objeto, es la secuencia de los enteros de las etiquetas de cada nodo, desde la raíz hasta el nodo en cuestión.

Como se ve en la Figura 3, todos los objetos de interés para SNMP están en la parte del árbol correspondiente a *iso* y cuelgan del nodo *internet*. Así, el nodo *internet* tiene el valor de identificador de objeto 1.3.6.1.

Los objetos de una MIB SNMP, y la estructura MIB completa se definen usando un subconjunto de ASN.1.

Cuando un fabricante desarrolla un dispositivo gestionable SNMP, típicamente suele diseñar una MIB que modela los datos que son importantes para la gestión de ese tipo de dispositivos. A este tipo de MIB se la denomina propietaria. Las MIBs propietarias generalmente cuelgan del nodo asignado a los fabricantes en la rama {...internet(1) private(4) enterprises(1)} del árbol de nombres. Por ejemplo, todas las MIBs de Hewlett-Packard cuelgan de la rama {...internet(1) private(4) enterprises(1) hp(11)}.

Las MIBs estándar cuelgan de la rama {...internet(1) mgmt(2)} en el árbol de nombres. La más importante es la MIB-2, que describe los objetos que deben ser implementados por todos los nodos que incluyan TCP/IP. Otro ejemplo de MIB

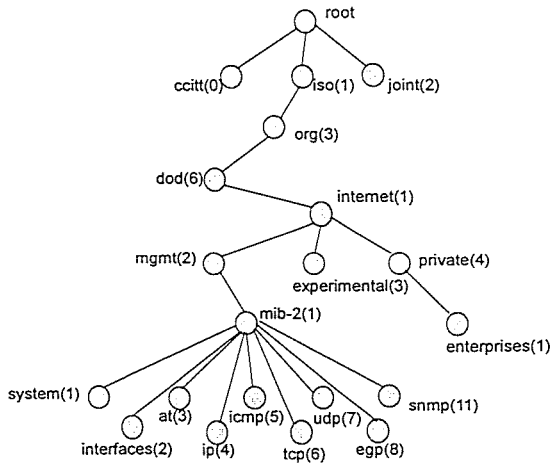


Figura 3

es RMON.

La MIB-2 está dividida en varios grupos, de manera que los nodos deben implementar o todos los objetos o ninguno dentro del mismo grupo.

En la tabla 1 se muestran los grupos de la MIB-2.

A la representación textual de una MIB se le denomina módulo y está escrito utilizando un subconjunto del lenguaje ASN.1. Estos módulos MIB se almacenan en ficheros ASCII.

A continuación se muestra un fragmento del módulo definido en RFC1213:

```
tcpConnTable OBJECT-TYPE
        SYNTAX  SEQUENCE OF TcpConnEntry
        ACCESS  not-accessible
        STATUS  mandatory
        DESCRIPTION
            "A table containing TCP connection-specific information."
        ::= { tcp 13 }

    tcpConnEntry OBJECT-TYPE
        SYNTAX  TcpConnEntry
        ACCESS  not-accessible
        STATUS  mandatory
        DESCRIPTION
            "Information about a particular current TCP connection.  An object of
             this type is transient, in that it ceases to exist when (or soon after) the
             connection makes the transition to the CLOSED state."
        INDEX   { tcpConnLocalAddress,
                  tcpConnLocalPort,
                  tcpConnRemAddress,
                  tcpConnRemPort }
        ::= { tcpConnTable 1 }

    TcpConnEntry ::=
        SEQUENCE {
            tcpConnState            INTEGER,
            tcpConnLocalAddress     IpAddress,
            tcpConnLocalPort        INTEGER (0..65535),
            tcpConnRemAddress       IpAddress,
            tcpConnRemPort          INTEGER (0..65535)
        }

    tcpConnState OBJECT-TYPE
        SYNTAX  INTEGER
                .
                .
        ::= { tcpConnEntry 1 }

    tcpConnLocalAddress OBJECT-TYPE
        SYNTAX  IpAddress
                .
                .
        ::= { tcpConnEntry 2 }

    tcpConnLocalPort OBJECT-TYPE
        SYNTAX  INTEGER (0..65535)
                .
                .
        ::= { tcpConnEntry 3 }

    tcpConnRemAddress OBJECT-TYPE
        SYNTAX  IpAddress
                .
                .
        ::= { tcpConnEntry 4 }

    tcpConnRemPort OBJECT-TYPE
        SYNTAX  INTEGER (0..65535)
                .
                .
        ::= { tcpConnEntry 5 }
```

3. Elementos del sistema

En esta sección se describen los componentes desarrollados para complementar la docencia en gestión de redes.

3.1. El entorno docente

El entorno docente está formado por tres componentes y cumplirán respectivamente los siguientes aspectos:

1. Enviar los diferentes mensajes básicos de SNMP, es decir *GetRequest*, *GetNextRequest* y *SetRequest*.
2. Combinar las diferentes operaciones de SNMP desde un entorno gráfico al que se le han añadido características adicionales para la gestión eficaz de una red con varios componentes equipados con el software de agente.
3. Como complemento de los puntos anteriores, se necesita la presencia de un agente que sea capaz de responder a las peticiones con los valores que tiene almacenados en su MIB.

Para poder cumplir con el primero de los puntos, se han implementado una serie de aplicaciones cuya única misión es recoger los parámetros necesarios, enviar la petición SNMP y mostrar el resultado de la operación. Estas herramientas se han implementado para entornos MS-DOS y Windows, pues al ser conocidos y utilizados por los alumnos no van a suponer un problema adicional.

Para cumplir con el punto segundo se ha desarrollado una estación de gestión ejecutable desde Windows que aprovecha la facilidad de manejo que aporta el entorno gráfico. Además de incorporar las características del punto anterior, se aportan una serie de utilidades adicionales que facilitan las tareas de gestión. Se destacan la detección automática de agentes, presentación gráfica de los mismos en una ventana de trabajo, realización de diferentes operaciones sobre los agentes utilizando el ratón, actualización inmediata y periódica de la red, peticiones de lectura de múltiples objetos tanto predefinidas como configurables con diseño personalizado de plantillas, conexión telnet con los nodos que lo permitan, personalización de los nombres de comunidad para cada nodo, traducción automática de secuencias de

Tabla 1. Grupos de la MIB-2

Grupo*	Descripción
system	Descripción del sistema
interfaces	Descripción de los interfaces del sistema
at	Traducción de direcciones físicas - IP
ip	Información del protocolo IP
icmp	Información del protocolo ICMP
tcp	Información del protocolo TCP
udp	Información del protocolo UDP
egp	Información del protocolo EGP
transmission	Otras MIB de los medios de transmisión
snmp	Estadísticas del propio protocolo SNMP

enteros a nombres y viceversa, y salida de los resultados por pantalla o fichero de texto configurable por el usuario.

Con lo que respecta al tercer punto, existen dos posibilidades: utilizar agentes reales o simulados. En el caso de utilizar los primeros, hay que observar que si se permite la modificación de valores utilizando *SetRequest* por parte de los alumnos, estos podrían causar disrupciones en el servicio prestado por los mismos. La estrategia consiste en utilizar los primeros en las operaciones de lectura siempre que sea posible, pues ofrece una visión más real de la gestión de la red, mientras que la modificación de objetos de la MIB se realizará sobre agentes con MIB estática. Para ello se ha implementado un falso agente, esto es, un programa capaz de responder a las diferentes peticiones del gestor, pero cuyos valores no representan a la maquina sobre la que se ejecuta, es decir, su MIB no se actualiza.

3.2. La programación de las aplicaciones

Para el desarrollo de las herramientas aquí descritas se han utilizado las siguientes librerías:

- PC/TCP OnNet Developer's Toolkit 3.0. de Ftp Software Inc. para el desarrollo de las diferentes herramientas que se ejecutan sobre el sistema operativo MS-DOS, como son *pcget.exe, pcnext.exe*, *pcset.exe* y *agent.exe*.
- Librería WinSNMP para la programación de las herramientas en entorno Windows, estas herramientas son la estación de gestión y una aplicación para realizar operaciones básicas en SNMP.

Se utilizó el compilador BorlandC++ 3.1 debido a que, de entre los que nos restringen las diferentes librerías, era de los pocos que permitían programar para y desde entornos DOS y Windows. Además, ofrece un control sobre el programa que está siendo desarrollado superior al ofrecido por otros compiladores.

4. Elementos del sistema.

En esta sección se describirá el funcionamiento de cada una de las aplicaciones desarrolladas.

4.1. Herramientas básicas para SNMP

Existen dos conjuntos de herramientas destinadas a realizar estas operaciones: aquellas que funcionan en entorno MS-DOS y aquellas que funcionan en entorno Windows.

Las herramientas que funcionan en entorno MS-DOS consisten en unos programas sencillos llamados ***pcget, pcnext*** y ***pcset*** a los que se les pasa los parámetros necesarios.

Para las operaciones *pcget* y *pcnext* se les pasan :

pcget <host> <variable> <community name>

En cambio, para la operación de *pcset* se necesitan más parámetros, con lo que la forma de llamarla es la siguiente:

pcset <host> <variable> <tipo de valor> <nuevo valor> <community name>

En el caso de que el usuario pase mal los parámetros requeridos, el programa muestra un mensaje de error indicando la forma correcta.

Existen dos formas de indicar la variable de la MIB sobre la que se desea trabajar:

- Formato numérico, donde la secuencia de números debe estar precedida por "_",
- Formato textual, precedido por el nombre del fichero que contiene el módulo mib compilado: <mib>;<nombre>_<índice>

Así, por ejemplo, la instancia de *sysName* se representará por "mib;sysName_0" o "_1.3.6.1.2.1.1.5.0".

La herramienta pensada para entorno Windows ofrece una ventana con los siguientes controles:

Caja de edición **host** en la que debe indicarse la dirección IP del nodo al que va a enviarse la petición SNMP.

Caja de edición **variable** en la que se indicará la variable requerida. En este caso solo está permitida la notación numérica.

Caja de edición **nuevo valor** en el que se indicará el valor que se desea que tome la variable indicada en la operación *SetRequest*. Para las operaciones *GetRequest* y *GetNextRequest* no debe ser completada.

Caja de edición **respuesta** en la que se mostrará la respuesta del agente.

Caja de edición combinada **tipo** en la que se indicará el tipo de variable devuelta. El tipo de valor debe ser indicado para poder realizar una operación *SetRequest*.

El botón **Get** realiza una operación *GetRequest* sobre los datos indicados mostrando el resultado en la caja de edición **respuesta**.

El botón **GetNext** realiza una operación *GetNextRequest* sobre los datos indicados mostrando el resultado en la caja de edición **respuesta**. Además, cambia el nombre del objeto pedido, que aparece en la caja de edición **variable**, por el que llega en la respuesta del agente. Esto permite recorrer en orden lexicográfico todas las instancias de una tabla, un grupo, etc. pulsando repetidamente este botón.

El botón **Set** realiza una operación *SetRequest* con los datos indicados mostrando el resultado en la caja de edición respuesta.

El botón **Salir** permite cerrar la aplicación.

La ventana principal de esta aplicación se puede observar en la Figura 4.

4.2. El agente

Este programa implementa las diferentes funciones de un agente, respondiendo a las peticiones SNMP que lleguen, pero los objetos de la MIB no reflejan el estado de la máquina sobre el que se ejecuta. Desde la consola se permite modificar los valores de las variables a gusto del usuario, facilitando de este modo la edición de la MIB.

En la ventana principal del programa **agent** se puede ver una lista de variables por las que puede desplazarse con los cursores. Junto a cada variable se puede ver el tipo de acceso que ofrece (RO ⇒ solo lectura, RW ⇒ lectura y escritura o NA ⇒ no accesible) y el valor que tiene en ese momento.

Las funciones ofrecidas al usuario son:

- Cambiar el valor de una variable.
- Cambiar el valor de todas las variables a uno predefinido.
- Salir del programa.

La operación de envío de traps no ha sido implementada todavía.

En la Figura 5 se puede ver la pantalla principal del programa **agent**.

4.3. La estación de gestión

Este programa, denominado **Gestired**, permite representar gráficamente los elementos de una red TCP/IP que tienen instalado el software de agente SNMP. La fácil identificación del tipo de nodo permite entender y gestionar fácilmente la red. Los diferentes agentes de la red están representados por su nombre y un icono representativo de la clase de equipo de red, tal como puede verse en la Figura 6. Sobre estos nodos se podrá realizar una serie de acciones orientadas a su gestión, tales como pedir o modificar los valores de los objetos de su MIB. También se puede realizar un Telnet sobre el nodo para modificar la configuración del dispositivo en caso de ser necesario.

Por defecto, a cada nodo de la red se le asigna un icono dependiendo del valor del objeto *sysServices*, cuyo valor indica el conjunto de servicios ofrecidos por la entidad; así, por ejemplo, los iconos "repetidor" representarán nodos que implementen funciones del nivel 1 de la pila OSI, "bridge" para los que implementen funciones de

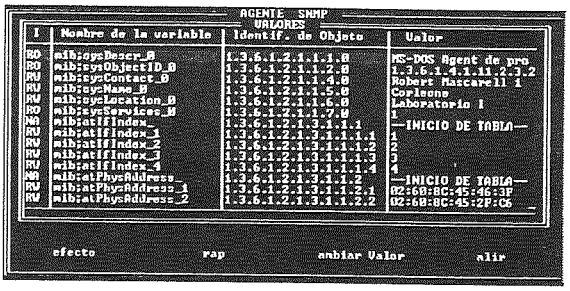


Figura 5

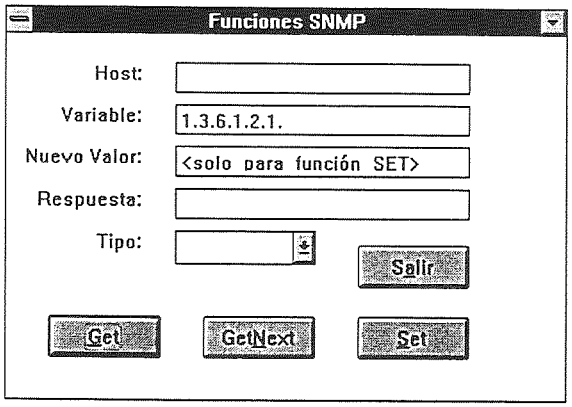


Figura 4

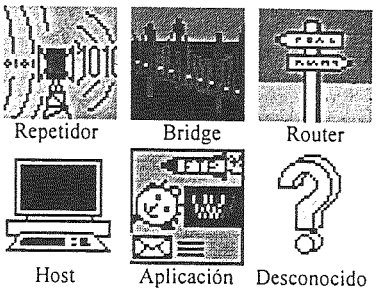


Figura 6

nivel 2, "router" para nivel 3, "host" para nivel 4 y "aplicación" para el nivel 7, tal y como está definido en [7].

El sistema de gestión implementa una función para el descubrimiento automático de dispositivos en la red. Para cada dirección IP correspondiente a un posible nodo se le envía un *GetRequest* en el que se le piden los valores correspondientes a *sysName* y *sysServices*. Si el nodo responde, se obtienen los datos de este host de la respuesta, dibujándose en pantalla el correspondiente icono. En el caso de no obtenerse respuesta, se trata, bien de un equipo que no implementa SNMP, de una dirección IP no asignada a ningún equipo, o bien, que el equipo está apagado.

El descubrimiento de nodos puede realizarse mediante dos operaciones: la inserción de un rango de direcciones IP, o la inserción de una subred.

Una vez representada la red en la ventana, se pueden añadir nuevos equipos de forma individual mediante la operación "host" del menú "insertar". También es posible crear la red insertando individualmente cada uno de los nodos.

Para poder visualizar más eficazmente los objetos de una MIB, se ha desarrollado una utilidad que permite visualizar un conjunto de objetos escalares y tabulares sobre una misma ventana. La aplicación más inmediata será la presentación de los objetos tabulares en forma de tabla, teniendo la posibilidad de consultar y representar las columnas que se desee. Para ello se ha desarrollado la utilidad PPM (Plantilla para Peticiones Múltiples) que permite editar un fichero ASCII en el que se indican los objetos a consultar. Un fichero PPM debe mantener una estructura predefinida que se muestra en una ventana de ayuda del programa.

Puesto que el funcionamiento de la red es dinámico, el programa debe de ser capaz de averiguar si un nodo de la red ha sido desconectado o ha dejado de funcionar. Para ello se utiliza el refresco, que utiliza las mismas operaciones que la función de descubrimiento de nodos. En el caso de que el dispositivo responda a un refresco se considera que continua activo; en caso contrario se supone que hay problemas con el nodo, y se representará gráficamente con una cruz roja sobre su icono; en el caso de un host, el icono será el representado en la Figura 7.

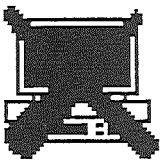


Figura 7

Existen dos tipos de refrescos: los refrescos sobre un nodo específico y los que se realizan sobre la red completa. El último se realiza tras el vencimiento de un temporizador controlable por el usuario. También se puede ordenar un refresco inmediato en cualquier momento.

En la pantalla principal del programa que podemos observar en la Figura 8, aparece un menú con las siguientes opciones:

- Archivo ⇒ permite abrir y cerrar el fichero que contiene la red de trabajo, crear una nueva red, almacenar la red actual en un fichero, abrir la aplicación Notepad del Windows, y salir del programa.
- Plantilla ⇒ se utiliza para crear nuevas plantillas PPM o abrir las definidas previamente.
- Insertar ⇒ permite descubrir los dispositivos SNMP de una red por cualquiera de los dos métodos indicados anteriormente e insertar nodos individualmente.
- Funciones ⇒ ofrece la posibilidad de realizar un refresco inmediato de la red completa.
- Configuración ⇒ se utiliza para configurar los diferentes aspectos del sistema, entre los que se destacan:
  1. El periodo de refresco del temporizador.
  2. El nombre de comunidad por defecto, utilizado para descubrir agentes.
  3. Destino de los datos recibidos: salida por pantalla o almacenado en un archivo de texto.
  4. Inclusión o eliminación de ficheros que contienen un cierto módulo MIB compilado; esto es útil para poder utilizar la representación textual de los objetos de las MIBs que implementen los agentes.
- Ayuda ⇒ ayuda del programa.

Cuando un usuario desee realizar una operación sobre un cierto dispositivo de la red, deberá hacer "clic" sobre su icono. Se abrirá una caja de dialogo como la mostrada en la Figura 9, en la que se aprecian una serie de funciones aplicables exclusivamente al nodo seleccionado.

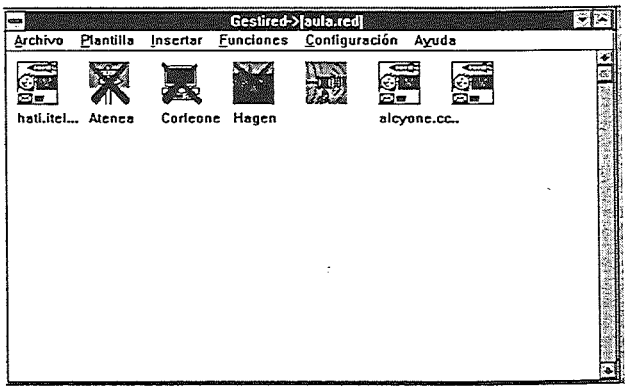


Figura 8

Estas funciones son:

1. Realizar un *GetRequest*, un *GetNextRequest* o un *SetRequest* sobre el nodo seleccionado.
2. Realizar un *GetNextRequest* recursivo sobre el nodo. Se utiliza por ejemplo para obtener todas las instancias de un objeto tabular.
3. Pedir una plantilla, ya sea de las predefinidas en el sistema, o definida por el usuario a través de la creación de una PPM.
4. Realizar una conexión tipo terminal virtual (telnet) en caso de que el nodo lo permita.
5. Efectuar un refresco exclusivamente sobre ese nodo y sin afectar al temporizador interno del sistema.
6. Cambiar el nombre de comunidad asignado al nodo en cuestión.
7. Cambiar el icono que representa al nodo, en caso de que el usuario considere que el dispositivo realiza otras funciones.
8. Eliminar el nodo de la red.
9. Cancelar la operación.

5. Conclusiones

Este sistema acerca al usuario la gestión de redes TCP/IP de una forma amena, amigable e intuitiva.

Como aplicación docente, se trata principalmente de un entorno económico destinado fundamentalmente a la realización de prácticas, de manera que ayude a la comprensión del protocolo SNMP, las funciones principales de una estación de gestión y de los agentes repartidos por toda la red. El software desarrollado permite por una parte suplir la carencia de agentes de gestión SNMP, y por otro poder realizar diversas operaciones de monitorización y control sobre los mismos. Para ello tan solo se precisa de un ordenador personal por usuario y al menos otro que actúe como agente en caso de no disponer de ninguno real.

Gracias a la personalización de las plantillas de visualización, la estación de gestión será de gran utilidad para la comprensión de los objetos tabulares de las mibs, pues permite seleccionar las columnas que realmente necesitemos, y su presentación se realizará en forma de tabla aunque sus valores se obtengan en varios PDUs.

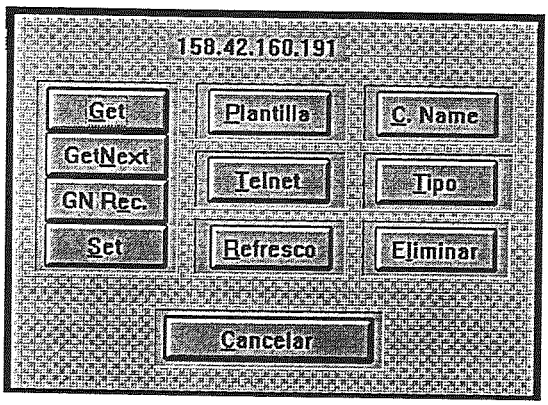


Figura 9

En el caso de disponer de una sonda RMON (bien real o simulada), con el sistema presentado se podrá estudiar con detalle el estándar RMON (RFC 1757), esto es, los fundamentos técnicos de la mib rmon.

Esta versión de la estación de gestión constituye una primera aproximación a lo que se considera básico para la realización de unas prácticas de laboratorio, siempre teniendo en cuenta el bajo coste de las mismas. Posteriores versiones incluirán más elementos y funciones importantes en la estación de gestión así como la realización de agentes que incluyan algunas de las características de RMON, lo que contribuirá a una mejor comprensión de estos temas tan importantes hoy en día las grandes redes informáticas. También deben ser extendidos para soportar el protocolo SNMPv2.

Referencias

[1] W. Stallings, "SNMP, SNMPv2, and CMIP. The Practical Guide to Network-Management Standards", Addison-Wesley Publishing Company, 1993

[2] M. Rose, "The Simple Book. An Introduction to Internet Management", Prentice Hall, 1994

[3] D. T. Perkins, "Understanding SNMP MIBs", SynOptics Communications Inc, September, 1993

[4] J. Case, "A Simple Network Management Protocol (SNMP)", Request for Comments 1157, SNMP Research, Mayo 1990

[5] M. Rose, "Concise MIB Definitions", Request for Comments 1212, Performance Systems International, Marzo 1991

[6] M. Rose, "Structure and Identification of Management Information for TCP/IP-based Internets", Request for Comments 1155, Performance Systems International, Mayo 1990

[7] K. McCloghrie, "Management Information Base for Network Management of TCP/IP-based Internets", Request for Comments 1213, Hughes LAN Systems Inc, Marzo 1991

Formación permanente en las PYMES usando herramientas multimedia

M. A. PÉREZ JUÁREZ, M. J. VERDÚ PÉREZ Y R. MOMPÓ GÓMEZ
DEPARTAMENTO DE TEORÍA DE LA SEÑAL Y COMUNICACIONES E INGENIERIA TELEMATICA
E.T.S.I. DE TELECOMUNICACIÓN DE VALLADOLID. UNIVERSIDAD DE VALLADOLID
C/ REAL DE BURGOS S/N, 47011 VALLADOLID. ESPAÑA.
TEL.: +34-83-423260. FAX.: +34-83-423261
Correo electrónico: marper@tel.uva.es, marver@tel.uva.es, jyr@dvnet.es

Abstract:

The purpose of the authors is to share their experience in which to SME's employees training through Computed Assisted Learning and Telematic Tools concers. We want to show why training is a need for any company that wants to survive in a competitive market, give then some good reasons for which this training must be done using distance learning, and finally explain why telematic tools and multimedia materials should be used in distance learning. After, we want to describe the multimedia interactive courses we have created for this purpose and the tool we have used to create them (Multimedia ToolBook Instructor). We also want to explain our experience in testing our software and also our experience in using it in real learning experiences. Finally we want to let you know our future working plans.

PARTE I: ¿Por qué la Formación Permanente debe hacerse a distancia pero empleando materiales multimedia y herramientas telemáticas?

1. Introducción

Los autores del presente artículo formamos parte de un grupo de trabajo denominado "Grupo Canalejas", uno de cuyos campos de trabajo es la multimedia educativa.

El "Grupo Canalejas" es un grupo interdisciplinario e interuniversitario que investiga las posibilidades y ventajas de la incorporación de la TI (Tecnología de la Información) en el proceso formativo. En concreto, en el seno del grupo, existen dos equipos de investigación, uno de la *Universidad de Salamanca* y otro de la *Universidad de Valladolid*. El primero lo constituyen profesores de la Facultad de Educación de Salamanca y el segundo profesores de la Escuela Técnica Superior de Ingenieros de Telecomunicación de Valladolid.

La razón fundamental que justifica la participación de los dos equipos en la tarea que estamos llevando acabo, es el carácter *interdisciplinar* de dicha tarea. Por la propia naturaleza de la multimedia educativa, en su elaboración se requiere la presencia de investigadores pertenecientes, al menos, a dos campos de trabajo: de un lado, el campo de la educación ya que los productos elaborados van a utilizarse en procesos de aprendizaje y, de otro, el campo de las telecomunicaciones, más concretamente el ámbito de la telemática, pues los materiales educativos a elaborar son multimedia e incorporan herramientas telemáticas y resultan por tanto radicalmente distintos a los basados en papel que tradicionalmente han venido confeccionado los docentes como apoyo para sus clases.

Hay que añadir que el presente grupo de trabajo cuenta con el apoyo de una empresa de nuestro entorno regional llamada Divisa Informática S.A.

Nuestro objetivo concreto en este artículo, es dar a conocer nuestra experiencia en lo que a formación permanente en PYMES (Pequeñas y Medianas Empresas) mediante CDRoms multimedia que incorporan herramientas telemáticas se refiere. En concreto queremos:

⇒ Justificar la necesidad de la formación permanente en la empresa en general y como consecuencia en las PYMES en particular. Plantear después las características particulares de la educación a distancia y también las ventajas que este método educativo puede aportar a la formación continua. Presentar posteriormente los inconvenientes intrínsecos a la educación a distancia y plantear como las herramientas telemáticas pueden corregirlos.

⇒ Describir la solución concreta empleada por nosotros. Y aquí queremos:

- Describir la herramienta utilizada en la elaboración de nuestros cursos multimedia interactivos: ToolBook Instructor II.
- Describir los cursos que hemos realizado.

⇒ Relatar nuestra experiencia en lo que a la utilización de dichos cursos en la formación permanente en PYMES se refiere.

2. ¿Qué es la Formación Permanente?

La formación permanente alude al proceso formativo que tiene lugar a lo largo de toda la vida de la persona, una vez finalizada su educación

formal. Como notas definitorias de la misma podemos señalar las siguientes:

- Educación **a lo largo de toda la vida.**
- Abarcadora de **todas las modalidades educativas.**
- Relacionada con **todas las gamas del saber.**
- Dirigida a **todas las gentes.**

3. Formación Permanente y Empresa

La educación y la formación son herramientas estratégicas que una sociedad necesita aplicar continuamente para mantener una ventaja global competitiva y para crear un mejor estándar de vida. En concreto, la formación permanente es, por diversos motivos, un factor estratégico que ninguna sociedad que quiera estar en los puestos de cabeza de la economía mundial puede descuidar.

La TI está evidentemente cambiando nuestras vidas. Es necesario sin embargo tener en cuenta, que dicho cambio no será a mejor para todos y cada uno de los individuos de la sociedad, aunque sí lo será, considerado globalmente.

Para algunos, la introducción de la TI en sus vidas, significará por ejemplo, el que le pongan un PC en su puesto de trabajo que le permitirá acceder a toda la información. Para otros, el cambio será más radical, pues perderán sus empleos.

Es éste, un aspecto de la TI que no debe ser ignorado.

A pesar de ser cierto que la introducción de la TI en las empresas, traerá como consecuencia a corto plazo, una pérdida de empleos, es ésta una transformación inevitable.

En una economía de mercado, aquella empresa que no incorpore la TI, *a corto plazo* salvaguardará los puestos de trabajo de todos sus empleados, sin embargo, *a medio o largo plazo*, el resultado podría ser que todos perdiesen sus puestos de trabajo al ser la empresa expulsada del mercado por otras, pertenecientes a la misma industria, más eficientes que sí incorporaron la TI a sus métodos de producción.

Ante este proceso continuo e inevitable de cambio tecnológico en el ámbito empresarial, uno de los retos de las sociedades modernas es el lograr que su capital humano avance al mismo ritmo para poder participar de manera eficaz en esas industrias que requieren personal bien formado y que representan, cada vez en mayor porcentaje la base de las economías modernas. Algunos ejemplos de este tipo de industrias son la aeroespacial, la bioquímica o la de telecomunicaciones.

Las economías modernas no pueden competir por más tiempo en industrias que desarrollan actividades intensivas en factor trabajo como la textil, que se han desplazado al lejano oriente y más recientemente a los denominados países del tercer mundo, donde la mano de obra es barata.

En estas economías modernas, uno de los resultados de la transición de una economía intensiva en factor trabajo a una intensiva en tecnología ha sido una gran masa de parados estructurales, con una formación escasa o nula, que necesita ser reeducada antes de poder reincorporarse al mercado laboral optando a los nuevos puestos de trabajo que la TI crea y que se espera que de manera global, aunque no industria a industria, compensen los que con su paulatina llegada destruyó.

La TI causa por tanto, una reestructuración en las tareas llevadas a cabo por el capital humano, una parte del cual necesitará ser reeducado antes de poder desempeñar las nuevas tareas que ahora se les asignen.

Será por tanto necesario, incrementar la habilidad de los trabajadores de **cuello azul** para aumentar su productividad al lograr que usen máquinas cada vez más complejas y ordenadores.

Sin embargo, el reto de la formación en las empresas no está tan sólo en los trabajadores de cuello azul, sino también en los **de cuello y bata blanca** que tienen día a día que vérselas con los últimos adelantos en sus respectivos campos de trabajo.

Todo lo dicho anteriormente nos lleva a concluir de manera general, que debido al proceso de cambio tecnológico en el que estamos inmersos, aquella porción del capital humano que desee sobrevivir en una economía competitiva y de mercado como la nuestra debe continuar formándose.

Sin embargo, a pesar de la importancia de la formación permanente, algunos países desarrollados como Estados Unidos invierten aún poco en ella. Así las empresas de los Estados Unidos gastan aproximadamente 30 billones de dólares al año, en formación permanente que afecta a aproximadamente un 10 % del capital humano del país. La cantidad de dinero parece elevada, no lo es sin embargo, comparada con la gastada por esas mismas empresas en factor capital (nuevas plantas y

equipamiento), la cual excede los 475 billones de dólares. [1]

Considerando la importancia creciente del capital humano en relación al otro tipo de capital, es evidente que las corporaciones necesitan encontrar un balance más adecuado entre las partidas de gasto dedicadas a ambos tipos de capital.

Si pensamos detenidamente en lo dicho hasta ahora, nos daremos cuenta de que el argumento empleado para presentar la formación permanente como una necesidad en las empresas es la necesidad que tienen éstas de sobrevivir al proceso continuo de cambio tecnológico en una economía competitiva.

4. ¿Por qué Formación Permanente a Distancia en las PYMES?

En primer lugar queremos hablar brevemente de la enseñanza a distancia y sus características para justificar que constituye un método de enseñanza válido para ser empleado en la formación permanente en PYMES, donde los alumnos por la propia naturaleza de este proceso formativo van a ser adultos.

Más que dar aquí una definición de enseñanza a distancia, preferimos reproducir dos de las muchas que a lo largo de la historia expertos en el tema han planteado:

"La educación a distancia es una estrategia educativa basada en la aplicación de la tecnología al aprendizaje sin limitación del lugar, tiempo, ocupación o edad de los estudiantes, Implica nuevos roles para los alumnos y para los profesores, nuevas actitudes y nuevos enfoques metodológicos." (J. L. García) [2]

"Educación a distancia es una modalidad mediante la cual se transfieren informaciones cognoscitivas y mensajes formativos a través de vías que no requieren una relación de contigüidad presencial en recintos determinados." (V. Guedez) [2]

La educación a distancia se caracteriza por una **gran flexibilidad**, pues se independiza el proceso de aprendizaje del espacio y del tiempo, es decir :

⇒ En la *Enseñanza Presencial*, los alumnos asisten a centros de formación, en los que, dentro de ciertos horarios, reciben lecciones magistrales impartidas por profesores más o menos buenos, dependiendo de la oferta docente de la zona en la que el alumno resida.

⇒ En la *Enseñanza a Distancia* el alumno aprende, desde el lugar que él decide -por ejemplo su hogar o su puesto de trabajo- en el horario que él decide -variable tantas veces como él quiera-.

La enseñanza a distancia logra **atender a una población educativa dispersa.**

Además fomenta el autoaprendizaje pues el **profesor** pasa a ser un mero guía del proceso de aprendizaje, adquiriendo el **estudiante** todo el protagonismo de su proceso formativo.

El alumno de la educación a distancia suele ser además, un individuo adulto y en el caso de la formación permanente en empresas así será siempre, y por tanto con una historia vivencial llena de experiencias, conocimientos, capacidades, hábitos y actitudes, lo cual le permite recorrer la mayor parte del proceso de forma autónoma e independiente.

El contacto del adulto con la realidad configura la "presencialidad" de la educación a distancia por lo cual podrá prescindir de las relaciones directas o presenciales.

Además el adulto puede encontrarse más a gusto recibiendo enseñanza a distancia, que volviendo a un aula tradicional, a la institución escolarizada.

Como el adulto no precisa de enseñanza presencial y se siente más incómodo con ella, la educación a distancia se perfila como el método más adecuado a sus necesidades y características.

Por todo lo anteriormente expuesto la enseñanza a distancia es un método no sólo válido sino también adecuado para la formación permanente en empresas.

Queremos destacar finalmente el factor económico que muchas veces constituye el factor fundamental por el cual las empresas optarán por la formación permanente a distancia.

La enseñanza a distancia ofrece la posibilidad de la **comunicación masiva** de un determinado mensaje cuyo diseño y producción ha comportado un determinado coste, dando lugar así a la aparición de **economías de escala** que conducen a una **reducción significativa de costes.**

La formación permanente a distancia resulta mucho más barata que la formación permanente presencial. Esto es algo obvio a poco que se reflexione sobre ello. Sin embargo, para aquellos a los que no les resulte obvio presentamos aquí unos presupuestos comparativos aproximados que hemos realizado para un curso de 1 semana de duración (a

razón de 5 horas al día) que se quiera impartir a 10 empleados.

Formación Presencial:

Alojamiento: 10*(10.000ptas/noche*5 noches) =500.000ptas

Dietas comida: 10*(5.000ptas/día*5 días)=250.000ptas

Desplazamientos: 10*50.000ptas/alumno=500.000ptas

Profesor: 10.000ptas/hora*25 horas=250.000ptas*

Material: 10*4.000ptas/alumno=40.000ptas

TOTAL:1,540.000 ptas

Formación a Distancia:

Profesor: 10.000ptas/hora*25 horas=250.000ptas*

Material: 10*4.000ptas/alumno=40.000ptas

TOTAL:290.000 ptas

*Nótese que consideramos en ambos casos el mismo número de horas de clase, esto es así puesto que una enseñanza no presencial no significa una dedicación menor del docente a los alumnos. Cuando la enseñanza a distancia es mediante vídeo-conferencia esto es obvio, por contra, cuando la enseñanza a distancia es mediante EAO (Enseñanza Asistida por Ordenador) es cierto que las horas de "clase" del docente son menores, pero es un hecho, comprobado por nosotros mismos, que se incrementan considerablemente las horas que el docente debe dedicar a contestar las consultas que los alumnos les plantean mediante correo electrónico.

Como se puede ver los costes son considerablemente mayores cuando el curso de formación se imparte de manera presencial. Además hay que tener en cuenta que hay costes como "los de oportunidad" que no han sido tenidos en cuenta en esta comparativa y que son significativamente mayores en el caso de la enseñanza presencial.

5. ¿Por qué la Formación Permanente a Distancia debe emplear la tecnología multimedia y las herramientas telemáticas?

No hemos hablado hasta ahora de las desventajas de la educación a distancia. Sin embargo, las hay y sería un error importante obviarlas. Como desventaja principal, se suele plantear tradicionalmente la escasez de la interacción de los alumnos entre sí y de éstos con su profesor.

Esto trae como consecuencia una posible sensación de aislamiento en el alumno que incapaz de resolver los problemas normales que el proceso de aprendizaje le crea puede en un momento determinado decidir abandonar.

La interactividad es un elemento clave en el proceso de aprendizaje, y es precisamente interactividad lo que proporcionan la tecnología multimedia y sobre todo las herramientas telemáticas que como el correo electrónico o los foros de debate permiten hablar de una **"verdadera aula virtual"** que no tiene nada que envidiar al aula real en el se imparten las clases presenciales.

La IDL (Interactive Distance Learning) tiene todas las ventajas de la educación a distancia, pero además soluciona algunos de sus problemas, a la vez que trae otras ventajas nuevas que la educación a distancia tradicional ni siquiera se atrevía a imaginar.

En una economía competitiva, las corporaciones además introducen **nuevos servicios y productos** y deben acelerar y reducir los costes del proceso de diseminar la información sobre nuevos productos entre todo el departamento de ventas. Es vital que los comerciales conozcan todas las características de los nuevos productos y el valor que los mismos pueden aportar a los clientes. La necesidad del empleo de IDL es este punto es cada vez si cabe más crítica al aumentar en las empresas los departamentos de comercial y atención al cliente y al estar éstos cada vez más diseminados entre múltiples enclaves que pueden estar diseminados a lo largo del estado, el país o el mundo.

PARTE II: NUESTRA SOLUCIÓN

Queremos ahora describir los cursos multimedia interactivos que nosotros hemos desarrollado y asimismo queremos describir la herramienta que para su desarrollo hemos utilizado: Multimedia ToolBook Instructor II.

6. Multimedia ToolBook

Una de las principales dificultades con la que se encuentran los formadores a la hora de confeccionar su propio material, es la carencia de la competencia informática necesaria.

Para solventar dicha dificultad aparecen en escena los **lenguajes de autor** que son lenguajes de programación creados con el objetivo de que puedan utilizarse con poco entrenamiento y permitan crear aplicaciones en poco tiempo.

Uno de los programas de autor existentes actualmente en el mercado para la creación de cursos multimedia interactivos es **Multimedia ToolBook II Instructor.**

La primera versión de ToolBook fue desarrollada en 1985 por la compañía norteamericana **Asymetrix** que dirige Paul Allen,

quién en su día fuera cofundador junto con Bill Gates de la empresa Microsoft. [3]

La idea surgió del vacío existente en el mundo del PC compatible, en el que no existía una herramienta de programación de autor que permitiese a cualquier usuario crear potentes aplicaciones, como sí ocurría ya en el mundo Apple, donde Bill Atkinson y un grupo de desarrolladores habían creado un lenguaje de programación llamado **Hypercard**, con el que cualquier persona era capaz de desarrollar en su Macintosh una aplicación que incorporase hipertexto, que por aquel entonces, era un elemento completamente novedoso. [3]

El objetivo era por tanto, crear un software de desarrollo *de autor*, que permitiese, por tanto, a cualquier usuario de un PC compatible crear una aplicación que en un determinado momento necesitase.

Para diseñar la estructura de un programa creado con ToolBook, su creador pensó en la metáfora de un libro impreso.

Un programa creado con ToolBook recibe el nombre de **libro.**

Cada libro consta de un cierto número de **páginas**, cada una de las cuales viene a representar lo que una página de papel es a un libro impreso.

Estas páginas se muestran en ventanas ToolBook que reciben el nombre de **visores.**

Asociado a cada página o grupo de páginas se encuentra el fondo. En efecto, existen en cada página dos niveles:

- el **frente**
- el **fondo**

En el fondo se coloca todo aquello que es común a aquellas páginas que lo comparten y en el frente lo que es específico de cada una de ellas. Esto no significa que todas las páginas de un mismo libro deban tener el mismo fondo: ToolBook permite que en un mismo libro, existan varios grupos de páginas, de manera que en cada grupo las páginas compartan un mismo fondo diferente para cada uno de ellos.

A los elementos que forman una página se les llama **objetos** y pueden ser de diversos tipos :

⇒ **Gráficos.**

⇒ **Campos** que contienen información.

⇒ **Palabras Activas** que nos permiten hablar de hipertexto en vez de simplemente texto.

⇒ **Botones** que se sitúan en el interfaz de usuario para que éste inicie diversas acciones (ir a otra página...).

⇒ **Cuadros Combinados y Cuadros de Selección** que almacenan listas en las que cada elemento es una opción que se puede elegir.

⇒ **OLE** (Link and Embeding Object) que son almacenados guardando la información necesaria para acceder a la aplicación que los creó.

ToolBook posee herramientas que permiten fácilmente la construcción de cada uno de los objetos. Cada objeto tiene diferentes **propiedades** asociadas que definen su apariencia y comportamiento. Los objetos pueden a su vez agruparse dando lugar a un **grupo.**

Asociadas a un objeto, pueden existir un conjunto de instrucciones que forman lo que se denomina su **guión.**

Los guiones se escriben en un lenguaje de programación *de autor* del tipo *orientado a objetos* denominado **Open Script.**

El éxito de la programación orientada a objetos, tan de moda en nuestros días, hay que buscarlo en sus múltiples ventajas entre las que cabe destacar su aspecto modular. Esta programación, como su nombre indica, se basa en objetos - totalmente coherente por tanto, con la estructura de un programa realizado con ToolBook- cada uno de los cuales posee un conjunto de instrucciones.

En una aplicación creada con ToolBook, el flujo del programa no está previamente fijado, sino que lo marca el usuario al actuar sobre los objetos ocasionando **sucesos.** De esta forma, el control pasa de un guión a otro dependiendo de esos sucesos.

Cuando ocurre un suceso se envía el correspondiente **mensaje.** Un mensaje por tanto, puede decirse que es un anuncio o señal enviada al entorno ToolBook, indicando que algo ha sucedido o está sucediendo, por ejemplo que el cursor ha entrado en un botón, que el botón del ratón está pulsado... Estas señales, cuando se producen, son detectadas por el entorno que se encarga de buscar el conjunto de instrucciones que deben ejecutarse como consecuencia del suceso anunciado (constituyen lo que se llama el **ejecutor del mensaje o receptor de mensaje**).

Para hacernos una idea de lo potente que es Multimedia ToolBook señalaremos que permite

diseñar cursos que incluyan herramientas telemáticas en el interfaz de usuario, lo cual significa por ejemplo que cualquier objeto puede convertirse en un hiperenlace a cualquier URL (Uniform Resource Locator) del World Wide Web.

Finalmente señalaremos que a la hora de trabajar con un programa en ToolBook existen dos formas de hacerlo :

⇒ **Nivel Autor** que es el que permite construir un libro diseñando las páginas, creando y modificando los objetos y escribiendo los guiones en Open Script.

⇒ **Nivel Lector** que permite ejecutar el libro pudiendo navegar por sus diferentes elementos...

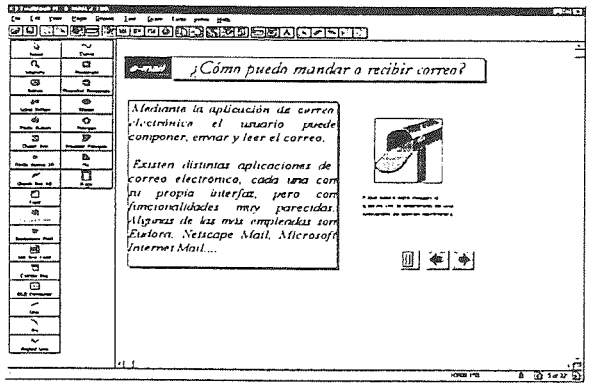


Figura 1: Aspecto de ToolBook cuando se trabaja en nivel de autor.

7. Los cursos que hemos realizado

Los cursos que hemos realizado están pensados para ser utilizados en combinación con el ATF (Asistente Telemático de la Formación) que la empresa Divisa Informática S.A. está desarrollando. Como consecuencia de este hecho, el diseño de la página se ha simplificado al máximo ya que las herramientas que permiten el contacto de los alumnos entre sí y con el profesor van integrados en el ATF.

Los cursos tienen la siguiente estructura:

- Introducción.
- Manual de uso.
- Indice.
- Contenidos.
- Glosario.
- Autoevaluación.

Queremos destacar que las preguntas planteadas en la autoevaluación son de diversos tipos. Algunas son del tipo verdadero/falso o de elección múltiple, pero en otras se exige que el alumno escriba sus propias respuestas. Además también se le plantean prácticas. Tanto en las preguntas como en las prácticas siempre existe una retroalimentación, cuya finalidad es orientar a los alumnos que no hayan logrado el éxito en el primer intento.

El diseño de página es distinto para cada una de las partes del curso anteriormente mencionadas, pero es el mismo independiente del curso del que se trate. A continuación mostramos el diseño de algunos de los tipos de página usados:

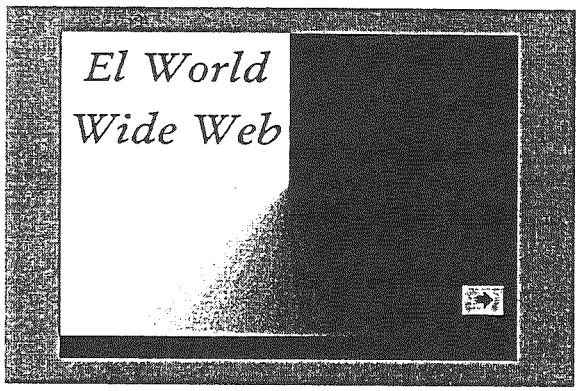


Figura 2: Diseño de página de la portada del curso sobre World Wide Web.

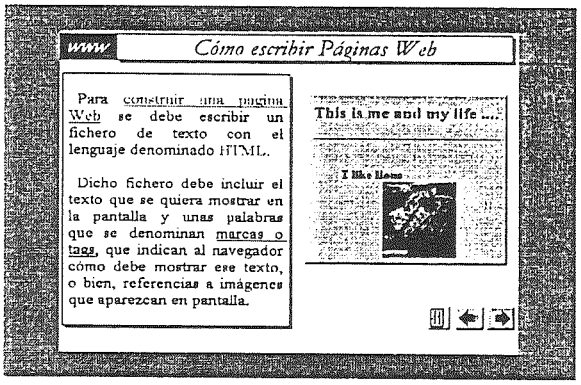


Figura 3: Diseño de página empleada en la parte de "Contenidos" del curso.

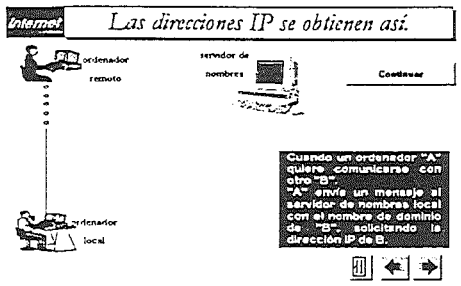


Figura 4: Diseño de página empleada en la parte de "Contenidos" del curso. Incluye Animación.

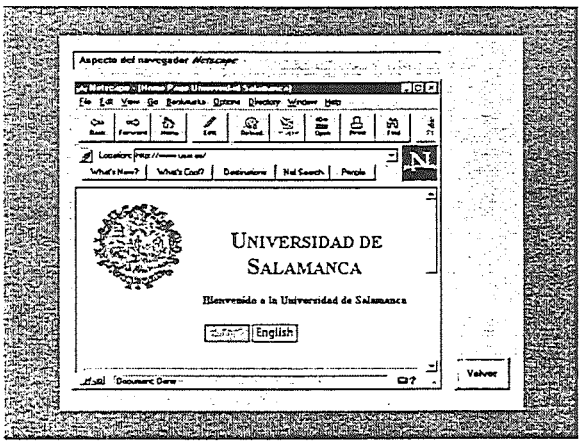


Figura 5: Diseño de página a la que se accede mediante una palabra activa. Desde aquí sólo se puede volver a la página desde donde se partió.

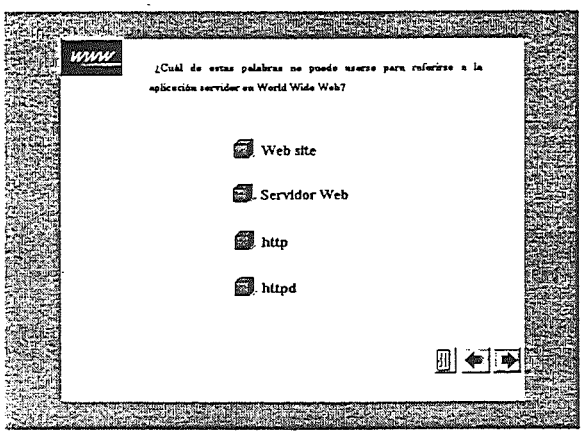


Figura 6: Diseño de página de la parte de "Autoevaluación", incluye una pregunta de múltiple elección..

Por último queremos mostrar algunos de los elementos que nos encontramos en una página tipo:

Botones de navegación:

- Salir. 
- Ir a la página anterior. 
- Ir a la página siguiente. 

Palabras o Imágenes activas y botones que constituyen hiperenlaces a:

- Otras páginas del mismo o de otro libro.
- Páginas del World Wide Web, recursos de Internet y programas.

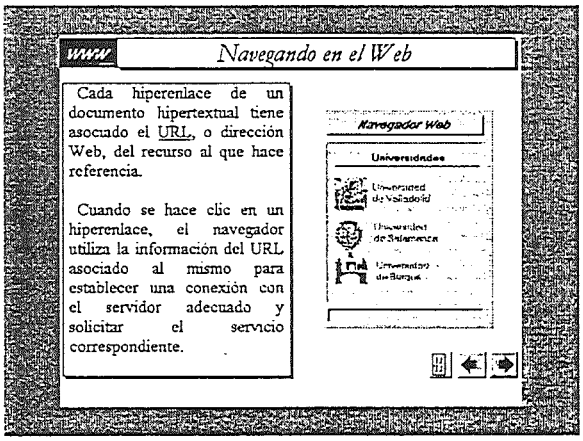


Figura 7: Las palabras escritas en rojo en la imagen de la derecha son palabras activas en las que al hacer click aparece una ventana con información adicional de un carácter más técnico.

8. Depuración de los Cursos

Los primeros cursos que hemos realizado versan sobre Internet y sus servicios y estaban destinados a empleados de PYMES "que no tuviesen conocimientos previos sobre el tema". Por ello, una vez tuvimos lista una primera versión de los cursos, organizamos una experiencia piloto en la que los alumnos eran profesores de la Facultad de Pedagogía de Salamanca.

Como resultado de dicha experiencia, la versión inicial de los cursos sufrió múltiples modificaciones. En concreto, se realizaron múltiples cambios en el texto, los gráficos, las animaciones e incluso en la estructura general de los cursos.

Todas estas modificaciones tenían como objetivo el mejorar la calidad de los cursos desde el punto de vista pedagógico. Hemos de tener siempre muy presente que, como ya hemos comentado, la enseñanza a distancia es radicalmente distinta a la enseñanza presencial.

Esto nos obliga a elaborar materiales que mediante el empleo de recursos multimedia en los

momentos apropiados logren mantener la atención del alumno hasta el final. Por otra parte, las herramientas telemáticas se vuelven imprescindibles siendo su misión el permitir la interacción del alumno con sus compañeros de curso y con su profesor.

Una vez tuvimos lista una segunda versión de los cursos, organizamos otra experiencia piloto, esta vez con algunos alumnos de doctorado de la Facultad de Pedagogía de Salamanca y en esta nueva experiencia pudimos comprobar que la TI nos permitía proporcionar una educación centrada en resolver problemas y buscar soluciones -en vez de en escuchar conferencias-.

También, como ya habíamos previsto, comprobamos que el profesor pasó, de ser un mero transmisor de información a ser un guía en el aprendizaje. El alumno por su parte, dejó de ser un receptor pasivo y se convirtió en el protagonista absoluto de su proceso formativo, alguien que activamente busca, procesa, analiza y evalúa hechos, gracias a las NTICs que le dan libertad para buscar aquella información que le interese y le permiten "aprender haciendo".

Además observamos que la formación mediante métodos que incorporaban la TI suponía ya una ventaja en sí misma, ya que la Nueva Sociedad de la Información, a la que está dando nuestra actual Sociedad Industrial, demanda cada vez con mayor intensidad capital humano diestro en el uso de las NTICs.

Como resultado de estas experiencias elaboramos un documento con una serie de consejos para la elaboración de cursos multimedia interactivos, algunos de los cuales recogemos a continuación:

⇒ **Lo simple es mejor.** No se debe abusar de los elementos multimedia, los cuales deben además repartirse equilibradamente por toda la aplicación diseñada. Debemos recordar que el mensaje es aún lo importante.

⇒ El **texto** debe ser breve -no debería ocupar más de un tercio de la pantalla-. Los ojos se cansan más leyendo en la pantalla, además, los bloques largos de texto le dan a esta un aspecto excesivamente estático. Asimismo, las frases deben ser breves y de estructura lo más sencilla posible (sujeto-verbo-predicado). Si se incluyen listas no ordenadas, cada una no debería tener más de unos seis puntos y cada punto no más de una o dos líneas.

⇒ No se deben usar muchos **tipos de fuentes diferentes**. Además, el tamaño y el estilo de la fuente usada debe emplearse para establecer una jerarquía. Se deben evitar asimismo **estilos, colores** y tipos de fuentes difíciles de leer (cursiva, azul o verde, Times New Roman).

⇒ Las **palabras activas** deben presentarse en un color llamativo y que resalte fácilmente.

⇒ En las **animaciones** en las que la aplicación vaya presentando textos secuencialmente, debemos elegir una secuencia de tiempos adecuada. En cualquier caso, debería incluirse un botón que permita repetir la animación y si esta es larga debería dividirse en partes que se vayan presentando al alumno a medida que éste las demande -al hacer click en un botón....-.

⇒ La **voz** debe sin duda utilizarse pues es uno de los recursos que obviamente diferencia a los libros de texto de los materiales multimedia.

⇒ Las **prácticas** no deben tener únicamente el objetivo de comprobar si el alumno recuerda la información que se le ha proporcionado, pues debe tenerse presente que se "aprende haciendo". Asimismo, la retroalimentación que se le proporcione al alumno que no ha sabido resolver la práctica, debe centrarse, no en hacerle ver que ha cometido un error, sino más bien en orientarle sobre qué estrategia debe seguir para obtener el éxito la próxima vez.

9. Uso de los cursos en la formación permanente en PYMES

Una vez depurado el producto lo utilizamos para impartir la parte correspondiente a Internet y sus servicios en el "Curso de Telecomunicaciones para la Mejora de la Competitividad de las Empresas Industriales". Este curso tuvo lugar en Cedetel que es el Centro para el Desarrollo de las Telecomunicaciones en Castilla y León y está situado en el Parque Tecnológico de Boecillo.

A dicho curso asistieron empleados de PYMES regionales. El curso se celebró a lo largo de seis sábados, a razón de cuatro horas cada sábado. Los últimos tres cuartos de hora se dedicaban a la evaluación de los conocimientos adquiridos por los alumnos en la sesión correspondiente. Esta evaluación se hacía mediante un examen y unas prácticas que primero realizaban los alumnos individualmente y que posteriormente se ponían en común.

Se observó que los alumnos podían en general resolver todas aquellas situaciones y preguntas que se les planteaban. Además notamos que la atención de los alumnos no disminuía a lo largo de la mañana, ni sábado tras sábado.

Durante tres meses los alumnos disfrutaron de conexión a Internet. La interacción entre el alumno y el profesor era presencialmente los sábados y mediante correo electrónico el resto de la semana. Nos sorprendió gratamente el ver que los alumnos interactuaban a lo largo de toda la semana con nosotros para plantearnos cuestiones, lo cual denotaba que se mantenía el nivel de motivación adecuado y que la posible sensación de aislamiento no era excesivamente importante, ya que el alumno sabía que contaba con una serie de herramientas para comunicarse con el profesor y el resto de los alumnos y hacía uso de ellas.

10. Conclusiones y Trabajos Futuros

Esta primera experiencia en el campo de la IDL ha sido para nosotros extremadamente positiva.

Entre nuestros proyectos futuros más inmediatos está el de realizar más experiencias de EAO utilizando los cursos realizados, no sólo en formación permanente para empleados de PYMES sino en actividad docente en general, ya que los cursos son apropiados para cualquier alumno adulto que posea habilidades lecto-escritoras.

En concreto lo vamos a emplear como apoyo a la docencia en las siguientes actividades:

⇒ Curso "Gestión Logística y Transporte" en la Cámara de Comercio Oficial de Valladolid en la primavera de 1998, para empleados de empresas del sector de Transportes.

⇒ Asignatura optativa "Introducción al empleo de las NTICs en procesos de formación autónomos". Esta asignatura con una carga docente de cinco créditos (50 horas lectivas) se impartirá en el segundo cuatrimestre del curso 97-98 en la Facultad de Pedagogía de Salamanca.

Además, queremos ver también si el documento de consejos para la elaboración de cursos multimedia interactivos que hemos redactado es o no una buena guía para dicho proceso. Para ello, profesores de distintas áreas generarán sus cursos apoyándose en tal documento. Estas experiencias han comenzado ya; sin embargo, aún es pronto para extraer conclusiones significativas.

Bibliografía

[1] Minoli, D., "Distance Learning Technology and Applications", Artech House, 1996.

[2] García L., *"La Educación a Distancia y la UNED"*, UNED, Madrid, 1996

[3] Mota, J. C., "Introducción a ToolBook y Multimedia ToolBook 3", Ra-ma, Madrid, 1996.

[4] Alvarez, E. y Alvaro, J. I. "ToolBook, crear multimedia con PC", Paraninfo, Madrid, 1996.

[5] Asymetrix, "A Guide to Creating Interactive Courses, Asymetrix ToolBook II Instructor", Asymetrix, Madrid, 1996

[6] Merill, M. D., "Don't Bother Me with Instructional Design, I'm busy Programming!. Suggestions for More Effective Educational Software", *Educational Technology Publications*, New Jersey, 1994.

[7] Alvarez, E., "Los Lenguajes de Autor en el Entorno Educativo. Experiencias diversas en en la U. Cantabria", Laredo, 30 de Agosto 1995.

[8] Ballesta J., *"Las Nuevas Tecnologías en la Enseñanza"*, Laredo, 30 de Agosto 1995.

[9] Fernández J. y Fernández L., *"Comunicación Multimedia Interactiva para Enseñanza a Distancia. Un medio para el diálogo"*, Laredo, 30 de Agosto 1995.

[10] Bartolomé A., *"Utilización de las nuevas Tecnologías en la Técnica Expositiva"*, Laredo, 30 de Agosto 1995.

[11] Martín P., *"Aplicaciones de la Telemática en Escenarios Educativos"*, Laredo, 30 de Agosto 1995.

NUEVAS FORMAS DE ENSEÑANZA EN INGENIERIA TELEMATICA

Manuel JUAN ESCRIVA
Fátima MARTI ADSUAR
Daniel PALACIOS MARQUÉS
Departamento de Sistemas Informáticos y Computación
Universidad Politécnica de Valencia
e-mail: palacios&dsic.upv.es

ABSTRACT

With this article, we have tried to carry out an analysis of the different aspects of the formation in the courses of Telematics, considering questions about the knowledge and its transmission. Our main aim has been to lay emphasis on the importance of the feedback of the obtained knowledge (they come from the artificial intelligence, visual languages and the design of interfaces). In this way, we seek to stand out the importance that for the student has to know deeply the knowledge of the systems that are developed or studied.

1.- INTRODUCCIÓN.

Las teorías más modernas nos conducen hacia el aprendizaje significativo y por descubrimiento. Se apoyan en el modelo constructivista, "el conocimiento lo construye cada persona" y el cognitivo "hay que aprender a aprender y descubrir cuales son los procesos del pensamiento".

Aunque los alumnos frecuentemente adquieren aquellos conocimientos que los profesores les transmiten de forma pasiva, se sabe que se aprende mucho más cuando uno se formula sus propias preguntas, las respuestas se buscan desde diversas perspectivas, se intercambian diferentes puntos de vista y se asocian los propios descubrimientos a los conocimientos ya existentes.

De acuerdo con el enfoque cognitivo, se deben dar a los alumnos las nociones que son necesarias para comprender cómo y por qué del proceso del conocimiento (metacognición). Aplicado a la formación este enfoque pone su énfasis en desarrollar habilidades del pensamiento y en comprender profundamente los procesos. Por ejemplo, algunas actividades que favorecen su desarrollo son:

- Propiciar la multiperspectiva en el estudio y enriquecer las diferentes vistas.
- Integrar el conocimiento de las diferentes áreas de estudio de un sistema.
- Hacer equivalencias y asociaciones de conocimientos.
- Mejorar las representaciones, y realzar la importancia del modelo simbólico y cualitativo.

El objeto de este artículo es mostrar los principales problemas relacionados con la docencia en informática y proporcionar pequeñas soluciones y ejemplos, a partir de los enfoques cognitivos anteriormente expuestos. En definitiva, se trata de poner en marcha en la actividad docente la acción de verbos conductuales de las taxonomías del conocimiento y de la comprensión , tales como: identificar, definir, redefinir, decir

de otra forma, traducir, reconocer, adquirir, reorganizar, distinguir, diferenciar, distinguir, completar, explicar, completar, determinar, predecir...

2.- CONSIDERACIONES GENERALES.

1.Desarrollo del razonamiento lógico.

La informática es por excelencia la ciencia que desarrolla el razonamiento lógico. Indicamos los factores que influyen en su desarrollo en la tecnología educativa: concentración , motivación, interés, reflexión, creatividad y expresión . Todos los factores comentados anteriormente se han de tener en cuenta, ya que todo ejercicio que los ejercite va a generar una serie de estrategias que se van a acumular en el cerebro y se van a aplicar sin ningún tipo de esfuerzo.

Como ejemplo vamos a referirnos, al desarrollo de un par de factores comentados anteriormente como son la expresión y la creatividad que llevan a la necesidad de buenas representaciones y diagramaciones en los sistemas de estudio.

2. *Feedback* en los conocimientos para la aplicación en formación.

Esta etapa tiene como objetivo la realimentación en los modelos, realizar diversos tipos de representaciones, estructuras de información, realización de algoritmos, análisis interdisciplinar, con el objetivo final de crear una mejor sinergia y descartar polos de conocimiento.

Como sobradamente es sabido, existe una disociación entre la formación que imparte el profesorado, entre los diversos métodos de implementación de sistemas y los conocimientos y metodologías que se usan en la exposición. El ejemplo clásico a lo comentado es el profesor de pedagogía que no enseña de forma pedagógica. Nos podemos preguntar ¿cómo se enseñan los modelos conceptuales?, si no tenemos en cuenta una mirada conceptual del problema. Las investigaciones cognitivas de los procesos del pensamiento se han desarrollado desde hace un tiempo en diversos campos de la informática, en donde ha habido más acercamiento a estas investigaciones cognitivas. Un ejemplo de estas investigaciones es el que se ha desarrollado dentro de la Inteligencia Artificial, y más concretamente dentro del campo de los Sistemas Basados en el Conocimiento (SBC).

Además de las aplicaciones citadas en Inteligencia Artificial, también hay que destacar por su gran importancia:

- Modelos cualitativos
- Modelos simbólicos
- Tutores inteligentes
- Diseño de interfaces
- Interacción hombre-máquina

La última aplicación actualmente se pone de relieve debido al interés generado en que el ordenador sea didáctico, inteligente, y ergonómico. Por lo tanto, en estos temas se pueden aplicar las realimentaciones. El concepto de realimentación de los conocimientos ha sido muy importante y productivo.

Un ejemplo bastante ilustrativo que explica lo comentado en el punto anterior lo constituye el hecho de que para crear un ordenador más inteligente y más cercano a los procesos de pensamiento que tiene el hombre, se han aprovechado ciertos estudios neurofisiológicos que se han realizado en el cerebro, y se han creado las redes y ordenadores neuronales, y debido a esto se han inventado algoritmos de comportamiento neuronal. Esto ha sido aprovechado por los psicólogos para explicar el comportamiento del cerebro a través de dichos modelos.

Una idea interesante para aprovechar la realimentación sería la de tener más en cuenta actividades de definición, adquisición, representación y formalización de los Sistemas Basados en el

Conocimiento, así como las herramientas que se han creado al efecto.

La idea que se pretende conseguir es la de hacer más expertos a los alumnos y que puedan manejar de una mejor forma estos sistemas expertos.

Otra idea interesante respecto a la realimentación, es la que proviene del campo de los diseños de interfaces Hombre-Computadora (H-C). La investigación que se ha realizado al respecto ha sido muy rica y ha llevado a la creación de novedosos tipos de análisis, serían por ejemplo: nuevos lenguajes visuales y S.O. tipo Windows, análisis como el A.O.O., y también los diseños cognitivos de interfaces de pantalla que están muy de actualidad con el fin de evitar o al menos minimizar los errores que el operador humano pudiese cometer.

En primer lugar, los modeladores se encargan de estudiar los sistemas y también de la tarea de crear modelos conceptuales para que posteriormente, los diseñadores puedan utilizar dichos modelos con el fin de implementar las interfaces cognitivas H-C que no lleven a error a los usuarios. De este modo, se consigue que los modelos implementados de los sistemas sean capaces de coincidir al máximo posible con los modelos mentales de los usuarios.

Se produce al mismo tiempo un proceso paralelo al descrito anteriormente. En el momento en que el profesor estudia los mismos sistemas con el fin de explicarlos, crea también sus propios modelos mentales y será su tarea fundamental el presentarlos a sus alumnos del modo más estructurado y conveniente que le sea posible. De lo que se trata con esto es de que los modelos mentales que el alumno se forma también coincidan con los del profesor.

Lo comentado hasta el momento de suma importancia, porque en el caso de que no se cumpliera tendríamos que lo recibido y lo asimilado por los alumnos no se parecería a lo transmitido. Al mismo tiempo, como lo que pretendemos en última instancia es acercarnos a la realidad, vemos que es necesario que las estudiadas anteriormente interfaces del usuario deberán también coincidir con las del profesor-alumno.

Con todo lo que acabamos de apuntar y de acuerdo con M.H. McLuhan cuando indica: "Nos convertimos en lo que contemplamos. La forma del portador de la información no carece de interés; al mismo tiempo dicta la clase de información aportada e influye en los procesos del pensamiento", podemos delimitar la figura del profesor. Consideramos pues al profesor como un ingeniero del conocimiento y al mismo tiempo como diseñador y modelador de la información.

En concreto, la figura del profesor en tanto en cuanto que ingeniero del conocimiento debe de reunir las siguientes habilidades:

- Efectivo comunicador.

- Conocer las técnicas de adquisición del conocimiento.

- Rápido aprendizaje del dominio en estudio.

- Tener buena atención y escucha.

- Disponer de experiencia en el diseño de bases de conocimiento.

- Ser muy organizado.

- Ser un buen conceptualizador.

- Tener buena memoria.

- Poseer un amplio conocimiento en diversas materias.

- Conocer las herramientas de software que son apropiadas.

Por otro lado, debemos contemplar las habilidades del profesor en cuanto a modelador y diseñador de la información. En función de la finalidad que se persiga, el profesor modelista suele considerar el sistema desde varios puntos de vista, y en el

momento en que modela establece un marco de observación, donde posteriormente va a enfatizar las características que son más importantes. Así pues, si en este punto obvia aquellas características que no considera como esenciales, más tarde las va a expresar a través de un conjunto de atributos que representa mediante variables. De este modo, al estudiar las causas y los efectos, tiene las observaciones ordenadas espacial y temporalmente y puede así organizar el conocimiento e intentar unificarlo.

3. Representaciones para el modelado de los problemas.

Llegados a este punto, es necesario es necesario emplear representaciones creativas basadas en el conocimiento.

Los modelos que existen para el modelado de los problemas, tienen altas propiedades cognitivas. Vamos a realizar un breve recorrido por modelos simbólicos más importantes y a la vez un comentario de sus principales posibilidades:

a. MODELOS CUALITATIVOS: son importantes desde la perspectiva de los avances en la representación del conocimiento y no desde el punto de vista de las simulaciones. Los modelos cualitativos, consiguen un conocimiento muy profundo de los sistemas.

b. MODELOS FUNCIONALES: para determinar las principales funciones del sistema necesitamos conocer los objetivos, el comportamiento, la estructura, y por último, el contexto del dispositivo. Estos modelos se caracterizan por mostrar la estructura, la función y el comportamiento de una forma conjuntada a través de niveles de representación. Se suelen especificar cuatro tipos de funciones: *hacer, mantener, prevenir* y *controlar.*

c. MODELADO ORIENTADO A OBJETOS: estos modelos se caracterizan por centrarse principalmente en los elementos del sistema, en sus parámetros más importantes y por organizar la influencia entre objetos. Podríamos decir que son modelos que representan la expresión inicial de un problema en términos de estructura, atributos y servicios. Son modelos connaturales totalmente al proceso jerárquico de conceptualización humana.

d. MODELOS DE COMPORTAMIENTO: en estos modelos a partir de la información común de que se dispone , lo que hacen es considerar múltiples vistas y luego relacionarlas entre sí para realizar una verdadera labor de integración.

e. MODELOS BASADOS EN TAREAS: el objetivo fundamental de estos modelos está en representar la estructura de las tareas basándose en los objetivos que se pretenden conseguir. Podemos definir varias categorías de tareas.

f. MODELOS CAUSALES: para realizar un diagrama causal es necesario que se realice un esfuerzo de análisis y de síntesis hecho que en sí mismo tiene un valor muy importante como técnica de integración del conocimiento.

4. Relacionar y unificar representaciones y modelos.

Cualquier informático en el desarrollo de su trabajo va a verse ante la necesidad de colaborar e integrarse con grupos multidisciplinares. Por ello el informático para poder comunicarse con los grupos implicados deberá entender y saber expresar un mismo problema de distintas formas.

De lo dicho hasta ahora, se desprende la necesidad de que el estudiante de informática se entrene en representar las soluciones desde distintas perspectivas o puntos de vista. El profesor por su parte, tiene la tarea de mostrar las analogías existentes entre los diagramas usados a lo largo de las distintas disciplinas.

5. Herramientas para la programación icónica-visual de funciones y de diagramación1.

Hay que tener presente las herramientas sencillas por cuanto son muy

útiles y además muy potentes en cuanto a la representación del conocimiento. Es importante no olvidarnos de estas herramientas y no despreciar el esfuerzo de otros programadores porque constituye un error muy común de profesores y alumnos en su afán de originalidad.

6. Equilibrio existente entre la implementación tipo software y hardware.

Es comúnmente utilizado en los diseños que se utilizan dentro del campo de la automática digital, ya que las soluciones implementadas en hardware y software se relacionan estrechamente. El profesor puede ejercitar al alumno para que pueda desarrollarse en ambas formas y motivarlo para que busque su propia solución, consiguiéndose de esta forma hacer que el alumno sea más creativo a través de una enseñanza inductiva.

7. Destacar errores y posibles diferencias entre los distintos diseños.

Las soluciones que se consiguen están obviamente condicionadas por las características implícitas de los problemas, las representaciones iniciales que se hayan elegido y las posibilidades de los lenguajes informáticos.

Normalmente, antes de comenzar a programar realizando una codificación propiamente dicha, se suelen utilizar pseudocódigos y diagramas de flujo, que nos ayudan a tener una mejor visión del problema pero por desgracia se termina cayendo en la rápida codificación, razonando más en cómo implementar que en las propias especificaciones del problema.

Los profesores se ven desbordados por listados interminables de los programas que realizan los alumnos, con lo que se pierde rigor en el análisis que éste puede hacer del programa, por lo que esto redunda en un perjuicio final para el alumno.

Los pseudocódigos ayudan en esta labor al no requerir tantas instrucciones para realizar una operación que las instrucciones que se podrían requerir en un lenguaje de alto nivel.

8. Enlace de distintas representaciones.

Para lograr una representación integrada, conviene que el alumno enlace los distintos modelos mediante uniones hipertextuales en los puntos de cada uno de ellos. Esto facilita una mejor comprensión global, ya que se nos permite navegar por los distintos planos de la información.

3- REFERENCIAS

[1] G. Fernández, F. Sáez, *Fundamentos de Informática.* Anaya multimedia, Madrid, 1995.

[2] D. W. Rolston, *Principios de I.A,* McGraw Hill, Madrid, 1990.

[3] J. D. Novack. *Aprendiendo a aprender.* Martinez Roca, Barcelona, 1988.

[4] Hudson K., *Enseñanza Asistida por Ordenador.* Ed. Díaz de Santos, 1986.

Herramienta SW para la simulación de un procesador con capacidad de conmutación de canales MIC.

S.G. GALAN (1), P.J. PEREZ (2), J.M. COLON
DEPARTAMENTO DE ELECTRÓNICA
E.U.P. DE LINARES, UNIVERSIDAD DE JAEN
ALFONSO X EL SABIO 28, 23700 LINARES (JAEN)
Correo electrónico: (1) sgalan@ait.ujaen.es, (2) pjperez@ait.ujaen.es

Abstract:

This paper describes an educational software to simulate the behaviour of a microprocessor which holds, in his features, the ability of switching MICs channels. Of course, this microprocessor is provided with a particular set of instructions and addressing modes. The software owns, too, an assembler (with all their directives included) and a loader.

1. Introducción

Se ha implementado un simulador de un procesador con capacidad de conmutación de canales MIC. La arquitectura del procesador, diseñada a tal efecto, permite un amplio juego de instrucciones; (entre las que se encuentran aquellas destinadas a la conmutación de canales MIC), así como una amplia gama de modos de direccionamiento. Igualmente el procesador permite los mecanismos de interrupción y de acceso directo a memoria, DMA.

El simulador incorpora un editor de textos para la realización de programas en un lenguaje ensamblador específico. Una vez escrito el programa, es ensamblado por el simulador y cargado en memoria. Posteriormente a estos pasos se puede ejecutar el programa de forma normal, paso a paso, e incluso introducir puntos de ruptura.

Evidentemente, el simulador ofrece resultados en pantalla, así como una visión de las zonas de memoria deseadas, del contenido de los registros, etc.

Como el simulador ha sido implementado para trabajar en entorno Windows (Win-3.11, Win-95 y Win-NT), dispone de un sistema de ayuda tanto para la programación de procesador, como para el manejo del simulador.

En realidad se trata de un paquete SW integrado que incorpora los siguientes elementos:

- Editor de textos.
- Conversor decimal <-> hexadecimal.
- Ensamblador, cargador y Simulador.
- Simulador de etapas de conmutación S-T.

2. Registros disponibles

Se utilizan para almacenar temporalmente direcciones, datos o instrucciones. El tamaño de los registros es de 16 bits, salvo el registro de instrucción que es de 32 bits. La arquitectura diseñada contiene los registros recogidos en fig. 1.

ACUMULADOR	0
PUNTERO DE PILA	1F4
REGISTRO DE ESTADO	0
REGISTRO DE INDICE	0
REGISTRO BASE	0
REGISTRO USO GENERAL	0
CONTADOR DE PROGRAMA	0
R. DE INSTRUCCION 1	0
R. DE INSTRUCCION 2	0

Fig. 1: Registros disponibles en la arquitectura.

3. Repertorio de instrucciones

La codificación de las instrucciones en memoria ocupa siempre dos palabras. La primera palabra contiene los datos necesarios para que el secuenciador entienda de que instrucción se trata, mientras que la segunda palabra corresponde al campo de dirección u operando.

El juego de instrucciones que maneja el procesador está construido por 38 abreviaturas nemónicas básicas. Las instrucciones pueden agruparse de la siguiente forma:

3.1 Instrucciones de transferencia de datos

Permiten el movimiento de datos entre registros y memoria. Aunque también se pueden englobar las instrucciones específicas para el maneo de la pila. Otro tipo de instrucciones que se pueden englobar en esta categoría, son aquellas que permiten cargar el contenido de los registros de índice, de base y el registro de estado.

3.2 Instrucciones aritméticas, lógicas y de desplazamiento

En lo relativo a instrucciones aritméticas, existen operaciones tanto en formato de coma fija como en formato de coma flotante.

Las operaciones lógicas consideradas son:

And, Or, Xor y Complemento a 1.

En cuanto a las instrucciones de desplazamiento, sólo se ha considerado el de un bit en ambos sentidos.

3.3 Instrucciones de control del programa y del sistema

Las instrucciones de control son utilizadas para alterar la secuencia normal de ejecución de un programa, mientras que las de sistema son capaces de modificar el funcionamiento del procesador.

3.4 Instrucciones de conmutación

Una de las características principales del procesador es que va a poder controlar una red de conmutación constituida por dos etapas (S-T).

Instrucción COM: Realiza la conmutación de un canal entrante perteneciente a un circuito MIC de entrada, hacia un canal saliente pertenenciente a un circuito MIC de salida. Esta instrucción, lógicamente tiene un formato especial que más adelante se tratará.

Instrucciones VI1 y VI2: Realizan un volcado a memoria los canales pertenencientes a los circuitos Mic 1 y Mic 2 de salida, respectivamente a partir de la posición de memoria indicada en el campo de operando.

3.5 Instrucciones de Entrada/Salida

Utilizadas, lógicamente, para el intercambio de datos con el exterior. Se han considerado sólo los datos numéricos en sus dos formatos: coma fija y coma flotante.

En la tabla 2 se puede observar el repertorio completo de instrucciones.

4. Modos de direcionamiento

Los modos de direccionamiento que se han considerado son los siguientes:

- Inmediato.
- directo.
- Indirecto.
- Relativo a Indice.
- Relativo a Base.

La tabla 2 muestra la forma de referenciar los modos de direccionamiento.

Tabla 1: Modos de direccionamiento.

DIRECCIONAMIENTO	SINTAXIS
Inmediato	#Op.
Directo	Por defecto.
Indirecto	iOp
Indexado	(Op)
Relativo a Base	[Op]

Tabla 2: juego de instrucciones

Transferencia de datos	
Instrucción	Operación
Alm	Mem ← (Reg)
Car	Reg ← (Mem)
Pus	Pila ← (Reg) P.Pila ← (P.Pila) +1
Pop	Reg ← Pila P.Pila ← (P.Pila) -1
Cri	R.I. ← (Reg)
Crb	R.B. ← (Reg)
Mrs	R. Estado ← (Reg)
Aritméticas	
Instrucción	Operación
Sum	Reg ← (Reg) + (De)
Res	Reg ← (Reg)- (De)
Mul	Reg ← (Reg) * (De)
Div	Reg ← (Reg) / (De)
Spf	Reg ← (Reg) + (De)
Rpf	Reg ← (Reg) - (De)
Mpf	Reg ← (Reg) * (De)
Dpf	Reg ← (Reg) (De)
Pac	Reg ← 0
Desplazamiento	
Instrucción	Operación
Dad	Desplz. Bit Derecha
Dai	Desplz. Bit Izquierda
Control de programa	
Instrucción	Operación
Par	Detener Ejecución
Sin	C.P. ← De
Ssn	C.P. ← De Si S=0
Ssc	C.P. ← De Si Z=0
Ssa	C.P. ← De Si A=0
Ssi	C.P. ← De Si Y=0
Sas	Pila ← (C.P) C.P. ← (De)
Ras	C.P. ← Pila
Control de sistema	
Instrucción	Operación
In1	Pila ← (C.P) Pil ← (R. Estado) C.P. ← M(1)
In2	Pila ← (C.P) Pil ← (R. Estado) C.P. ← M(2)
Rte	R. Estado ← Pila C.P. ← Pila
Lógicas	
Instrucción	Operación
And	Reg ← (Reg) and (De)
Orn	Reg ← (Reg) Or (De)
Ore	Reg ← (Reg) xor (De)
Cmp	Reg ← (Reg)
Conmutación	
Instrucción	Operación
Com	Conmutación de canales
VI1	Volcado a mem. De MIC 1
VI2	Volcado a mem. De MIC 2
Entrada/Salida	
Instrucción	Operación
Mte	Reg ← entero por teclado
Mtp	Reg ← flotante por teclado
Ime	pantalla ← (De) entero
Imp	pantalla ← (De) flotante

5. Directivas del ensamblador

Lógicamente, se ha implementado un

ensamblador específico para el procesador diseñado, y éste consta de las siguientes directivas:

ORG: Utilizada para indicar la dirección de memoria donde se cargará el código que le sigue

PIL: Se utiliza para indicar hasta que posición de memoria llega el bloque de memoria reservado para la pila.

CEN: Esta directiva se utiliza para asignar a una etiqueta un valor entero. Es por tanto una definición de una constante tipo entero.

CPF: Esta directiva es utilizada para asignar a una etiqueta un valor en coma flotante.

REM: Esta directiva sirve para reservar posiciones de memoria.

6. Memoria principal

El tamaño de la memoria principal es de 65536 posiciones de memoria de 16 bits cada una de ellas. En cuanto al mapa de memoria, las cuatro primeras posiciones, aquellas que van de la posición 0 a la posición 3, están reservadas para cuatro vectores de interrupción.

Desde la posición de memoria 4 a la posición de memoria que indique el programador mediante la directiva PIL.

Esta zona de memoria es la que pueden utilizar los programadores para cargar los programas que posteriormente serán ejecutados por el simulador. Evidentemente comienza donde acaba la zona reservada para la pila, y llega hasta la última posición de memoria.

En Fig. 1, se puede apreciar gráficamente el mapa de memoria.

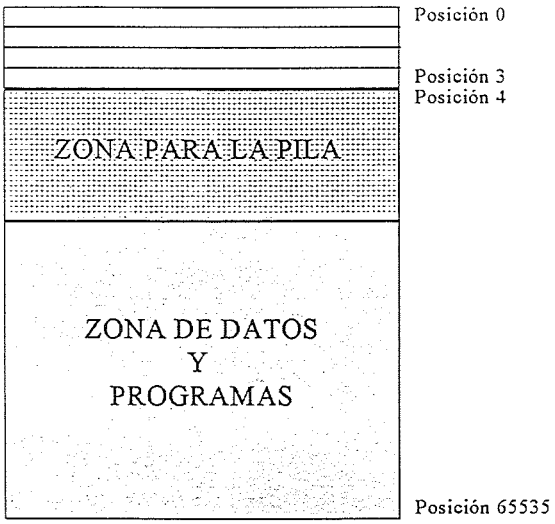


Fig. 1: Descripción memoria principal.

7. Arquitectura del procesador y etapas de conmutación

La arquitectura que permite todo lo expuesto hasta ahora, al nivel de transferencia de registros (RT), es la que se muestra en Fig. 3. Para favorecer una mejor comprensión, se ha evitado unir a la arquitectura de Fig. 3, la estructura de la red de conmutación de dos etapas S-T, la cual se muestra en Fig.3.

La red de conmutación de canales MIC de dos etapas S-T, está formada (ver Fig. 2) por dos etapas de conmutación, una etapa de conmutación espacial (S); donde el canal cambia de circuito pero no de rango, y otra de conmutación temporal (T); donde el canal cambia de rango dentro del mismo circuito. De esta forma puede conmutarse canales entre distintas tramas MICs (etapa S) y entre distintos intervalos temporales (etapa T); ver Fig. 4.

Como puede observarse en Fig. 2, sólo existen 2 Circuitos MICs de entrada y dos de salida, por lo tanto el dimensionamiento de esta estructura será el siguiente:

- Memorias Tampón: 32 posiciones de 8 bits cada posición.
- Memorias de control Temporales: 32 posiciones de memoria de 5 bits cada posición.
- Memorias de control espaciales: 32 posiciones de 2 bits cada una de ellas.

La posibilidad de bloqueo que presenta esta estructura de conmutación es la siguiente:

"No se pueden conmutar canales del mismo rango a la entrada al mismo circuito MIC de salida"

Es decir, con esta estructura, no se puede conmutar el canal 5 de los MICs entrantes 1 y 2 a los canales 10 y 11, respectivamente, del MIC saliente 1.

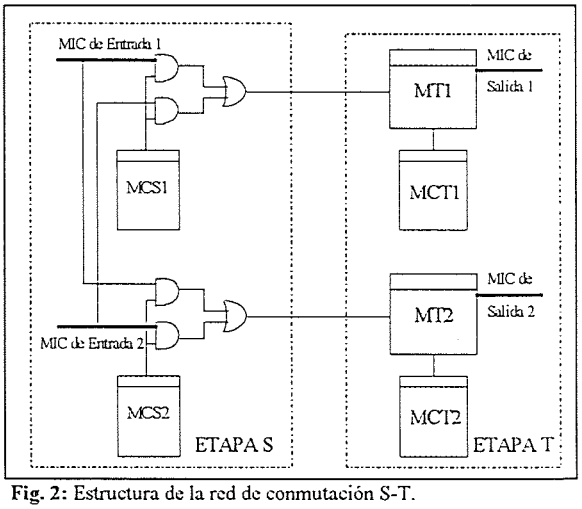


Fig. 2: Estructura de la red de conmutación S-T.

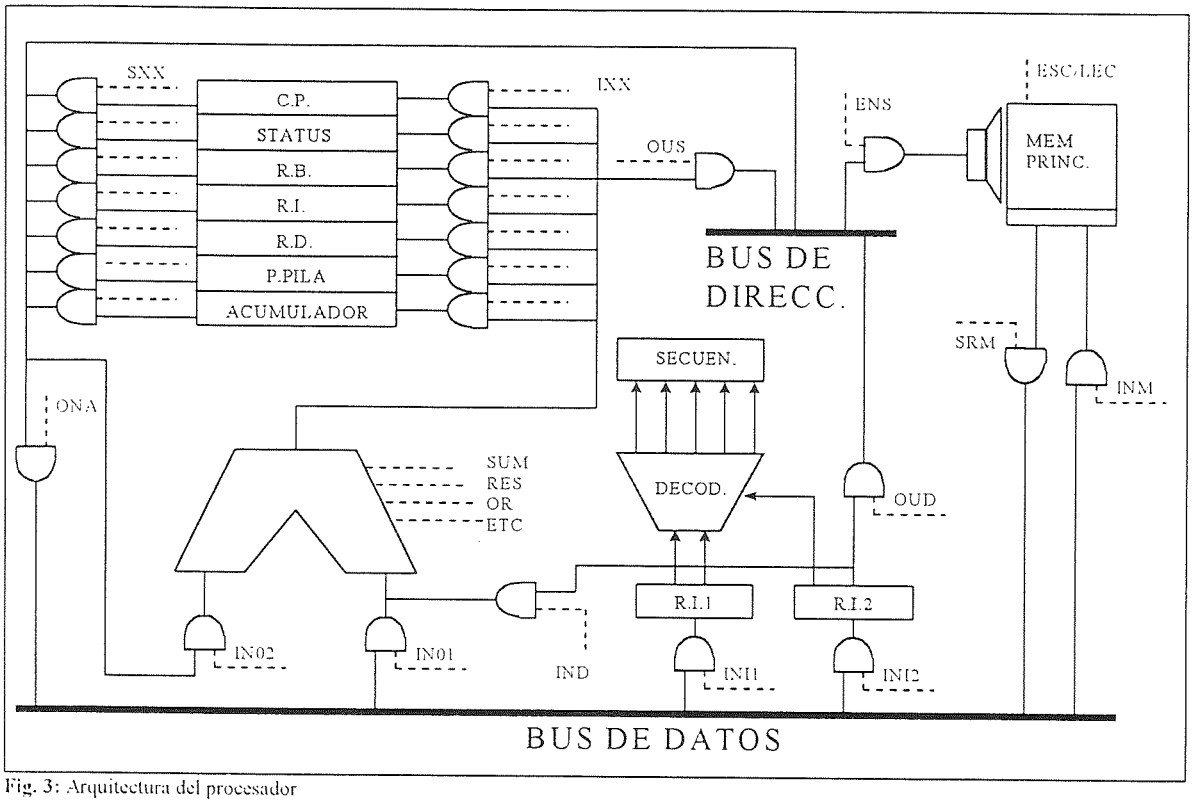


Fig. 3: Arquitectura del procesador

8. Ejemplo:

En este ejemplo se expondrán los resultados obtenidos al realizar la conmutación de dos canales MICs, concretamente del canal 5 del MIC 1 con el canal 8 del MIC 2.

Para una mejor visualización del efecto de la conmutación sólo tomarán valores distitntos de cero los canales implicados en la conmutación. El canal 5 del MIC 1 tomará el valor "FF" y el canal 8 del MIC 2 tomará el valor "AA".

Lógicamente, al finalizar la conmutación, el canal 5 del MIC 1 debe tener el valor "AA", y el canal 8 del MIC 2 debe tener el valor "FF".

El contenido de la memoria de las memorias de control de las etapas espaciales es el siguiente:

MIC 1
Posición 5: contenido = 2.
Resto de posiciones: contenido = 1.

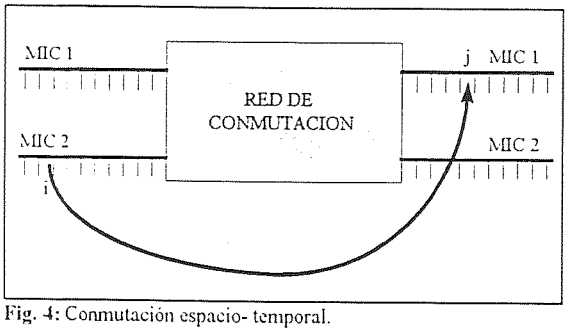


Fig. 4: Conmutación espacio- temporal.

MIC 2
Posición 8: contenido = 1.
Resto de posiciones: contenido = 2.

El contenido de las memorias de control de las etapas de conmutación temporales es el siguiente:

MIC 1
Posición 8: contenido = 5.
Resto de posiciones: contenido = nº de canal.

MIC 2
Posición 5: contenido = 8.
Resto de posiciones: contenido = nº de canal.

Para ilustrar los resultados de este ejemplo se presentan a continuación las salidas de los canales de ambos MIC a la salida de las etapas.

Tabla 3: Canales de salida.

CANAL	MIC 1	MIC 2
00	0	0
01	0	0
02	0	0
03	0	0
04	0	0
05	AA	0
06	0	0
07	0	0
08	0	FF
09	0	0
...	...	---
31	0	0

9. Conclusiones

Puede decirse que la herramienta implementada es una ayuda importante en la enseñanza tanto en la conmutación de canales MIC, como en la programación en ensamblador.

En lo relativo a la arquitectura de procesadores, ayuda a la comprensión del funcionamiento de los modos de direccionamiento, del los distintos tipos de instrucciones, del funcionamiento de la pila, de los mecanismos de interrupción; (experimentalmente se ha comprobado que este es un aspecto en el que hay que hacer especial énfasis); mecanismo de DMA, etc.

Es también digno de mención el aporte didáctico de la herramienta a la comprensión del funcionamiento de las operaciones en punto flotante, y a los problemas de redondeo en la representación.

En lo referente a la conmutación de canales MIC, es importante resaltar el buen comportamiento que tiene la herramienta implementada para la comprensión de la probabilidad de bloque que tienen, de forma implícita, las etapas de conmutación S-T, permitiendo la posterior extensión del concepto de bloqueo a otro tipo de redes de conmutación.

En cuanto a las posibles vías abiertas, cabe destacar la ampliación a más redes de conmutación. También la posible implementación real del procesador es una vía de interés.

Agradecimientos

Queremos agradecer a Mario Fernández Pantoja sus siempre interesantes observaciones.

Referencias

[1] Colón Mendoza. "Diseño e implementación de un procesador con capacidad de conmutación, a nivel R-T". PFC. E.U.P. Linares (1997).

[2] G. Fernández, F. Sáez Vacas. "Fundamentos de los ordenadores". Publicaciones E.T.S.I.T. UPM. (1985).

[3] Beck. "Software de sistemas". Addison-Wesley. (1988).

[4] W. Stallings. "Computer organization and architecture". Prentice-Hall. (1996).

Modelo WWW en la Enseñanza

WebTutor. Enseñanza adaptativa a través de WWW

Julián Gutiérrez, Tomás A. Pérez, José A. Carro, Iñaki Morlán & Philippe Lopistéguy
GRUPO DE HIPERMEDIA Y MULTIMEDIA
DEPARTAMENTO DE LENGUAJES Y SISTEMAS INFORMÁTICOS
FACULTAD DE INFORMÁTICA, UNIVERSIDAD DEL PAÍS VASCO / EUSKAL HERRIKO UNIBERTSITATEA
APTDO. 649, 20080 SAN SEBASTIÁN, GIPUZKOA, ESPAÑA.
Correo electrónico: <tomas, gutierrez, boti, jiplopop>@si.ehu.es
http://www.ji.si.ehu.es/hyper/
TFNO: +34 (43) 21 80 00 FAX: +34 (43) 21 93 06

Abstract:

This paper presents WebTutor, an Adaptive Hypermedia System for education purposes based on World Wide Web Technology. It is highly known that Internet is a technological phenomenum that has become part of end users. Networks are quickly increasing their bandwidth to become the utopic information highways. The World Wide Web can be accessed from almost everywhere. Its technology can also be used without being connected to the Internet and also disregarding the platform one is using at the moment. This versatility attracts a lot of work including educational applications.

Our work uses the representation of hypertexts used by the WWW to provide the students an intelligent learning environment that adapts the hyperspace to their ability either in navigation through the pages or their cognitive flair. The system has some advantages, first the output generated by the system is text, avoiding the need to include window managers that make the programming task harder and without avoiding the possibility of including images, animations, movies, etc. Besides, since the system behaves as a normal hypertext, the usability is higher because the student does not have to understand another brand new system.

The pages are enhanced with some modules that provide exercises to the student, evaluate them, evaluate the navigation, the knowledge acquisition, maintain a user model and adapt the hyperspace according to that model.

Finally, the system has given us the possibility of using distributed programming. A program (client) where the client can interact with a number of modules that can reside (probably) in different machines, defining interaction protocols between them.

1. Introducción

Internet está siendo, sin duda, el fenómeno tecnológico de más envergadura de finales del siglo XX. Desde su inicio como una red de investigación y de uso militar, ha pasado a convertirse en la auténtica precursora de las utópicas superautopistas de la información por donde se transmiten imágenes en movimiento, dibujos, sonidos, voz y por supuesto una cantidad tal de datos que en breve tiempo superará el tráfico telefónico existente [4] [18].

Hace algún tiempo, podía leer en un artículo sobre un sistema de información turística y cultural sobre WWW una frase profética [1]: *"Internet hoy es una realidad, y mañana será una necesidad"*. Efectivamente, por un lado estamos asistiendo al crecimiento de la interconexión de las redes informáticas, ya sean locales, metropolitanas o mundiales. Paralelamente, los conductos principales entre las diferentes redes están aumentando su capacidad de trasvase de información de manera que cada vez contamos con conexiones más eficientes. En España, por ejemplo, en 1995 se firmó un convenio entre Telefónica y la Comisión Interministerial de Ciencia Y Tecnología (CICYT) [8] para dotar en el plazo de dos años a RedIris, el proveedor de interconexiones de redes para instituciones científicas españolas, de una infraestructura de comunicaciones más avanzada.

Por otro lado, la tecnología ha permitido la inclusión de todo tipo de informaciones multimedia en el ordenador, dotándole de más expresividad y capacidad de comunicación. Así, ya no es extraña la utilización de imágenes, videos, animaciones, etc. y la aparición cada vez más numerosa de utilidades para crearnos nuestra propia aplicación multimedia. De entre todas las aplicaciones multimedia destaca una por la relevancia que se le da a la manera de organizar y acceder a la información: los sistemas *Hipermedia*. En un documento hipermedia, pueden aparecer enlaces entre ideas que pueden estar reflejadas en otro documento o en el mismo documento, pero no necesariamente a continuación. Es el propio usuario el que decide cuál es el orden de recorrido por la información dependiendo de los enlaces que siga. Esta organización de la información es adecuada para la creación de aplicaciones para la educación, ya que el alumno puede decidir cuáles son los aspectos que quiere aprender y en qué orden quiere hacerlo, dándole mucho margen de libertad. Esta popularidad es debida, en parte, a una de sus características principales: la flexibilidad, que permite que el usuario obtenga la información que él quiera en el orden que prefiera deambulando a través del hiperespacio intuitivamente.

Lamentablemente, si pensamos en dar a estos sistemas un uso educativo vemos que no son adecuados para tal fin, asociados básicamente a la orientación del alumno entre una gran cantidad de nodos [14]. En primer lugar, la organización de los nodos no suele ser evidente para alguien que está aprendiendo de la información almacenada en dichos nodos y se puede "perder" en la cantidad de información disponible. En segundo lugar por su escasa adaptación al usuario: la información que contienen los hipermedia tradicionales es fija, esto es, no depende de las características del usuario, ni del conocimiento adquirido.

Estos problemas se solucionan en gran medida haciendo que los hipermedia se adapten al usuario que los está utilizando proporcionándoles información adecuada a su grado de absorción de conocimientos e incluyendo todo tipo de herramientas que faciliten la navegación a través de los nodos. Para ello, se puede incluir *scripts* que modifiquen el comportamiento del hipermedia dentro de los enlaces entre los diferentes nodos. Esto hace que las estrategias pedagógicas estén distribuidas a lo largo de todo el documento y suponga una tarea repetitiva y de gran complejidad la modificación de los contenidos [9].

La tecnología hipermedia y las redes de ordenadores quedan integradas cuando se habla del proyecto de la telaraña mundial hipermedia [1], más conocida por su acepción inglesa, World Wide Web, o por sus siglas, WWW, Triple W o W3. En ella se permite la creación de manera sencilla de documentos hipermedia que incluyen todo tipo de informaciones y repletos de enlaces entre ellas. En estos documentos, los enlaces no tienen por qué referirse a documentos que residan en la misma máquina sino que se pueden establecer entre documentos que puedan estar en cualquier otra parte del mundo con la única condición de que estén accesibles a través de la red.

En este artículo utilizamos HTML [6], la base de los hipermedia en WWW para construir documentos hipermedia para la educación. Además, el sistema hipermedia se convierte en adaptativo mediante la colaboración de un sistema tutor con el hipermedia, dando lugar al sistema WebTutor, que se engloba dentro de lo que ha dado en llamarse Sistemas Hipermedia Adaptativos (SHA) [4][5].

En WebTutor se produce una simbiosis muy interesante entre las partes que lo componen. El diseño del sistema ha sido realizado de manera que tanto la parte Tutor como la parte Hipermedia sean perfectamente identificables e incluso separables. Cada una de ellas se beneficia de las funcionalidades específicas que le proporciona la otra. De esta forma, la parte Tutor se beneficia de la flexibilidad y utilización de distintos medios audiovisuales que motivan al alumno y proporcionan distintas formas de presentar la información que provee la parte Hipermedia. Y esta última se beneficia de la adaptación al alumno que realiza el Tutor haciendo que el resultado sea un sistema más educativo. Viéndolo desde otro punto de vista esta simbiosis también ayuda a superar los principales problemas de cada una de las partes del sistema. Así, el sistema Tutor consigue no controlar ni dirigir excesivamente su instrucción, problema que se achaca a muchos de los Sistemas Tutores que existen en la actualidad y que desmotiva completamente a aquellos alumnos que prefieren una instrucción más libre y autónoma. La libertad de navegación sobre la información que proporciona la parte hipermedia resolverá este problema. Y por otro lado, la parte hipermedia resuelve los problemas que se le presentan, como la pérdida del alumno en el hiperespacio, gracias a la actuación del Tutor que adapta las zonas de navegación a las características concretas de cada alumno y restringe su acceso a zonas en las que se encuentre cómodo y no se pierda. Así se proveerá de mayor dirección a aquellos alumnos que presenten mayor tendencia a perderse.

Por lo tanto, WebTutor reúne las ventajas de los Sistemas Tutores Inteligentes y de los Hipermedia: adapta el hiperespacio disponible dependiendo del conocimiento del alumno y, a medida que éste va aprendiendo, la accesibilidad crece, dándole la oportunidad de alcanzar nuevas informaciones. Como se puede observar en la Fig. 1, el sistema se estructura en dos partes bien diferenciadas: el *Componente Hipermedia* y el *Componente Tutor*. En los siguientes apartados se describirá brevemente la arquitectura del sistema.

2.- El componente hipermedia

El Componente Hipermedia se basa en MADAME, un sistema hipermedia orientado a objetos descrito en [22] y como se puede ver en la Fig. 1, está compuesto por tres módulos diferentes: el módulo interfaz, la hiperbase y el control de la navegación. Todos ellos se describen a cointinuación.

2.1. Módulo interfaz

Este módulo se encarga de presentar los nodos y obtener las reacciones del usuario. Implementado por cualquier browser que soporte HTML. La interacción se realiza mediante el uso de enlaces dentro de la descripción de las páginas.

Esta elección tiene asociadas ventajas y desventajas. Por un lado, la descripción de las páginas multimedia se reduce a un fichero de texto que incluya referencias a los contenidos multimedia que contenga. Además, la gestión de las selecciones del usuario son gestionadas por el propio browser de

manera que el sistema global, por parte del usuario sólo necesita de un programa de amplia difusión que está al alcance de cualquier usuario de ordenadores con la tecnología de Internet evitando así la programación de un gestor de eventos y de ventanas capaces de mostrar y manejar ficheros multimedia.

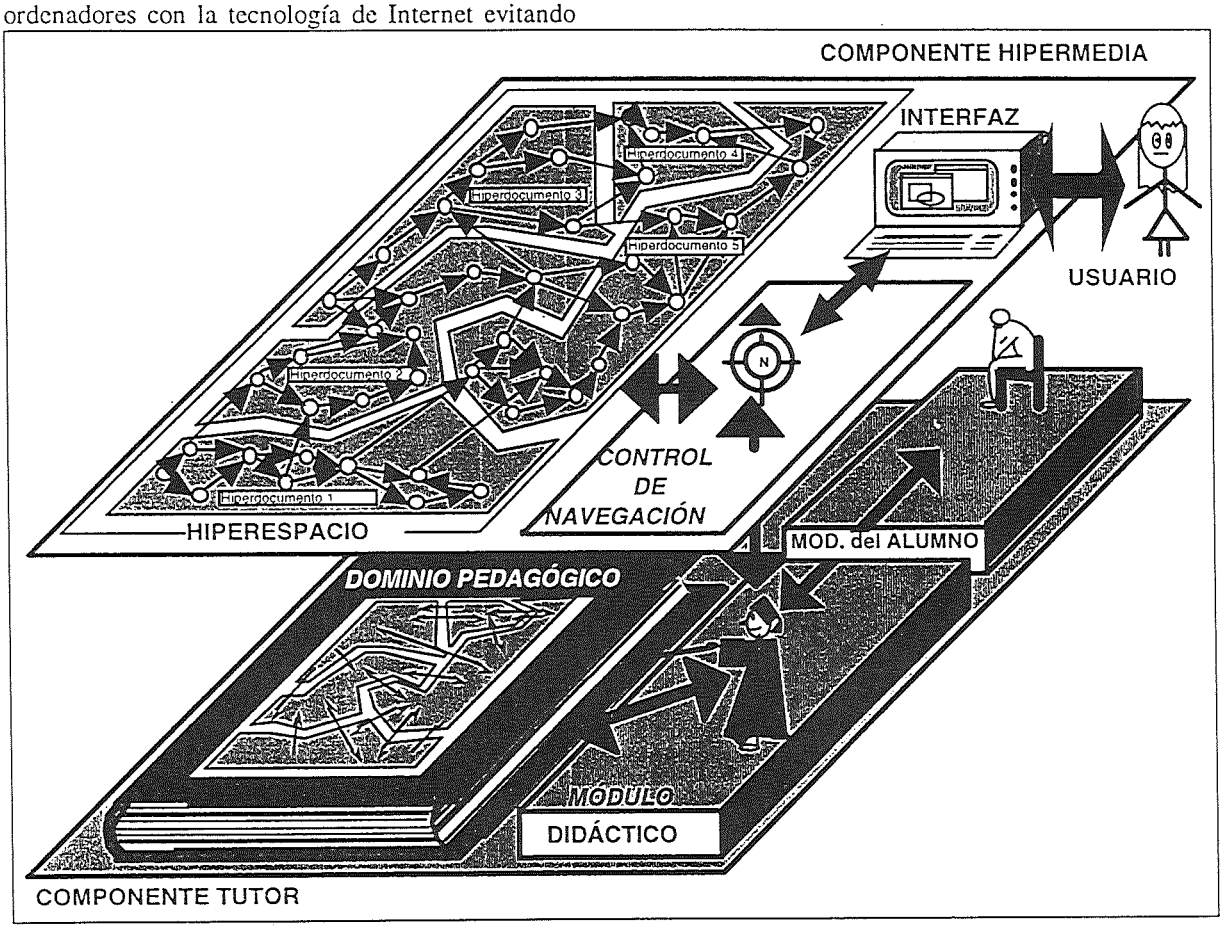


Fig. 1. Arquitectura del sistema WebTutor

Aunque el usuario no dispone de toda la funcionalidad que pudiera tener un hipermedia, al menos se puede suponer con gran que la interfaz es conocida y que el usuario no necesita un periodo de aprendizaje del nuevo sistema. Además, es fácil la descripción de plantillas a utilizar por el sistema hipermedia, e incluso la inclusión de programas específicos en lenguajes de script (como java o javascript) asociados a ellas para implementar algún comportamiento necesario como, por ejemplo, corregir una página con un ejercicio que se le presente a un alumno.

2.2. Hiperespacio

Está almacenado en una base de datos llamada *hiperbase* y contiene la información que se le va a presentar al alumno. Básicamente está formado por *nodos* (unidades de información a presentar al alumno) y *enlaces* (que relacionan diferentes nodos, estableciendo caminos entre ellos). Se distinguen varios tipos de nodos: *nodos de presentación de contenidos* y *nodos de ejercicios.* El hiperespacio está organizado en hiperdocumentos (Fig. 1). Un hiperdocumento es un conjunto de nodos y enlaces que componen una unidad conceptual de información. El Tutor influirá, haciendo uso de la información que dispone acerca del alumno concreto, sobre la forma de utilización del hiperespacio diciendo en cada momento qué parte está accesible, es decir, qué hiperdocumentos son los que puede ver el alumno. Los hiperdocumentos no son fijos. Dependiendo del tipo de alumno, el tutor establece las dimensiones de dichos documentos (tal y como veremos en el apartado 3.2.1). Por otro lado, los contenidos de los nodos de ejercicios son generados dinámicamente atendiendo a las indicaciones del tutor, basadas en el tipo de alumno y el recorrido del hiperespacio realizado.

2.3. Módulo de control de la navegación

Este módulo se encarga mantener comunicación con el Componente Tutor, ya sea para informarle de los movimientos que realiza el alumno, como para recibir de éste instrucciones para readaptar el hiperespacio al alumno (Fig. 1).

Para realizar este cometido, ha sido necesario crear un protocolo de comunicación entre los dos componentes de manera que sea simple el intercambio de información. La comunicación se hace a través de sockets, permitiendo así que la colocación en diferentes máquinas si fuese requerido de ambos componentes.

3.- El Componente Tutor

Este componente mantiene toda la actividad inteligente del sistema. Se encarga, además, de realizar un seguimiento de la interacción del alumno evaluando los conocimientos que éste adquiere a mediante ejercicios. Con esta información se decide cual será la accesibilidad del alumno dentro del hiperespacio, diciendo qué conocimientos estarán disponibles y cuáles no.

Siguiendo la arquitectura clásica de un sistema Tutor, se encuentra dividido en tres módulos principales: Dominio Pedagógico, Módulo Didáctico y Modelo del Alumno (Fig. 1), que describiremos en los siguientes apartados.

3.1.- Dominio pedagógico

El dominio no sólo ha de representar los conceptos a aprender por el alumno, sino que debe estar organizado de forma que la enseñanza de los mismos resulte sencilla, clara y eficaz. En el proceso de instrucción resulta necesario conocer el "orden" en que se van a presentar los conceptos, las relaciones que existen entre ellos, de que forma ayudan esas relaciones en el proceso de instrucción, la dificultad de aprendizaje de ese concepto, sus prerrequisitos, los diferentes puntos de vista en que puede ser expresado y explicado un concepto, etc. [11]. Por tanto existe gran cantidad de información que debe estar asociada a cada concepto o cada grupo de conceptos, que se debe conocer a la hora de enseñar y que debe estar plasmada en la descripción del dominio a enseñar: el Dominio Pedagógico.

Para organizar este dominio hemos seguido la aproximación utilizada en [11] definiendo los conceptos a enseñar, sus niveles de dificultad y las relaciones pedagógicas que existen entre estos conceptos, como ya hicimos en HyperTutor, su sistema precursor [16] [17]. La Fig. 2 muestra una pequeña parte del Dominio Pedagógico creado para la aplicación que se ha desarrollado para HyperTutor para enseñar Dibujo Técnico Industrial. En ella se puede ver algunas de las relaciones existentes entre los conceptos y la dificultad asociada a cada uno de ellos. Por ejemplo, el concepto *generalidades* está relacionado con otros conceptos (*condiciones*, *objetivo* y *clasificación*) por la relación *parte-de* que hará que para considerar entendido el concepto se tengan que haber entendido cada una de las partes. Además, estos conceptos están relacionados entre sí mediante la relación *siguiente*, que indica cuál debe ser el orden de visita. Los relacionados con la relación *es-un* se pueden visitar de manera desordenada, etc. Así cada una de las relaciones tiene asociada un comportamiento pedagógico por parte del Tutor.

Además, cada uno de los conceptos del dominio posee asociadas referencias a informaciones sobre ejemplos y ejercicios. Los ejemplos servirán para clarificar el concepto y los ejercicios para evaluar los conocimientos del alumno sobre el mismo. Los ejercicios se elaborarán a partir de preguntas (ítems de evaluación) de tipo test de selección múltiple [18].

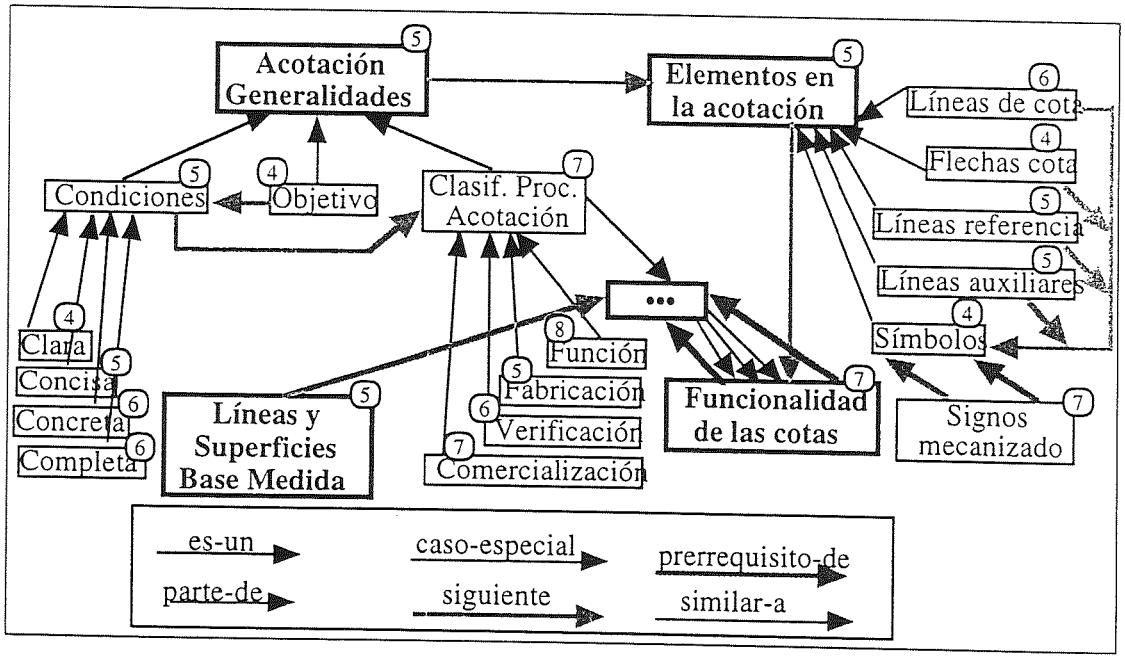


Fig. 2. Pequeña parte del Dominio sobre Dibujo Técnico Industrial

Para explicar brevemente este paralelismo entre el Dominio Pedagógico y el Hiperespacio hemos construido la figura 3. El contenido de un concepto del Dominio Pedagógico será presentado por uno o varios nodos del hiperespacio (ilustrado en la figura para los conceptos *clara*, *condición*, *verificación* y *fabricación*) utilizando el material y los medios audiovisuales que se hayan creído convenientes a la hora de diseñar dicho hiperespacio. Un mismo nodo del hiperspacio puede presentar información de distintos conceptos del Dominio Pedagógico. Como se puede ver en la figura, los conceptos *fabricación* y *verificación* comparten un nodo del hiperespacio. Así mismo, cada nodo posee un porcentaje de representación sobre la cantidad de conocimiento correspondiente a el/los concepto/s a el/los que aparece asociado.

La información contenida en el Dominio Pedagógico, junto con la información disponible sobre el alumno (en el Modelo del Alumno), será utilizada por el Tutor para comunicar al Componente Hipermedia cuáles deben ser los partes del hiperespacio accesibles al alumno, tal y como veremos en el apartado siguiente.

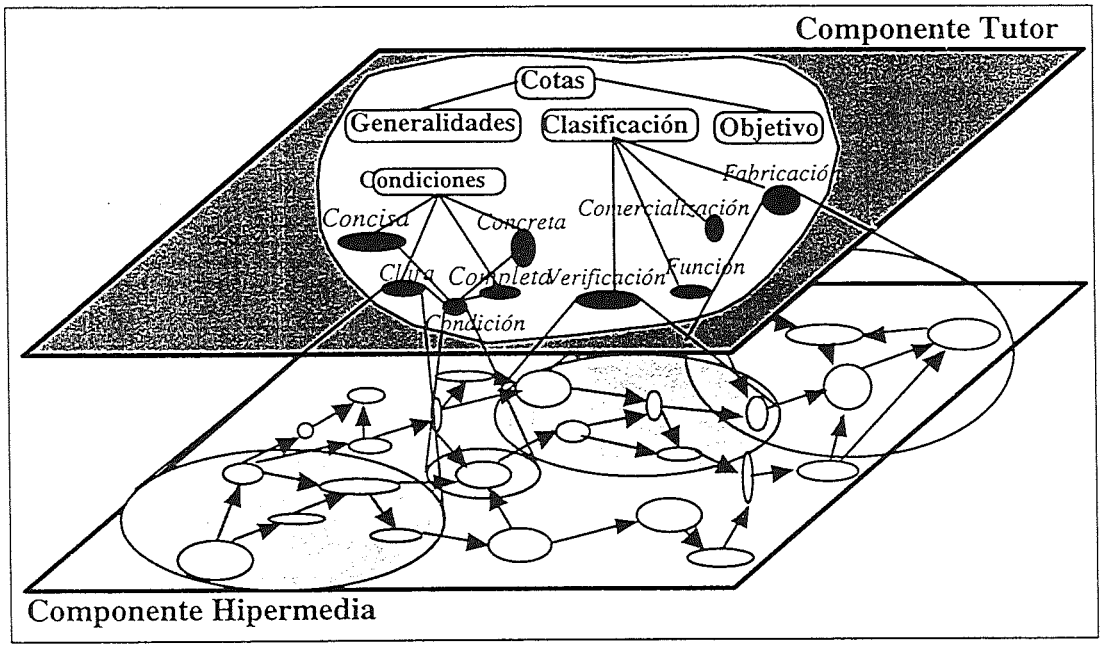


Fig. 3. Correspondencia entre los conceptos del Dominio Pedagógico y los nodos del Nivel Hipermedia

3.2.- Módulo didáctico

Este módulo es el cerebro del sistema WebTutor. Es el encargado de adaptar el sistema al alumno. Esta adaptación posee dos vertientes: la *adaptación de la actuación* del sistema, que decide:

- la accesibilidad del alumno al hiperespacio y el nivel de guía que dará el Tutor; y,
- la *adaptación de la presentación* que selecciona el material didáctico que más se adapta a las características concretas del alumno.

Este módulo se divide a su vez en otros tres:

a) *Decisión de curriculum.* encargado de la adaptación del nivel de actuación del sistema

b) *Selección de Material Didáctico.* Encargado de la adaptación de la presentación de ejercicios y ejemplos.

c) *Evaluación de la actividad del alumno.* Realiza inferencias sobre los conocimientos que posee el alumno a partir de los resultados que éste ha obtenido en los ejercicios y de los recorridos que ha realizado dentro del hiperespacio.

A continuación se describe cada uno de ellos con más detenimiento.

3.2.1.- Decisión de curriculum

Tal y como hemos señalado este submódulo se encarga de decidir qué partes de hiperespacio son accesibles en cada momento adecuándolas a las capacidades del alumno, evitando que se pierda en él. Su forma de actuar está basada en una idea muy simple: si un alumno se puede perder en un hiperdocumento, lo que se ha de hacer es dividirlo y presentar ese hiperdocumento por partes (sub-

hiperdocumentos) y de la forma que pedagógicamente resulte más oportuna dadas las características concretas del hiperdocumento y del alumno.

Así, el hiperdocumento más grande que podría encontrarse en nuestro sistema sería el correspondiente a todo el hiperespacio diseñado, que a su vez posee su organización pedagógica paralela en el Dominio Pedagógico. Si sobre este hiperdocumento el sistema decide hacer una primera división, pues así lo considera oportuno para presentárselo al alumno, lo hará atendiendo a las divisiones que en su momento ha establecido el experto del dominio cuando se ha diseñado y construido el Dominio Pedagógico.

Luego, en este momento, tenemos nueva información sobre el Dominio Pedagógico que no contamos en el apartado 3.1, pero que resulta necesario conocer. Hemos de saber que el Dominio Pedagógico está compuesto no sólo por los conceptos de orden inferior que podríamos llamar básicos y que hemos presentado en la figura 2, sino que el experto ha establecido los agrupamientos de conceptos básicos que ha considerado oportunos formando nuevos conceptos (conceptos *aglutinadores*) y ha estableciendo relaciones pedagógicas entre ellos y sus niveles de dificultad. Se han establecido distintos niveles de agrupamiento de los conceptos básicos: aquellos que ha considerado oportunos, dada su experiencia en la docencia de la materia, el experto del dominio. Así en los niveles de agrupamiento más bajo (niveles 1, 2, 3, etc. de la figura 5) es menor el número de conceptos básicos agrupados por cada concepto aglutinador. El último nivel de agrupamiento (nivel **n**) posee un único concepto aglutinador y que agrupa todos los conceptos básicos del Dominio Pedagógico. Es importante darse cuenta también que en la construcción de conceptos aglutinadores de un nivel (por ejemplo el nivel 3, ver figura 5) no tienen por qué verse agrupados conceptos aglutinadores de su nivel inferior (en nuestro ejemplo ilustrativo el nivel 2 contiene conceptos organizados de manera totalmente diferente a los del nivel 3, y sin considerar inclusiónes de conceptos). Tal y como hemos dicho antes, los niveles de agrupamiento se han hecho de acuerdo a las experiencias del experto humano y en ellos se encontrarán definidas implícitamente sus estrategias de guía y adaptación al alumno como veremos más tarde.

Siguiendo con la idea de paralelismo entre el Dominio Pedagógico y el Hiperespacio debemos señalar que cada concepto aglutinador del primero posee una correspondencia con un hiperdocumento (compuesto por nodos y enlaces) en el segundo (fig. 4).

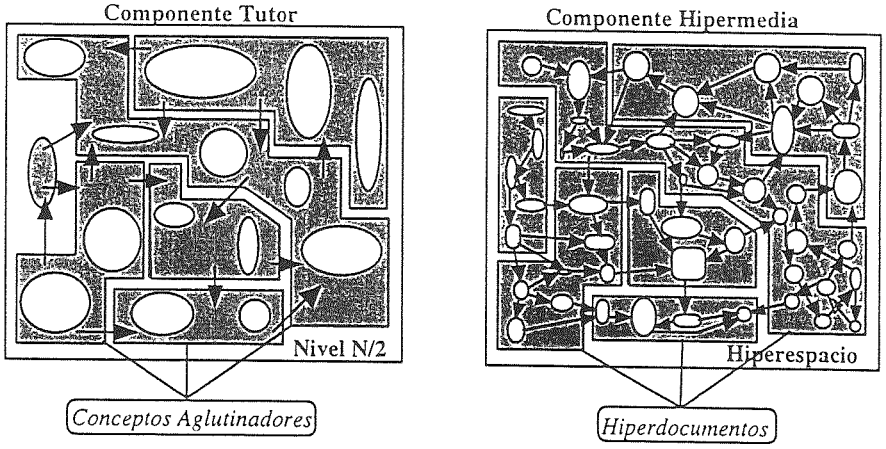


Fig. 4 Los conceptos aglutinadores y su correspondencia con los hiperdocumentos

De esta forma cada nivel de agrupamiento de conceptos establece un nivel de actuación del sistema que corresponde con un nivel de accesibilidad en el Hipermedia y con un nivel de guía por parte del Tutor (fig. 5). Así, en el concepto que corresponde a todo el dominio establece el nivel de actuación **n**, que organiza el nivel de accesibilidad más alto, el alumno puede llegar a cualquier nodo del hiperespacio (el hiperdocumento más grande posible), y el nivel de guía más bajo dado que puede ver la información del dominio (los conceptos) como el quiera sin ninguna dirección por parte del Tutor. El nivel de actuación **1** corresponde con el nivel de accesibilidad más bajo, es decir, sólo puede acceder al grupo de nodos (al hiperdocumento) que corresponden a un concepto básico del dominio, y al nivel de guía más alto, dado que el sistema irá presentando al alumno el contenido del dominio concepto a concepto y en el orden que considere más oportuno de acuerdo con la organización pedagógica del mismo.

En un nivel de actuación intermedio **n/2** tendrá un nivel de accesibilidad medio, donde puede acceder sólo al grupo de nodos correspondientes a las conceptos agrupados bajo un concepto

aglutinador medio (accesibilidad sólo a ese hiperdocumento medio) y un nivel de dirección medio pues el sistema solo dirigirá la forma de presentar los conceptos aglutinadores medios en este nivel pero sin poder dirigir la forma en la que el alumno va a ver los conceptos básicos agrupados por cada uno de estos conceptos aglutinadores medios. Así cuanto mayor sea el nivel de actuación mayor será el nivel de accesibilidad y menor el nivel de guía y al revés cuanto menor sea el nivel de actuación menor será el nivel de accesibilidad y mayor el de guía.

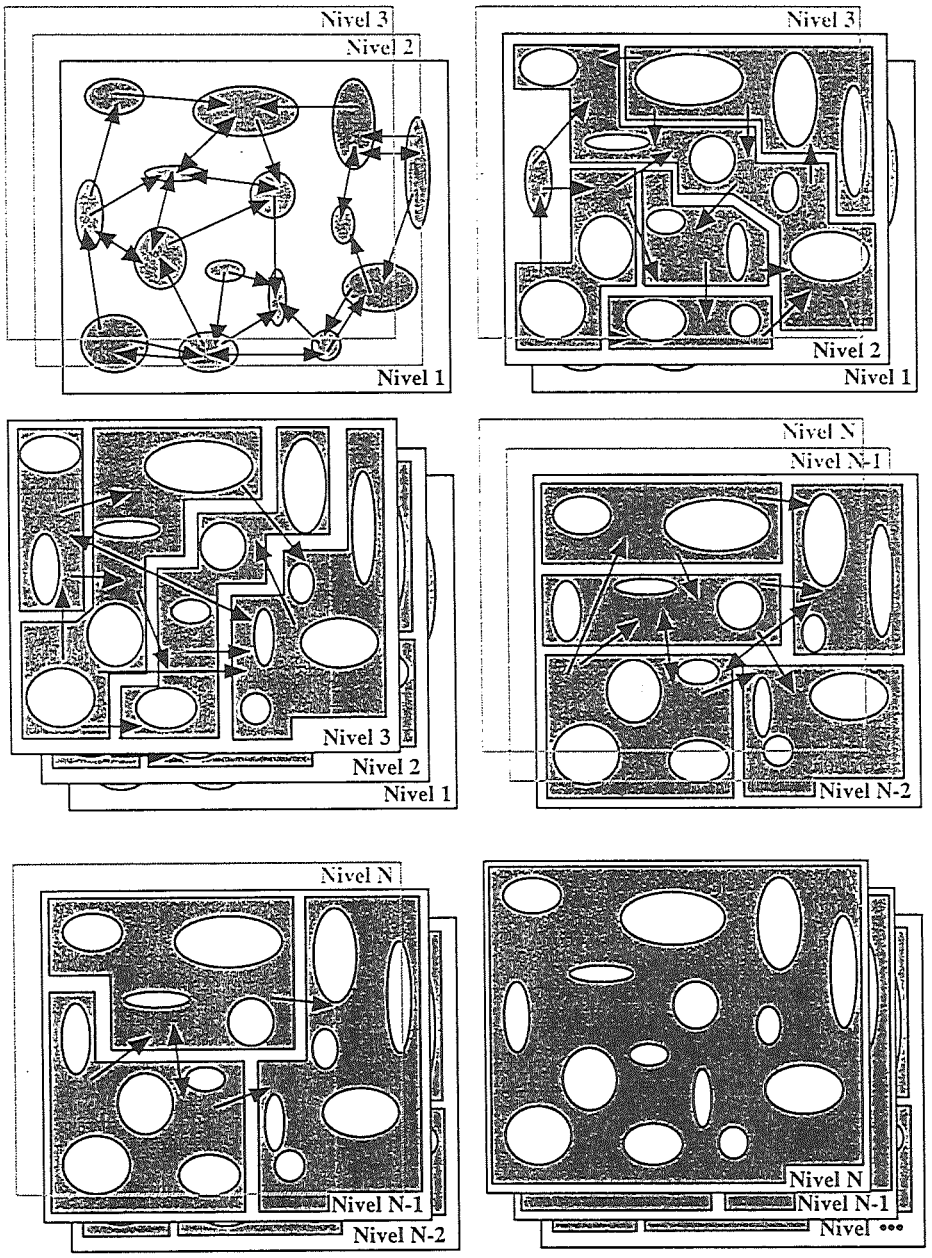


Fig. 5. Un Dominio Pedagógico estructurado en diferentes niveles de actuación.

El nivel de actuación que corresponde con el nivel **n** sería el de un sistema hipermedia clásico, mientras que el nivel **1** constituye una forma de actuar propia de un sistema tutor tradicional. En los niveles intermedios se establecen por tanto actuaciones conjuntas (hipermedia/tutor) cuyo fin es mantener las características de ambos sistemas, la flexibilidad de aprendizaje del hipermedia y la guía en el aprendizaje del tutor, de forma que se adapte lo más posible a las características concretas de cada alumno. Así, el módulo de Selección de Curriculum decide qué nivel le resulta más adecuado de manera que no se pierda navegando dentro de los hiperdocumentos correspondientes y que posea un nivel de guía adecuado que le presente los conceptos aglutinadores de dicho nivel de forma que pueda construir sus mapas conceptuales del dominio de la manera más sencilla y eficaz.

3.2.2.- Selección de material didáctico

Este submódulo se encarga de seleccionar el material didáctico, ejercicios y ejemplos, que más se adapta a las características concretas del alumno. Los ejercicios le servirán para evaluar los conocimientos que ha adquirido el alumno y los ejemplos para explicarle contenidos del dominio de la forma que mejor pueda comprenderlos.

En realidad la selección de material didáctico debería incluirse en lo que denominamos faceta de adaptación en la presentación del sistema y se puede dividir en dos partes: la correspondiente a la adaptación del material didáctico que sería tarea exclusiva del componente Tutor, y la de adaptación de los medios audiovisuales que implicaría a los dos componentes sistema y que posee aspectos más complejos y que aún son materia de investigación.

Como hemos señalado, por ahora se han considerado dos tipos de materiales a presentar y adaptar al alumno: los ejercicios y los ejemplos. Ambos se encuentran asociados a los conceptos recogidos en el Dominio Pedagógico. Su selección se hará dependiendo de las características de aprendizaje del alumno, de los ejercicios y ejemplos que se han presentado anteriormente y de los resultados obtenidos en la utilización de este material concreto. Para más información sobre la forma en que se seleccionan y organizan consulte [18].

3.2.3. Evaluación de la actividad del alumno

Este submódulo analiza los datos del alumno que llegan desde el componente Hipermedia gracias a la comunicación que existe a través del Control de Navegación. Como consecuencia de este análisis se actualizará el Modelo del Alumno guardando aquella información que más tarde será utilizada por el Tutor para tomar decisiones de adaptación al alumno. La actividad de este módulo se divide en tres tareas:

Evaluación de ejercicios

Se comparan las respuestas del alumno con las respuestas correctas almacenadas en el Dominio Pedagógico. Esta tarea resulta sencilla dado que los ejercicios se plantean son combinaciones de preguntas de tipo test de selección múltiple [17]. Cada ejercicio se puntúa entre 0 y 10 de acuerdo al número de respuestas acertadas. Las respuestas erróneas no se utilizan para reducir la puntuación de cara a aumentar la motivación del alumno evitando actitudes de castigo.

Evaluación del conocimiento del alumno sobre un concepto

Se obtiene por medio de una fórmula diseñada heurísticamente. Esta fórmula utiliza dos parámetros principalmente: el porcentaje de conocimiento visto de ese concepto (basado en la información que recibimos desde el componente hipermedia sobre los nodos asociados a ese concepto que se han recorrido y en la representatividad de dichos nodos) y la media de las puntuaciones que se han obtenido en los ejercicios presentados para evaluar el conocimiento de ese concepto. Así, el valor de la fórmula se calcula multiplicando el valor de ambos parámetros. Mientras que el segundo parámetro aparece como la medida de más confianza a cerca del conocimiento del alumno, el primero se utiliza para saber cual ha sido el porcentaje de conocimiento que se supone tiene el alumno en base a su navegación sobre los nodos del hiperespacio.

Evaluación de la actividad global del alumno

Esta información es utilizada para decidir si se debe cambiar el nivel de actuación del sistema (nivel de navegación y de dirección) o para saber si cambiar de categoría cognitiva al alumno. Consideramos tres categorías cognitivas o tipos de alumnos en base a su experiencia previa sobre el dominio: novatos, medios y expertos. Todos estos cambios, tanto de categoría cognitiva como de nivel de actuación siempre se harán de forma que al alumno no le desaparezcan nodos de información del hiperespacio que anteriormente haya visitado o visto. Esto podría desorientar completamente al alumno [2].

3.3. Modelo del alumno

Uno de los aspectos fundamentales a considerar dentro del diseño de un Sistema Tutor Inteligente es la construcción del Modelo del Alumno. Existen trabajos ([19] [23]) que remarcan la importancia de utilizar modelos fáciles de construir y modificar, y que reflejen fielmente las características de los diferentes alumnos. En nuestro caso particular, se ha diseñado un modelo descriptivo [23], individual, que se va modificando a medida que transcurre el proceso de aprendizaje del alumno. Como podremos apreciar a lo largo de este apartado la integración de un Sistema Hipermedia con un sistema Tutor ha llevado a introducir información que hasta ahora no se había necesitado en ningún otro sistema Tutor y que recoge aspectos sobre el alumno que consideran los sistemas Hipermedia y que resultan fundamentales de cara a tener un sistema adaptativo. Clasificamos la información en el Modelo del Alumno bajo cuatro apartados.

Características del alumno.

Se recoge información sobre sus características de aprendizaje, se utiliza por el Tutor para seleccionar los niveles de actuación (niveles de guía y de accesibilidad) y los materiales didácticos que se van a presentar: ejercicios y ejemplos. Este son por tanto los principales parámetros que utiliza el Tutor para adaptar el sistema al alumno. El prototipo construido ha utilizado un atributo para representar la información relativa a la categoría cognitiva, *tipo alumno*, el cual establece tres niveles de experiencia en el dominio: novel, medio y experto. También se utiliza otro atributo para el nivel de actuación que recoge información sobre los niveles en los que ha aprendido el alumno remarcando especialmente el nivel en el que se encuentra. De cara a lo que podríamos denominar habilidades cognitivas del alumno, podrían haberse considerado otro tipo de características como: la velocidad de aprendizaje del alumno, que tipo de aprendizaje utiliza más (memorístico, inductivo, deductivo), etc.; estas cuestiones necesitan de un estudio más profundo y en el prototipo actual no se han tenido en cuenta. Por otro lado y dado que se ha conjuntado un sistema Tutor con un sistema Hipermedia será necesaria la introducción de atributos que guarden información acerca de las preferencias del alumno sobre los distintos medios de presentación de información (vídeo, sonido, animación, etc.) y sobre los parámetros de los mismo (colores, contrastes, etc.). Esta información será utilizada para adaptar al alumno no sólo la información, tal y como hace el Tutor, sino la forma de presentarla adecuándola a sus habilidades audiovisuales tan importantes en el aprendizaje dentro del mundo moderno.

Conocimiento del dominio.

Refleja los conceptos que el alumno ha adquirido a lo largo de su interacción con el sistema y la forma en que los ha adquirido. Esta información constituye un modelo *Overlay* [11] modificado del Dominio Pedagógico. Sin embargo, la representación de los conceptos adquiridos por el alumno aunque refleja la estructura particular del dominio, incorpora nuevos atributos para controlar las características de adquisición. Estos atributos son: nivel de adquisición para reflejar el nivel de conocimiento que posee el alumno sobre el concepto, fecha de adquisición, que servirá para conocer la secuencia en la que han sido adquiridos los conceptos y lo "olvidado" que pueda estar un determinado concepto dependiendo de lo lejana que se encuentre la fecha de su aprendizaje.

Material Didáctico utilizado.

Se guarda información sobre los ejercicios y ejemplos que han sido utilizados para enseñar al alumno. Con ella el Tutor seleccionará el próximo material a presentar al alumno, así podrá: plantear un ejercicio parecido al último visto, presentar un ejemplo más amplio, repetir de nuevo un ejercicio, no repetir un ejemplo hasta que haya transcurrido un determinado tiempo desde su última presentación, etc. Además, por cada ejercicio se guarda información de la puntuación obtenida.

Historia.

Mantiene información sobre el desarrollo del proceso de interacción del alumno con el sistema (datos tanto de la parte Hipermedia como de la parte Tutor). El desarrollo histórico de todas las sesiones se elabora como una colección de notas que relatan lo que ha sucedido en cada sesión. Estas notas serán de gran utilidad para los diseñadores y constructores del sistema, puesto que de ellas podrán obtener informaciones interesantes que servirán para evaluar el sistema. Estos datos también podrían resultar interesantes para el Tutor humano como herramientas para conocer más acerca de sus alumnos, en el supuesto de que el sistema se utilizase de forma complementaria al profesor.

4.- Conclusiones

Los Sistemas Hipermedia permiten un acceso flexible a la información pero presentan alguna desventaja que es especialmente grave si se les considera como entornos educativos: el usuario puede perderse navegando a través de esa información (la mayoría de las veces muy abundante). Para facilitar la navegación se han presentado distintas alternativas tan simples como la provisión de Ayudas a la Navegación representando gráficamente el hiperespacio, como los *ojos de pez* [10], dando recorridos prefijados para recorrer los nodos, como las *guías* [20], la recomendación de los nodos que deben visitarse a continuación según los objetivos que se tengan [13], la adaptación de los contenidos de los nodos dependiendo del tipo de usuario de unos prefijados [2], etc.

WebTutor, además de incluir algunas de las facilidades antes mencionadas, ha considerado una aproximación novedosa que hace que el sistema sea interesante para la educación integrando dentro de un Sistema Hipermedia características importadas del mundo de los Tutores Inteligentes, formando un Sistema Hipermedia Adaptativo. El sistema consta de dos partes bien diferenciadas: el Componente Hipermedia que gestiona el comportamiento hipermedia del sistema y el Componente Tutor que adapta el hiperespacio al usuario (el alumno). Es por esto que WebTutor conjuga las ventajas de los Sistemas Hipermedia (libre navegación, contenidos expresados por diferentes medios) con las de los sistemas tutores

(adaptación al alumno y elaboración dinámica de contenidos a presentar) minimizando las desventajas de los mismos.

La adaptación que proporciona el Tutor se basa en la selección del nivel de actuación (nivel de accesibilidad y nivel de guía) que más se adecúa al alumno. De esta forma y tal y como hemos explicado en el apartado 3.2 se resuelven problemas como la perdida del alumno en el hiperespacio o la excesiva dirección por parte del Tutor para aquellos alumnos que prefieren una aprendizaje autoguiado. Así, se enriquece la capacidad de enseñanza del Sistema dado que es mayor el número de alumnos a los que puede enseñar, desde los que necesitan de un proceso de instrucción completamente guiado a aquellos que aprenden con completa libertad explorando de manera autónoma las posibilidades que presenta el hiperespacio. Por otro lado hemos de decir que la adaptación del Tutor al alumno no termina con la selección del nivel de actuación adecuado, sino que también se selecciona el material didáctico, ejercicios y ejemplos, que más coinciden con las características de aprendizaje de cada alumno y que le ayudarán en el proceso de adquisición de conocimiento.

Agradecimientos

Este trabajo se encuadra dentro del proyecto "Sistemas Hipermedia Adaptativos" financiado por el Gobierno Vasco.

Referencias

[1] O. Astier, J. A. Carro, A. Urkizu & D. Rodríguez. "PACTE: Una nueva generaciónde sistemas en la Web". Novatica, Num. 120, marzo-abril, 1996.

[2] T. Berners-Lee y R. Cailliau. "Proposal for a hypertext project". Documento electrónico. (1989). Disponible en: http://www.w3.org/pub/WWW/Proposal.html

[3] Boyle, C. y A. O. Encarnacion. "Metadoc: An Adaptive Hypertext Reading System". *User Modeling and User-adapted Interaction* 4, 1-19. (1994).

[4] P. Brusilovsky, L. Pesin & M. Zyryanov. (1993). Towards an Adaptive Hypermedia Component for an Intelligent Learning Environment. En L. J. Bass, J. Gornostaev & C. Unger (Eds.): *Lecture Notes in Computer Science #753, Human Computer Interaction.* Springer-Verlag: Berlin (Alemania), pp. 348-358.

[5] P. Brusilovsky. (1994). Adaptive hypermedia: the state of the art. *Proceedings of the East-West International Conference on Multimedia, Hypermedia and Virtual Reality, MHVR'94*, Moscú (Rusia).

[6] R. Chamorro. "La red de redes: Internet". Documento electrónico. (1995). Disponible en: http://www.bne.es/internet.html

[7] D. W. Connolly. "HyperText Markup Language (HTML)". Documento electrónico (1995). Disponible en: http://www.w3.org/pub/WWW/MarkUp/

[8] "Convenio CICYT-Telefónica". Documento electrónico. (1995). Disponible en: http://www.rediris.es/rediris/red/convenio.html

[9] O. El Hani. "Un modèle d'architecture pour les hypermédia éducatifs". Tesis doctoral, Université Paul-Sabatier,Toulouse, France (1993).

[10] G. W. Furnas. "Generalized Fisheye views". *CHI'86 Conf. Proceedings.* ACM Press. (1986).

[11] I. P. Goldstein. "The Genetic Graph: A Representation for the Evolution of Procedural Knowledge". D. Sleeman and J. S. Brown Eds. Intelligent Tutoring Systems. London Acad. Press (1984).

[12] J. Gutiérrez. "INTZA: Un Sistema Tutor Inteligente para entrenamiento en entornos industriales". Tesis Doctoral. Universidad del País Vasco UPV-EHU, San Sebastián, España. (1994).

[13] C. Kaplan, J. Fenwick y J. Chen. "Adaptive Hypertext Navigation on User Goals and Context". User Modeling and User-adpted Interaction 3, 193-220 (1993).

[14] J. Nielsen. (1993). *Hypertext & Hypermedia.* Academic Press, Cambridge, MA, USA.

[15] T. A. Pérez y J. Gutiérrez. "¿Pueden ser los Hipermedia más Educativos?" M. Ortega, J. Bravo, F. Ruiz y J. Ruiz (eds.). *Informática Educativa.* Colección Ciencia y Técnica. Ediciones de la Universidad de Castilla-La Mancha. (1995).

[16] T. A. Pérez, J. Gutiérrez, P. Lopistéguy & I. Usandizaga. (1995). HyperTutor: From Hypermedia to Intelligent Adaptive Hypermedia. En H. Maurer (Ed.) *Educational Multimedia and Hypermedia, ED-MEDIA'95.* AACE: Charlottesville, USA.

[17] T. A. Pérez, J. Gutiérrez. & P. Lopistéguy. (1995). An Adaptive Hypermedia System. En J. Greer (Ed.) *Artificial Intelligence in Education, AI-ED'95.* AACE: Charlottesville, USA.

[18] T. A. Pérez, J. Gutiérrez, & P. Lopistéguy. (1995). The Role of Exercises in a User-Adapted Hypermedia. Proceedings of the 3rd Computer Aided Engineering Education, CAEE'95, Bratislava (Slovakia), pp. 57-62.

[19] A. M. Rutkowsky. (1994). *The Present and the Future of the Internet: Five faces.* Keynote Address in Networld + Interop'94. Disponible en: http://www.isoc.org/speeches/interop-tokyo.html

[20] G. Salomon, T. Oren y K. Kreitman. "Using Guides to explore Multimedia Databases". Actas del 22nd Int. Conf. on System Science. IEEE Computer Society Press (1989).

[21] Sleeman, D. "UMFE: A User Modelling Front-End Subsystem". International Journal of Man-Machine Studies. Vol. 23, pp. 71-88 (1985).

[22] Usandizaga, I. & P. Lopisteguy. (1995). Una nueva tendencia en Sistemas Hipemedia Educativos. M. Ortega, J. Bravo, F. Ruiz & J. Ruiz (eds.). En *Informática Educativa.* Colección Ciencia y Técnica. Ediciones de la Universidad de Castilla-La Mancha.

[23] Usandizaga, I.; P. Lopisteguy & T. A. Pérez. (1995). MADAME: An Object Oriented Hypermedia Prototype. *Basque International Workshop on Information Technology, BIWIT'95,* IEEE Press, 1995.

[24] M. F. Verdejo. "User Modeling in Knowledge-Based Systems". Actas del 1st International Colloquium on Cognitive Science. San Sebastián. (1992).

Sistema de enseñanza de Televisión

Carlos Fernández Baladrón, Xulio Fernández Hermida
DEPARTAMENTO DE TECNOLOGÍAS DE LAS COMUNICACIONES [††]
ETSIT. Universidad de Vigo
Ciudad Universitaria s/n 36200 Vigo
Correo electrónico: carlos@gpi.tsc.uvigo.es xulio@gpi.tsc.uvigo.es

Abstract:

In this paper, we show a remote educational system, oriented to Television teaching, available trough the Internet, using the WWW interface. The problem in developing educational systems in this enviroment lies in the branched structure of the hipertext , i. e. the taking of decisions. A guide system is, therefore, necessary.

1. Introducción

Afirmar que el WWW constituye una de las aportaciones tecnológicas de mayor repercusión en los últimos años no creará excesiva polémica. Desde su aparición quedó claro que tenía todas las cartas ganadoras: capacidad de gestión en red, facilidad de edición de documentos y, sobre todo, una potente interfaz gráfica.

Como consecuencia natural de las posibilidades del entorno se han realizado diversas experiencias educativas.

El problema fundamental radica en que, debido a la estructura propia del Web, la información pierde la linealidad propia de los sistemas de enseñanza tradicionales para adoptar una disposición ramificada. En los elementos "centrales" se producen las tomas de decisión.

Si no se asumen estos principios durante la creación de sistemas de enseñanza, difícilmente se podrá rentabilizar el uso del entorno Web.

2. Soporte para el curso de Televisión

Como consecuencia de la naturaleza del modelo Web se han desarrollado diferentes elementos para la consecución del sistema de enseñanza de Televisión.

La posibilidad de acceso mediante Internet permite que usuarios puedan seguir el curso desde literalmente cualquier parte del mundo, siempre que se disponga de una conexión. Como soporte se ha elegido un servidor de HTTP sobre Linux.

3. Webscopio

Como se expuso en la introducción, la estructura ramificada del sistema Web hace difícil su uso para la enseñanza. Como solución a este problema han surgido sistemas que basan su funcionamiento en la gestión de los enlaces y que, en mayor o menor medida, son dirigistas. Es la aproximación del WWW como el acceso a un base de datos.

Como alternativa se propone una solución basada en varias observaciones:

1. En la estructura del Web hay páginas centrales a partir de las cuales se ramifica la estructura. Es usual en los servidores web elaborar más las páginas centrales.
2. La tendencia en informática, especialmente la dirigida al gran público, es evolucionar hacia sistemas con interfaces gráficas elaboradas. Se deberá prestar especial atención, en consecuencia, a la creación de un entorno atractivo.
3. Es habitual en el desarrollo de una clase tradicional, partir de un dibujo o ayudarse con instrumental de medida, para explicar los conceptos de un tema, en lugar de seguir un esquema lineal.

Todo lo expuesto sugiere la utilización de ciertos elementos que ayuden a la navegación; estos son applets en Java a partir de las cuales se lleva al alumno a diferentes páginas. En el sistema de enseñanza realizado es posible elegir uno de los varios temas relacionados con la Televisión, lo que sería una visión próxima a los libros tradicionales; buscar en el índice general las relaciones que existen con un tema, que previamente se han indexado, o utilizar el Webscopio como punto de partida para consultar los diferentes temas.

El Webscopio (figura 1) permite representar la señal de televisión que correspondería a una imagen que se le introduce en formato JPEG o GIF.

3.1. Objetivo del Webscopio

Una de las cosas que más atrae a los estudiantes es poder representar la señal de televisión en un osciloscopio.

El inconveniente es la necesidad de disponer de un equipo caro y con una cierta dificultad de manejo, como es un monitor de forma de onda u osciloscopio y un generador de patrones o cámara de vídeo.

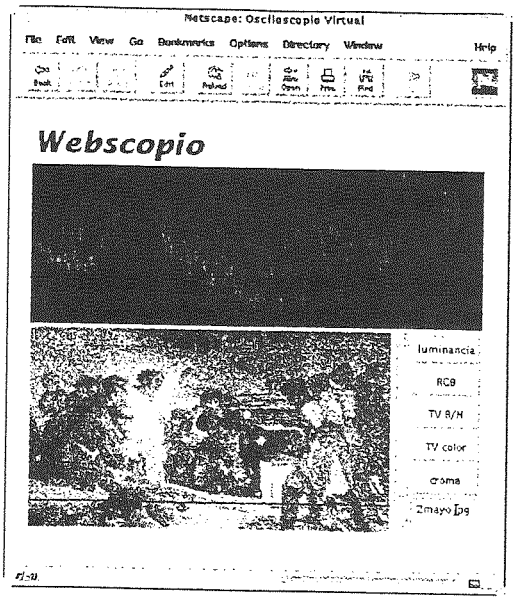


Figura 1

La solución idónea para fines educativos es, en consecuencia, una simulación de ordenador. El problema es la elección de plataforma y el acceso en red. El lenguaje Java de Sun Microsystems soluciona estos problemas.

3.2. Opciones de representación

A continuación se describen las opciones de representación que permite el programa.

- Luminancia.
- Componentes RGB.
- Señal de televisión en blanco y negro.
- Señal de televisión en color.
- Señal de croma.

La luminancia y componentes RGB son características de la imagen y se representan sobre un fondo blanco. Para las señales eléctricas generadas (televisión en B/N, televisión en color y croma) se representan en verde sobre fondo negro, para reforzar la idea de un monitor de forma de onda.

3.3. Generación de saltos

Uniendo la capacidad de representación del Webscopio a la idea de unas páginas centrales muy elaboradas que ayuden a la navegación, se dio un paso más, haciendo que la imagen de la señal representada (que se corresponde con la señales de televisión, luminancia y componentes) fuese sensible a los clics de ratón. de esta forma se provoca la consulta de las diferentes páginas en función de la zona de interés de la señal representada.

En la figura 2 se muestran las diferentes zonas que corresponden a la señal de televisión en color.

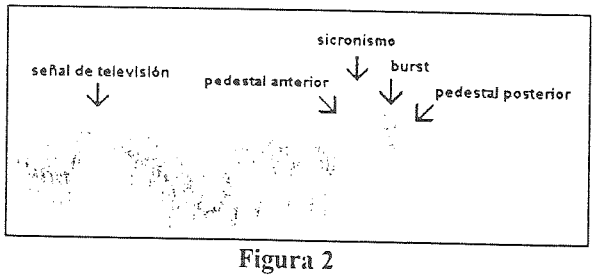


Figura 2

4. Consultas

El objetivo ha sido en todo momento la generación de un sistema de enseñanza de Televisión, no simplemente un libro en soporte electrónico. Como tal debe contemplar la bidireccionalidad que es usual en un sistema presencial: por un lado la información fluye del profesor al alumno, pero posteriormente el alumno genera información hacia el profesor en forma de cuestiones, que a su vez son contestadas

Como consecuencia, se tomó la decisión de diseñar un soporte para el curso, que consistía en un servidor web con ciertos mecanismos adicionales para el control de usuarios, autenticación, y seguimiento de la evolución.

Para añadir interactividad se han seguido dos métodos: incluir programas de simulación, "applets" y por otro lado permitir realizar las consultas y sugerencias habituales en una clase tradicional mediante "*forms*" o formularios. Como estas consultas pueden ser de interés para futuras actualizaciones, o incluso para elaborar una lista de *FAQ's* (preguntas más frecuentes), resulta interesante que se guarden, de forma que puedan ser consultadas en el futuro.

Para esto se generan páginas web automáticamente mediante scripts que se ejecutan en el servidor que se adaptan al estándar CGI (figura 3 y figura 4).

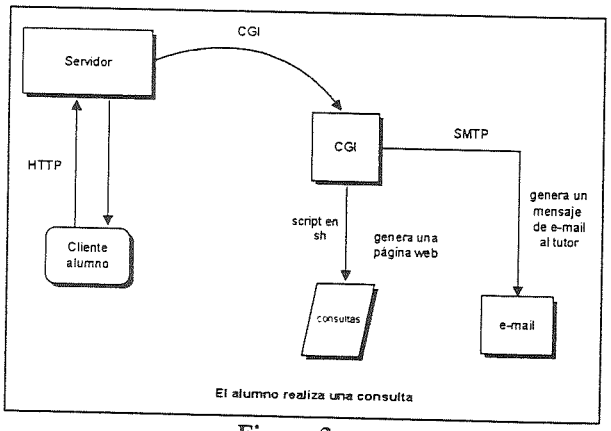


Figura 3

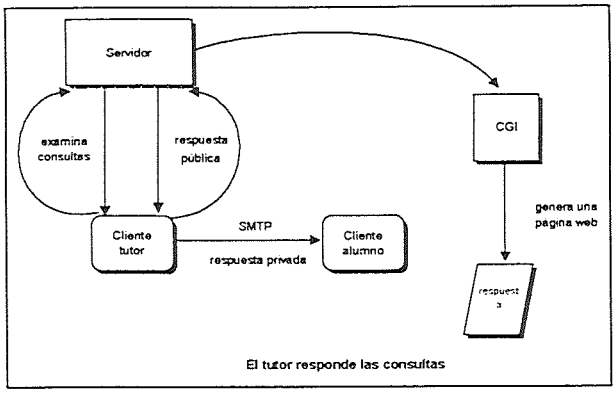


Figura 4

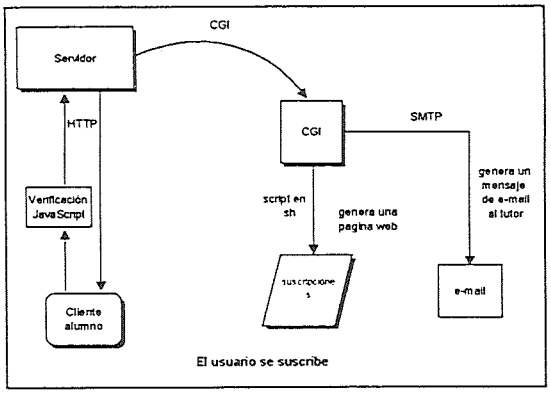


Figura 5

No obstante las sugerencias recibidas no aportaban suficiente información acerca de los usuarios. Se ideó entonces una forma de "forzar" las sugerencias, utilizando un sistema de autenticación de usuarios.

5. Autenticación y creación de perfiles de usuario

El propósito de incluir un sistema de autenticación de usuarios no pretende una restricción del acceso si no obtener información acerca de los usuarios que ayudase al desarrollo del curso de Televisión.

Cuando se accede por primera vez al curso se presenta la misma página principal, pero cualquier enlace que se elija presentará un mensaje de petición de *login* (identificador de usuario) y *password* (palabra clave). Los únicos accesos libres serán los índices de cada tema, el índice temático y la ayuda donde se explican todos los pormenores de la suscripción.

Si el usuario suministra los datos adecuados el servidor de HTTP responderá con el fichero pedido y en la mayoría de los browsers no volverá a realizar la petición dentro de la misma sesión. Si los datos suministrados no son correctos o bien se elige la opción CANCEL, el servidor responderá mostrando una página donde se deniega el acceso y que contiene un form para suscribirse al curso.

La suscripción genera un fichero HTML mediante un programa que se ejecuta en el servidor (figura 5) y se utiliza únicamente para realizar un seguimiento de las consultas que realizan los alumnos, mediante los ficheros de *logs* (entradas al sistema realizadas por los usuarios).

La observación de la información almacenada en el fichero permite conocer:

- interés que suscitan los temas.
- orden de consulta de las páginas.
- relación del perfil del alumno con todo esto.
- deficiencias en la secuenciación del curso
- páginas con una secuenciación defectuosa.

Antes del envío al servidor se realiza una validación del formulario en JavaScript, para evitar retardos en la respuesta, en caso de que faltase un dato y se tuviese que volver a cubrir el formulario.

De la misma forma que se hace con las sugerencias, se genera un mail automáticamente al administrador del curso para que sepa que se ha suscrito un nuevo alumno.

Las páginas generadas con los datos de los alumnos son accesibles a través de Internet, pero solo por el tutor del curso. Aunque se tenga acceso a los contenidos del curso no se tiene acceso a los datos de los alumnos.

6. Contenidos

La generación de contenidos para el curso quizá sea la parte más importante, ya que un sistema de enseñanza sin contenidos carece de sentido.

Un sistema por bueno que sea solo será útil como soporte de unos contenidos. Estos habrán de ser de interés y actualizados periódicamente.

Es importante resaltar que la generación de contenidos es la parte más difícil de la elaboración del curso, a pesar de ser la menos espectacular. Por un lado está el problema de la selección y redacción de los contenidos, común a la edición en otros soportes tradicionales.

Pero existe un problema añadido. a estructura propia del Web es contraria a la organización lineal que se usa en los textos tradicionales.

El curso se organiza, como es habitual, en temas que agrupan cierto número de páginas web. Debido a los saltos que se introducen entre los diferentes temas, esta estructura no es tan rígida como en un texto tradicional.

Los temas que no están disponibles pero van a ser objeto de redacción inmediata aparecen en el cuadro de temas con letras sobreimpresas que los identifican. Esto evita consultas inútiles al tiempo que anuncia temas que pueden ser de interés

Los temas desarrollados hasta el momento son:

Historia. Este tema es de carácter introductorio. En él se describe la evolución de la televisión desde los primeros sistemas experimentales hasta la televisión en color y los sistemas en componentes. Dado que abarca la práctica totalidad del proceso de desarrollo de la televisión, es un buen punto de partida para consultar los diferentes temas que se encuentran como enlaces a lo largo de este.

Actual. Desde el planteamiento del proyecto, siempre se consideró de gran interés tratar temas de actualidad, ya que estos son menos frecuentes en la bibliografía existente. La razón principal de la escasez de estos temas es que en los sistemas de edición tradicional el periodo que transcurre desde la elaboración hasta que el libro está finalmente disponible es tan largo que con frecuencia supera el tiempo que permanecen estables los estándares. Además, la edición comercial tiene unos costes de lanzamiento y producción inicial que restringen la posibilidad de editar cierto tipo de información de interés minoritario, como es la información técnica. Estos costes se reducen de forma espectacular cuando se realiza la edición en entorno Web.

Los temas de actualidad tratados son:

- PALplus
- MMDS

Vídeo. Desde los inicios de la televisión un objetivo constante fue el desarrollo de medios para conservar las imágenes captadas por las cámaras electrónicas.

En este tema se tratan los diversos sistemas de almacenamiento de la señal de televisión en soporte magnético, desde los domésticos hasta los profesionales.

Tecnología de la Televisión. La explicación de los fundamentos técnicos de los diferentes sistemas es imprescindible para la comprensión de la televisión y su evolución.

Procesado de Imagen. La televisión digital involucra principios de procesado de imagen. Dado que nuestro grupo tiene una amplia experiencia en este campo, nos ha parecido una ocasión ideal para introducir estos temas en un curso de televisión.

Por los datos obtenidos de los alumnos suscritos al curso hemos podido constatar que este es uno de los temas que suscita mayor interés.

Índice. Ideado como una búsqueda al estilo de los índices de los textos habituales. A pesar de que el uso del sistema se basa en seguir los enlaces a partir de puntos centrales de interés, como applets en Java o imágenes sensibles, también se ha previsto que el usuario desee buscar los contenidos que hagan referencia a un término en concreto.

Ayuda. Una de las principales razones del éxito del WWW es la utilización de una interfaz de usuario atractiva y de uso muy intuitivo. No obstante, por la experiencia adquirida durante el desarrollo del proyecto, se ha comprobado que es deseable introducir ciertas recomendaciones.

7. Normas de edición

A medida que los sistemas de gestión documental crecen es necesario el trabajo en equipo. De esta forma, los métodos y sistemas que son válidos para pequeños sistemas se vuelven demasiado lentos en la elaboración de grandes sistemas.

Es necesario elaborar normas y protocolos para la edición de contenidos para conseguir un resultado homogéneo en un tiempo mínimo.

Los temas relacionados con la tecnología están en constante revisión. Una condición indispensable era, en consecuencia, concebir el sistema de forma que se pudiese ampliar o actualizar fácilmente. Para facilitar esta labor se han redactado unas normas para la edición de contenidos, que recogen la experiencia adquirida en la elaboración del curso y permiten el trabajo en equipo.

8. Conclusiones

Generar contenidos en entorno web es complejo. La anarquía, el grado de desorden, que introduce el modelo cuando se explota correctamente hace difícil su uso, sobe todo con fines docentes.

La elaboración debe abordarse desde una perspectiva diferente a los libros tradicionales.

Utilizar applets no solo como elemento de demostración sino también como guía para motivar la consulta de las diferentes páginas constituye una solución alternativa al problema.

El uso de los datos obtenidos con la suscripción y de los logs permite determinar el perfil del alumno y evaluar el curso.

Referencias

[1] Anuff, Ed.,. "Java Sourcebook". *John Willey & sons* (1996).

[2] December , John.. "HTML & CGI Unleashed" *Sams.net Publishing.*.(1996).

[3] Stevens , Richard W. "TCP/IP Illustrated Volume 3", Adison Wesley (1996)

Elaboración de recursos para el aprendizaje de las matemáticas en entorno Web

José Luis Hueso Pagoaga, Ana Martínez Vidal, Roberto Romero Llop, Juan Ramón Torregrosa Sánchez
DEPARTAMENTO DE MATEMÁTICA APLICADA
E.T.S.I. Telecomunicación, UNIVERSIDAD POLITÉCNICA DE VALENCIA
Camino de Vera, 14, 46071 VALENCIA
Correo electrónico: jlhueso@mat.upv.es, anmarvi1@teleco.upv.es, rorollo@teleco.upv.es, jrtorre@mat.upv.es

Abstract:

The increasing availability of computers in the classroom has produced an important revision of the objectives and contents of the mathematical subjects taught in the first years of Engineering Schools. Nowadays, the development of new computer communication tools is triggering new initiatives to restructure teaching at university level, with the aim of creating Virtual Universities which provide the cyberspace as 'campus' where learning interaction takes place.

This paper presents two electronically supported teaching applications. The first one uses the university computers network to perform administrative tasks and mainly to distribute learning resources to the students via Internet. The other application consists of developing interactive environments for learning Mathematics. These environments are intended to complement the information received in the classroom or the prerequisites for the subject.

1. Introducción

La Universidad Politécnica de Valencia promueve desde principios de los 90 el llamado Proyecto de Innovación Educativa (PIE) que acoge las propuestas de grupos de profesores y de Centros tendentes a mejorar la calidad de la enseñanza adaptando los contenidos a la actualidad tecnológica y científica y la metodología docente a las últimas tendencias didácticas.

Nosotros hemos participado en le PIE desde sus inicios con diversos Proyectos de Innovación Docente (PID) en los que hemos impulsado la inclusión de las nuevas tecnologías en la enseñanza de las matemáticas.

2. Prácticas de Matemáticas con Ordenador

La Escuela de Ingenieros de Telecomunicación fue la pionera dentro de la UPV de la realización de prácticas con ordenador en matemáticas, lo que fue el objetivo de nuestros primeros PID's. Actualmente, en casi todas las titulaciones de nuestra Universidad, la docencia de las matemáticas del primer curso tiene una componente práctica en la que el alumno aplica los conocimientos adquiridos a la resolución de problemas con ordenador.

Esta experiencia inicial se plasmó en 1994 en la publicación de un libro [1] usado por los alumnos como guía para la realización de las prácticas frente al ordenador. El breve tiempo transcurrido desde el inicio de la experiencia no ha sido óbice para que nos hayamos planteado y realizado una profunda revisión metodológica de las Prácticas de Matemáticas. Los contenido han variado ligeramente, pues se refieren a métodos numéricos elementales de la Ingeniería (resolución de ecuaciones lineales, algebraicas y diferenciales, integración numérica, interpolación polinómica, aproximación mínimo-cuadrática), pero la experiencia nos ha sugerido la unificación de las herramientas informáticas utilizadas en las prácticas. Inicialmente, utilizábamos distintos programas: hojas de cálculo, lenguajes de programación, programas de cálculo numérico y simbólico, según el tema tratado.

La implantación el pasado curso 96-97 del nuevo Plan de Estudios, nos obligó a concentrar las prácticas, antes vinculadas a las asignaturas de Álgebra y Cálculo, en una nueva asignatura formalmente independiente, Laboratorio de Matemáticas, con 4.5 créditos prácticos. Aprovechando el cambio de status, decidimos modificar las prácticas en el sentido de utilizar en ellas un solo programa que nos permitiese cubrir todos los objetivos docentes. El programa elegido fue el Matlab, por sus reconocidas cualidades numéricas y gráficas, su programabilidad y su gran implantación en el ámbito del cálculo numérico en la ingeniería.

Durante este curso hemos realizado la adaptación a Matlab de los algoritmos que antes desarrollábamos en distintos lenguajes y la reelaboración del la documentación docente de las prácticas para uso de los alumnos.

Al tiempo que actualizábamos los contenidos, decidimos ponerlos a disposición de los alumnos por nuevos medios, por lo que aparte del tradicional papel impreso, recurrimos a la difusión de los documentos que íbamos elaborando mediante la red local de la Universidad y la Web.

3. Utilización docente de los recursos telemáticos

Los alumnos de nuestra Escuela tienen acceso a un servidor local, al correo electrónico y a la World Wide Web y disponen de un aula de ordenadores con un amplio horario de acceso libre. Algunos alumnos tienen en casa conexión a Internet mediante proveedores privados. Esto ha hecho que nos planteemos la utilización docente de los recursos telemáticos.

3.1 Gestión docente por correo electrónico

En la asignatura de Introducción a las Aplicaciones Científicas Técnicas e Informáticas, que es optativa del segundo cuatrimestre, los alumnos hacen un trabajo por grupos. Cuando se forma un grupo, el responsable nos manda un correo indicando los componentes del grupo y el tema elegido. Nosotros les contestamos aceptando el tema o indicándoles las modificaciones oportunas. Si a lo largo del curso se producen alteraciones en los componentes del grupo, o quieren cambiar el tema, lo comunican por correo.

También envían como anexos de correo, las presentaciones preliminares de los trabajos, que son ficheros de Power Point de unas pocas diapositivas. Finalmente, se pueden enterar por el mismo medio de las fechas de examen.

3.2 Página Web de Laboratorio de Matemáticas

Con respecto al uso de la Web, hemos creado una página de la Unidad Docente de Telecomunicación del Departamento de Matemática Aplicada (http://www.etsit.upv.es/depart/dmat/homepage.htm) con información sobre sus componentes y las asignaturas impartidas en ella. Esta página en construcción (¿y cuál no?) contiene información general de todas las asignaturas (organización docente, programa desarrollado de la asignatura, profesores que la imparten), mientras que la documentación docente sólo está desarrollada en alguna de ellas, como Laboratorio de Matemáticas, a la que nos referimos en adelante.

Además de los datos antes mencionados, en esta asignatura disponemos de información más completa en los apartados de Apuntes y Transparencias.

El primero de ellos contiene información en formato de texto (documentos de Word) de parte de las 12 prácticas de las que consta la asignatura. En alguna de ellas la documentación no está aún completamente desarrollada y se reduce a un guión de los contenidos de la práctica. En otras incluimos junto al guión una colección de ejercicios a realizar por los alumnos. En las más trabajadas se dispone del desarrollo completo del tema, con la explicación detallada de los conceptos teóricos y de los algoritmos numéricos. Típicamente, estos algoritmos se aplican a ejemplos concretos y se incluyen segmentos de código de Matlab que permiten resolver el problema planteado así como otros análogos.

Para ver las transparencias utilizadas por el profesor en la explicación de las prácticas, basta seguir el correspondiente enlace de la página Web de la asignatura Laboratorio de Matemáticas (http://www.etsit.upv.es/depart/dmat/asig/labormat/) y elegir las de la práctica deseada. En las transparencias, el alumno puede encontrar las ideas fundamentales de la práctica, los algoritmos utilizados e indicaciones de cómo usar Matlab para aplicarlos. Este apartado está totalmente desarrollado, disponiéndose así de transparencias de todas las prácticas.

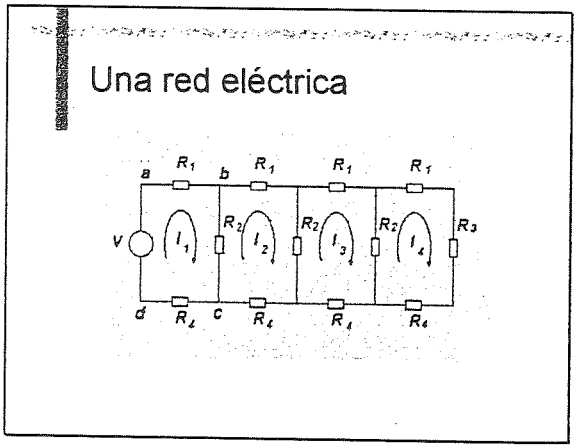
El alumno tiene toda la información al alcance de una tecla, que puede pulsar en la sesión de prácticas mientras está elaborando un programa del algoritmo explicado, en su casa, para completar los ejercicios propuestos o para imprimir las transparencias en color, o el día del examen, para recordar las ideas o algoritmos que debe aplicar para resolverlo.

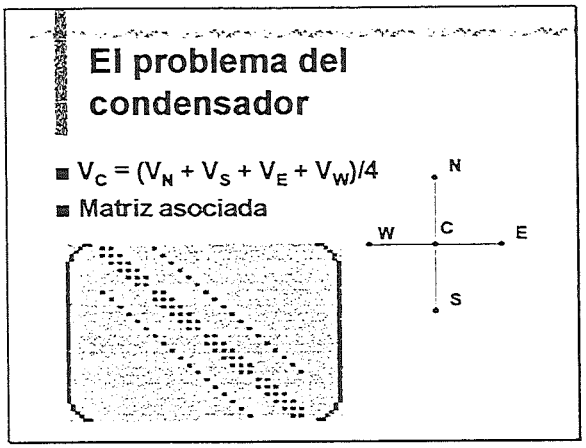
3.3 Contenido del Laboratorio de Matemáticas

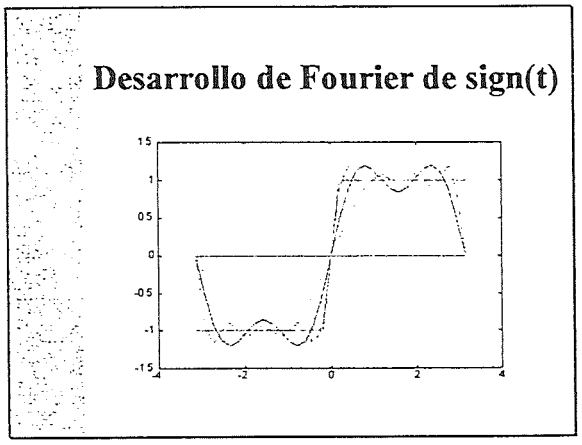
Como ya hemos indicado, las prácticas de Laboratorio de Matemáticas versan sobre métodos numéricos elementales y su programación en Matlab.

En la primera práctica, de carácter introductorio, explicamos los rudimentos del uso de Matlab, haciendo especial hincapié en que las variables y las operaciones son, por defecto, de tipo matricial, con las ventajas e inconvenientes que ello supone. La principal ventaja, desde nuestro punto de vista, es la simplificación en la escritura de muchos algoritmos, al sustituir los bucles necesarios en un lenguaje que usa variables indexadas por operaciones con vectores o matrices. La otra cara de la moneda es que este proceso exige un cierto grado de abstracción al que hay que ir habituando poco a poco al alumno. Otra ventaja, de cara a alumnos de primeros cursos, es la relativa facilidad con que se obtienen los distintos tipos de gráficos en dos y tres dimensiones, que permiten representar curvas, superficies y campos de vectores.

Las prácticas 2 y 3 tratan de métodos iterativos para la resolución de ecuaciones no lineales (para ecuaciones de una variable). Partiendo de la seguridad del método de bisección, exploramos otros algoritmos más arriesgados, con el fin de conseguir mayor velocidad de convergencia. En el marco del método del punto fijo, examinamos el de Newton y sus variantes, como el método de la secante.







Los problemas de circuitos eléctricos lineales o de tráfico en una red de calles ilustran la necesidad de disponer de métodos rápidos y fiables para resolver sistemas lineales. En las prácticas 4 y 5, comparamos las características de los métodos directos basados en la eliminación gaussiana con las de los métodos iterativos elementales (Jacobi, Gauss-Seidel y sobrerrelajación), indicando sus limitaciones y ámbitos de aplicabilidad.

La discretización de ecuaciones diferenciales y en derivadas parciales da lugar a sistemas de gran tamaño y con matriz dispersa que se resuelven generalmente por métodos iterativos. Remitimos a cursos posteriores de análisis numérico en los que se dan algoritmos más sofisticados para resolver las dificultades computacionales que surgen al tratar problemas de gran tamaño o mal condicionados.

En la práctica 6 se generaliza la regla de Simpson de integración numérica, obteniéndose el método de Romberg.

Las prácticas siguientes, de la 7 a la 10, giran en torno a las distintas facetas de un concepto que se nos antoja fundamental: la aproximación funcional.

El caso más sencillo, desde el punto de vista teórico es el de la interpolación polinómica que es resuelta en la práctica 7. Mediante distintos ejemplos mostramos al alumno que esta solución no es razonable en ciertos casos, debido a los efectos indeseables del aumento del grado del polinomio con el número de puntos a interpolar.

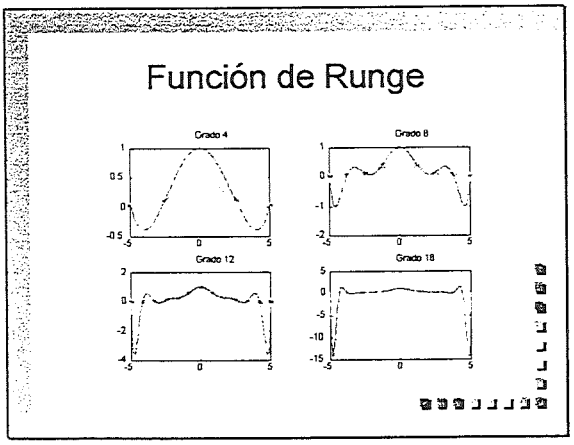
Introducimos así la práctica 8 sobre interpolación polinómica segmentaria (los famosos splines) y la 9 sobre ajuste polinómico mínimo cuadrático. Las ideas de aproximación en el espacio euclídeo utilizadas en esta práctica se generalizan a la práctica 10, en la que se trata de aproximar funciones, en lugar de puntos, mediante polinomios. El problema de aproximación se formula como la minimización de una norma convenientemente elegida. Como caso particular notable mencionamos las aproximaciones de Fourier, que tantas veces encontrará el alumno de Telecomunicación a lo largo de su carrera.

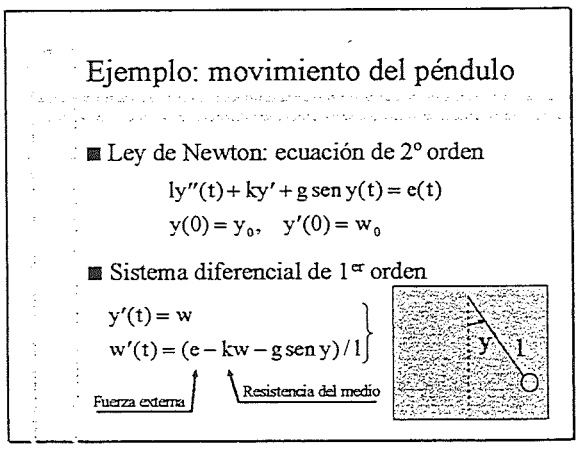
En las dos últimas prácticas se analizan los métodos de resolución de problemas de valor inicial y problemas de contorno para ecuaciones o sistemas diferenciales. Estudiamos los métodos de Euler, Euler modificado, Heun y Runge-Kutta, poniendo especial énfasis en el orden de convergencia de cada uno de ellos en relación con el coste de cada iteración. Estos métodos se aplican a ejemplos como el del péndulo, que permite analizar visualmente la convergencia e interpretar gráficamente el resultado.

La extensión del tema a las ecuaciones en derivadas parciales remite de nuevo a la asignatura de cálculo numérico.

4. Multimedia educativo.

La preocupación por utilizar diferentes





medios a la hora de dar clase, hacer una presentación, etc., no es nueva. Ya en los años 70 se incluía el uso simultáneo de diferentes productos en una clase. Estos productos eran texto, diapositivas, sonido, vídeo y música entre otros. Pero el profesor estaba limitado porque el texto estaba en papel, las diapositivas en película 35mm, y el vídeo la música y el audio en cinta magnética. Las ideas que tenía en un medio eran difíciles de implementar en los demás, con lo cual era tedioso trabajar de esa manera.

Con el desarrollo de las tecnologías digitales este problema está resuelto, al menos en gran parte. Es ahora el momento de diseñar herramientas que aprovechen todas las ventajas de las nuevas tecnologías.

Hace unos años, multimedia entró en escena reivindicando su protagonismo en el proceso enseñanza-aprendizaje. Pero, la transformación del sistema educativo es lenta y no siempre va por el camino adecuado. No es suficiente con convertir las páginas de los libros de texto en pantallas de un ordenador con unas gráficas excelentes. Nuestro conocimiento de distintas técnicas de aprendizaje, modelos de enseñanza y teorías sobre la educación nos pueden servir para desarrollar herramientas multimedia que faciliten el aprendizaje. La tecnología de multimedia interactivo, junto con nuestra experiencia docente, nos proporciona una buena oportunidad para incrementar los niveles de eficacia y eficiencia en el proceso educativo.

La disponibilidad a medio plazo de dispositivos de almacenamiento de alta capacidad a precios razonables favorecerá el crecimiento de servicios multimedia educativos. La oferta europea de estos productos está constituida por algunos grandes grupos industriales y una multitud de pequeñas empresas que tienen problemas de distribución por la fragmentación del mercado. Las productoras americanas ocupan actualmente una posición de privilegio en este mercado incipiente.

Numerosas experiencias han demostrado la utilidad pedagógica del multimedia. Sin embargo, la generalización del uso del multimedia en la educación, en cualquier nivel del proceso educativo se ve dificultada por una serie de obstáculos:

- Escasez de equipos y software multimedia para profesores y alumnos
- Equipos poco potentes para soportar software multimedia.
- Insuficiencia, tanto en calidad como en cantidad, de programas informáticos educativos.
- Dificultad de integración del multimedia educativo en la práctica pedagógica del profesor.
- Falta de información y formación del personal docente.
- Coste de elaboración del material.

Por regla general, las universidades producen ellas mismas materiales educativos multimedia para la formación de sus alumnos. Están empezando a utilizar redes de telecomunicación para la difusión de cursos y la colaboración con otros centros en temas de investigación.

5. Elaboración de recursos multimedia

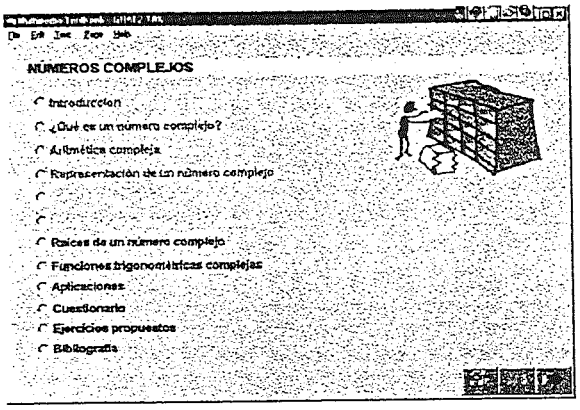
Actualmente son escasos los recursos multimedia disponibles para la enseñanza universitaria de la mayoría de las materias y prácticamente nulos los existentes para la enseñanza de las matemáticas. Existen algunos proyectos en determinadas universidades europeas pero que todavía están en fase de elaboración.

Nosotros hemos entrado en este mundo de la multimedia con la puesta en marcha de dos proyectos en los que se pretende el desarrollo de entornos de aprendizaje interactivo en soporte multimedia para diferentes temas de matemáticas. Estos entornos son flexibles y permiten tanto acceder a los conceptos de un tema, como explorar aplicaciones, realizar cuestionarios con evaluación de resultados, resolver problemas, etc. La herramienta que hemos utilizado en ambos trabajos es el programa **ToolBook**, versión 4.0.

ToolBook [2] es un programa que permite realizar aplicaciones Windows y forma parte de los que se conocen con el nombre de lenguajes de autor. Esta herramienta posee un lenguaje de programación de fácil comprensión pero muy potente llamado OpenScript, de manera que podemos construir de forma rápida aplicaciones que contienen los elementos típicos de una aplicación Windows (ventanas, botones, menús,....). Con **ToolBook** se pueden construir entornos de gran complejidad que superan a muchas de las aplicaciones desarrolladas en lenguajes convencionales. También se utiliza para la creación de prototipos de programas y demostraciones ya que permite combinar de forma rápida y sencilla texto, gráficos, vídeo, animación y sonido.

La utilización de **ToolBook** en el desarrollo de programas educativos permite la creación de aplicaciones en las que, de forma sencilla y rápida se tiene la posibilidad de cambiar el flujo de la información según las necesidades del usuario, relacionar palabras, incluir cuestiones y evaluar los conocimientos alcanzados, activar animaciones, incorporar lenguaje hablado, etc. Todos estos elementos aplicados a la enseñanza hacen que el aprendizaje se pueda realizar de forma individualizada, ajustándose a las necesidades particulares de cada persona y a las específicas de cada tema.

5.1. Unidades interactivas de Matemáticas



El objetivo final de este proyecto es la elaboración de un texto de matemáticas multimedia que contemple los siguientes temas: Números Complejos, Funciones de una variable, Integración, Algebra Matricial y Geometría.

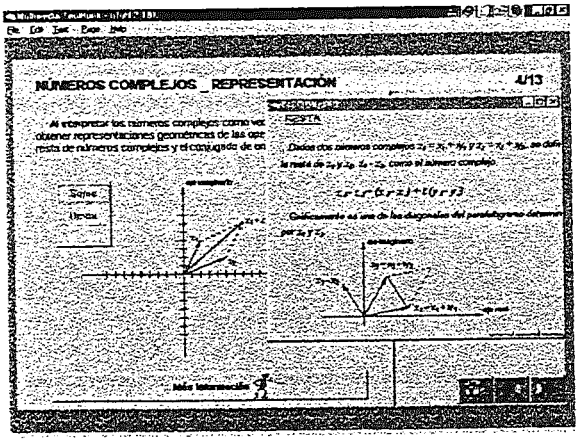
Cada tema pretende cubrir desde los prerrequisitos de entrada a la Universidad hasta el nivel de primer año de carrera. Cada uno de ellos se inicia con un índice de contenidos a los que se puede acceder con enlaces hipertexto, y dispone además de un glosario de conceptos, una colección de problemas resueltos, que permiten al alumno su autoevaluación, una colección de problemas propuestos y una bibliografía.

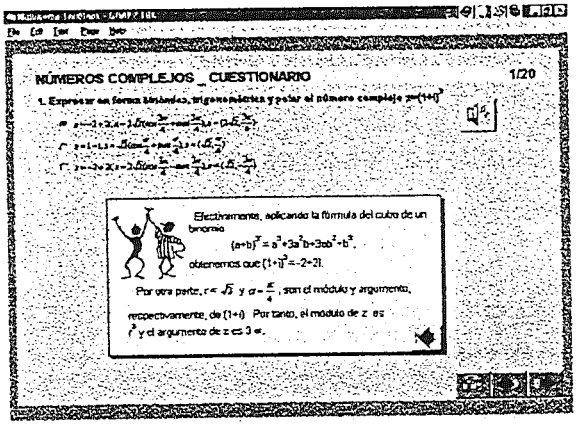
La elección de un tema nos lleva automáticamente a la tabla de contenidos. Estos están distribuidos en orden creciente de dificultad, lo que se distingue en la pantalla por la utilización de diferentes colores.

Los elementos multimedia utilizados en este trabajo son: texto, gráficos, sonido y animación.

Dentro del texto debemos resaltar el sistema de hipertexto. Las posibilidades de búsqueda y recuperación de información que proporciona este sistema son enormes pero creemos que deben ser utilizadas adecuadamente sin abusar de ellas. En lugar de diseñar un sistema de hipertexto muy elaborado, hemos optado por establecer vínculos directos entre las principales palabras (resaltadas en el texto) de forma que pinchando en ellas se llegue a temas relacionados o al diccionario.

Como ya hemos indicado, cada tema





contiene problemas resueltos que permiten al alumno autoevaluarse. Cada problema dispone de tres posibles respuestas de las que sólo una es cierta. Sea cual sea la elección el programa proporciona una explicación, bien para confirmar que la respuesta elegida es la correcta o para indicarnos por qué no lo es y que debemos seguir intentándolo.

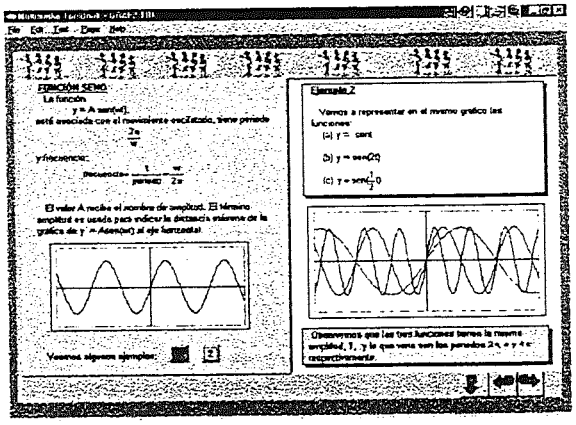
El sonido es quizás el elemento multimedia que más estimula los sentidos. La forma en que se utilice el sonido puede significar la diferencia entre una presentación multimedia normal y otra espectacular.

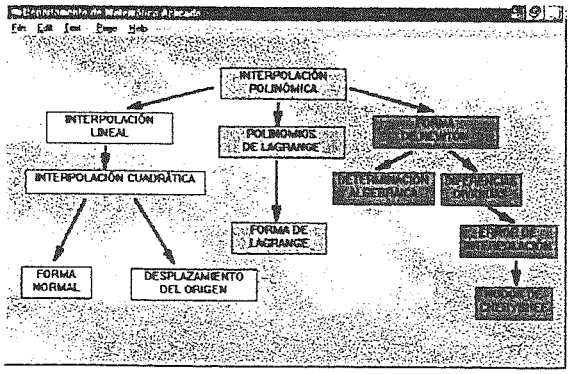
En este trabajo el sonido se limita, de momento, a música al arrancar el programa y a unos "aplausos" cuando se acierta la respuesta correcta en cada problema.

Los gráficos son posiblemente el elemento más importante de un proyecto multimedia, ya que el alumno juzga el trabajo, sobre todo, por su impacto visual.

Los elementos gráficos normalmente pueden dimensionarse, colorearse o hacerse transparentes, colocarse enfrente o detrás de otros objetos, o hacerse visibles o invisibles con una orden. Hemos utilizado gráficos para representar funciones, para dar una interpretación geométrica de los conceptos, en algunos efectos de animación, en definitiva, siempre que la utilización del gráfico refuerza la explicación dada en el texto.

La animación agrega impacto visual a un proyecto multimedia. Se puede animar el proyecto completo, o se puede animar ciertas partes, acentuando determinadas cosas y dándoles más





movilidad.

5.2. Interpolación Polinómica

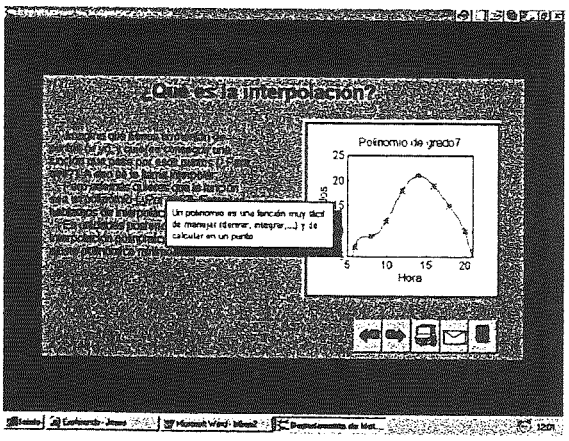
Esta es la primera de una serie de unidades temáticas que hemos empezado a desarrollar en entorno multimedia y que van a cubrir el contenido de la asignatura Laboratorio de Matemáticas, de primer curso de la E.T.S.I. de Telecomunicación.

El objetivo de este recurso es que el alumno disponga de una referencia actualizada que le permita aprender, recordar, ampliar, ..., a lo largo de la carrera, cualquier tema de carácter básico que o bien no se ha explicado o no se ha hecho con la suficiente profundidad.

El desarrollo y los elementos multimedia utilizados en esta unidad temática son totalmente análogos a los descritos anteriormente. Quizás convenga resaltar que, en este caso, hemos optado por presentar los contenidos con una estructura de árbol. Ello permite al alumno seguir un orden lógico a la hora de abordar dicho tema

El resto de elementos utilizados: enlaces hipertexto, gráficos, etc., siguen una filosofía similar a la comentada en el otro trabajo.

En definitiva, en estos proyectos intentamos crear recursos didácticos complementarios o, en su caso, alternativos a la docencia presencial, promoviendo la elaboración y utilización de herramientas multimedia, estimulando la colaboración y el intercambio de experiencias con profesores de otros centros nacionales e internacionales y fomentando el desarrollo de actividades como la "tutoría por correo electrónico", el tablón de anuncios en la Web, etc.



Referencias

[1] Hueso, J.L., Roca, A., y Torregrosa, J.R. "Matemática Aplicada. Prácticas con ordenador". *SPUPV 92.674*

[2] Multimedia Toolbook, 4.0. Manual del usuario

HERRAMIENTAS MULTIMEDIA PARA LA GENERACIÓN DE TUTORES INTELIGENTES SOBRE LA WEB BASADA EN MODELOS.

*P. Domingo, *A. García Crespo, **V. Martínez Orga, *B. Ruiz, *F. García
* Departamento de Informática
Universidad Carlos III de Madrid
c/ Butarque, 15. 28911 Leganés (Madrid).
Tel.: 91.624.99.17
Fax: 91.624.94.30
e-mail: pdgar@rioja.uc3m.es
**Departamento de Inteligencia Artificial
Facultad de Informática
Universidad Politécnica Superior de Madrid.

One of the most relevant implementation of Artificial Intelligence in Education are Intelligent Tutoring Systems (ITS). ITS developments started 20 years ago. Nowadays, ITS comes from labs to classrooms offering spectacular improvements in learning process. Unfortunately, they are several problems that haven't allowed these systems to be spread as much as it would be desirable. First of all, ITS development is very expensive. Secondly, pedagogical component is very small because it is quite difficult to be implemented since this kind of knowledge is poorly formalised. Finally there is an additional problem, as in any software development, is how to update different software versions.

Auto-regulated Intelligent Tutoring Systems are the newest systems in computer based education. AITS offers a very broad capacity in education considering its perfect synchronisation with pedagogical models.

This paper presents design and development of EDU-EX a tool for AITS generation based on models on Web environment. System knowledge is stored in a Knowledge Base. The knowledge modelisation is an areas tree. The tool has been developed on the Web in order to facilitate learning and teaching materials distribution. Web environment also, allow us to run our system in almost every current platform. Finally, it is necessary to mention that user interface tool is very easy to use. The tool offers also a help on-line similar to Windows'95 and avi files in which step by step demos are presented to student.

Key Words: Auto-regulated Intelligent Tutoring Systems, Knowledge Based System, Intelligent systems for learning and training.

1. Características Básicas.

Podemos definir un Tutor Inteligente como el sistema de software que es capaz de adaptarse al estudiante durante su proceso de aprendizaje para sacar el máximo rendimiento del esfuerzo realizado por el alumno. Los Tutores Inteligentes también permiten el desarrollo de cursos interactivos simplificando la transmisión de conocimiento de los expertos humanos sin ninguna restricción.

Un Tutor Inteligente detecta y controla las variables que caracterizan a cada estudiante: capacidad de asimilación, frecuencia de estudio, tiempo de respuesta, cambio de los parámetros de estudio debido a diferentes factores, etc. El objetivo no es únicamente la detección de estas variables sino el diagnostico de las que se han modificado y cual ha sido la causa de esta variación. Dependiendo de este diagnostico, podemos deducir las necesidades de los alumnos en cada momento de las etapas de aprendizaje. Estos mecanismos de diagnostico deben evolucionar y estar en continua modificación para que se adapten de la mejor forma posible al perfil de aprendizaje del estudiante en cada momento.

Un Tutor Inteligente también tiene que ser capaz de detectar el nivel de comprensión de la materia que esta siendo explicada al alumno. Este nivel de comprensión será función de los contenidos del curso y de los valores de las variables que el propio alumno introduce durante el desarrollo de una sesión con el sistema. El Tutor Inteligente debe de ser capaz de ir adecuando la estrategia pedagógica en cada instante. Esta estrategia pedagógica incluye tanto la granularidad de los contenidos que van a ser presentados como su soprte físico (texto, voz, imagen ...). Debe de ser capaz por ejemplo de elegir el momento correcto para efectuar la revisión de un tema teniendo en cuanta valores de variables como los ratios de estudio, la complejidad de los contenidos, cuantas veces o con que frecuencia explicar un contenido, etc. De acuerdo con todo lo dicho anteriormente un Tutor Inteligente debe de ser capaz de adaptarse a las características del estudiante utilizando mecanismos dinámicos. Es

esencial para el proceso de aprendizaje sea lo mas fructífero posible que la enseñanza este completamente personalizada. El mayor éxito de un Tutor Inteligente es el ser capaz de obtener la máxima eficiencia del esfuerzo realizado por el alumno. Es necesario encontrar el momento más adecuado para que una lección se revise o cuando introducir un concepto nuevo. De todo lo anteriormente expuesto parece obligado que el sistema sea desarrollado por profesores y pedagogos utilizando la herramienta que es presentada en este articulo. Otro punto muy importante a considerar es que el sistema debe de ser capaz de detectar los errores cometidos por el estudiante tan pronto como sea posible. La detección precoz de estos conceptos mal comprendidos por los alumnos reduce considerablemente las dificultades de los estudiantes. De este modo es posible evitar que los errores se hagan crónicos, explicando y profundizando los puntos más complicados a los estudiantes.

EDU-EX ha sido desarrollada sobre la Web para poder así aprovechar todas las ventajas que ofrece esta nueva forma de trabajo. Entre estas podremos citar: Distribución de los materiales de enseñanza en todo el mundo y la posibilidad del desarrollo de Tutores Inteligentes que utilicen esta herramienta en casi todas las plataformas que existen en la actualidad.

El Tutor Inteligente se crea en un entorno PC/Windows. El curso es desarrollado por expertos: Experto en el dominio del curso que queremos desarrollar, experto en pedagogía y experto en informática. Una vez que se ha completado el contenido del curso se almacena en el servidor tal y como se muestra en la Fig. 1. Posteriormente, diferentes alumnos pueden utilizar este curso sin más que conectarse a las páginas Web del curso y las características más importantes de cada uno de los alumnos se almacenaran en un servidor permitiendo así la enseñanza personalizada que es más importante objetivo de este tipo de sistemas. Con este nuevo enfoque se palian algunos de los problemas de implementación mas clásicos. Cada estudiante puede conectarse a su propio Tutor Inteligente personalizado en cualquier punto geográfico y en casi todas las plataformas actualmente mas extendidas.

La Fig. 1 nos muestra como se implementa un sistema y como es utilizado por los estudiantes accediendo a las páginas Web del curso.

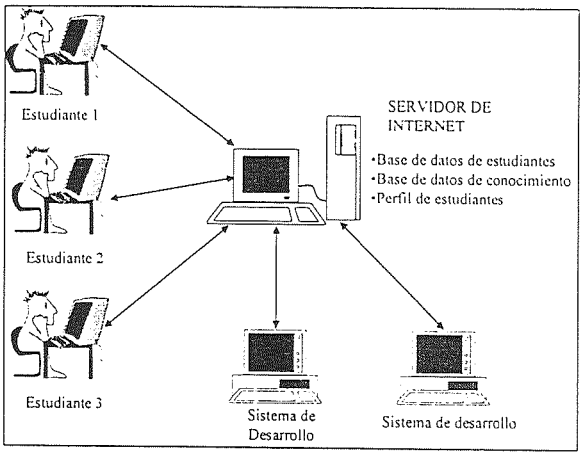


Figura 1. Implementacion del Sistema

2.- Herramienta para el desarrollo del Sistema.

La herramienta para el desarrollo de Tutores Inteligentes EDU-EX esta basada en la tecnología para el desarrollo de sistemas expertos, en concreto los sistemas de soporte a la toma de decisiones. Cualquier materia puede ser modelizada utilizando esta herramienta, sin mas que tener en cuenta los niveles de conocimiento adaptados para cada área y a cada estudiante y verificando la asimilación del conocimiento del estudiante. EDU-EX utiliza objetos para poder estructurar la Base de Conocimientos. El Conocimiento esta almacenado en las propiedades de los objetos. El planificador de áreas es el modulo del sistema que procesa la información que previamente se ha almacenado en la Base de Conocimiento. El objeto mas importante de EDU-EX son las ÁREAS. El resto de los objetos tales como ACCION, DEMOS, TEST, DECISIÓN etc. sirven como soporte. El planificador es el modulo del sistema que procesa toda la información previamente almacenada en la Base de Conocimiento. Mediante el uso de las áreas , el planificador realiza el soporte a las decisiones pedagógicas y la gestión de dichas decisiones.

Las funciones principales del planificador son: Realizar un diagnostico del nivel de conocimientos adquiridos por el estudiante. Sugerir cuales son los conocimientos necesarios que deben mostrarse al alumnos es decir intentar adaptar mejor los contenidos al alumno. Adaptar estos contenidos a cada estudiante en cada momento. Verificar que las decisiones tomadas resuelven y mejorar los ratios del estudiante. EDU-EX es un sistema no monotonico. Esto significa que en un momento de la sesión, el usuario puede cambiar las respuestas dadas previamente. El planificador comienza otra vez el proceso de razonamiento automáticamente, teniendo en cuanta los valores de los test realizados hasta el momento por el estudiante y la anulación de

algunos de esos test que están situados en la red por debajo del valor que se ha modificado.

3. Descripción del diagnóstico.

Para cada una de las características del estudiante que se identifican al comienzo, el planificador identifica su conocimiento asociado y determina cual área se debe ejecutar. Esta fase de diagnóstico es la mas importante de todo el proceso y se realizará en función de los valores previos del conocimiento del alumno y el perfil del alumno considerado por el sistema.

3.1. Orden en el que el planificador examina los objetos.

La información previa almacenada en la base de conocimientos esta organizada de una forma jerárquica. Esta organización es en forma de red de áreas cuyo objeto padre es el objeto inicio. El planificador examina esta red primero en profundidad y luego en amplitud, es decir primero lo examina de arriba hacia abajo y después de izquierda a derecha comenzando por el objeto inicio, hasta que encuentra un área se ha superado o no. Cuando esto sucede, debido a la forma en que esta organizada la información, el planificador sabe que algún nodo hijo va a ser activado. Entonces elige o ignora esta rama de la red y continua examinando la próxima rama derecha. El planificador continua repitiendo este proceso hasta que encuentra un área que no esta superada o hasta que no hay mas objetos que puedan ser examinados. Una vez que el planificador a encontrado algún área no superada pasa a analizar el objeto decisión_pedagógica.

3.2 De que forma el Planificador entra a explicar un área.

Los objetos del tipo área tienen una propiedad llamada "condición_de_entrada" cuyo valor es una serie de funciones con sus argumentos respectivos. Cada vez que el planificador quiere conocer si un área esta no superada (el estudiante necesita adquirir más conocimiento), evalúa esta serie de funciones y obtiene un valor. Este valor puede ser:

- Verdadero: Significa que este área no esta superada, luego entra en esta parte de la red y le muestra al alumno el conocimiento que le hace falta.
- Falso: Lo cual significa que el área esta superada, luego no entra en esta parte de la red. El alumno ya tiene un conocimiento suficiente.
- Desconocida: No podemos saber si el área esta superada o no debido a que existen valores desconocidos como argumentos de las funciones.

Si la propiedad "condición_de_entrada" del área esta vacía, es decir no tiene valor consideraremos que el área no esta superada, luego el alumno tiene que entrar a estudiar el área.

4. Resolución del área.

Una vez que se ha comprobado que el área no esta superada, se deberá mostrar al estudiante los conocimientos oportunos del tema en concreto y para este tipo de estudiante. El objeto acción, asociado a este área, permite determinar que demostraciones y test serán presentados al alumno para que consiga la correcta comprensión de los temas.

4.1. Estrategia de decisión_pedagógica.

Cada área puede tener una o varias decisiones. Una decision_pedagogica básicamente contiene información acerca de los contenidos o materias que necesita aprender el estudiante y de que forma se la vamos a presentar. Una decisión_pedagógica puede ser ejecutada si y solo si cuando se evalúa la propiedad requisitos_previos tiene como resultado "verdadero" y si se ha tratado de ejecutar menos veces que el valor de la propiedad "numero de decisiones". La propiedad "número de decisiones" indica el numero máximo de veces que vamos a mostrar un contenido determinado a un estudiante.

Si un área que es nodo final tiene varias decisiones, estas se estudiaran por orden realizando la primera que pueda ser ejecutada.

Una vez que esta decisión_pedagógica pueda efectuarse, ya sea porque se han cumplido los requerimientos o bien porque no se ha efectuado el numero máximo de veces, el sistema sugiere al alumno que acciones de esta decisión_pedagógica se deben efectuar. Estas acciones incluyen las demostraciones a realizar y los test con los cuales evaluaremos el nivel de conocimientos adquiridos.

5. Verificación del nivel de conocimientos adquiridos por el alumno.

Una vez que se han ejecutado las acciones (demostraciones y test) sobre una determinada materia comienza la fase de verificación del nivel de conocimiento adquirido por el alumno. Es una fase muy importante, ya que hay veces en las que el alumno no ha adquirido correctamente los conocimientos o bien tiene algunos errores previos en su conocimiento de la materia que estamos impartiendo. En esta etapa se decide que va a pasar

y en que lugar de la red seguiremos presentando las diferentes materias al alumno Si no se ha superado el área, el área inicial no superada comienza de nuevo.

Una vez que se ejecutan las acciones de una decisión_pedagógica pueden entonces ocurrir dos supuestos:

- Que la evaluación de como resultado verdadero. Entonces el alumno ha adquirido los conocimientos necesarios y el alumno puede pasar a otra área, finalizar la sesión, etc dependiendo del valor de la propiedad estrategia.
- Que la evaluación de como resultado falso. En este caso el sistema examina primero la propiedad si_suspende y después la propiedad estrategia.

5.1. Propiedad si_suspende.

Cuando el alumno no ha adquirido los conocimientos necesarios, el sistema pasa a examinar la propiedad si_suspende. Si esta propiedad esta vacía, el sistema examina la propiedad estrategia. En el caso de que la propiedad si_suspende contenga un nombre de un área, el sistema tomara el objeto inicio como área de inicio y comenzara a examinarlo.

Una vez de que se ha efectuado la propiedad si_suspende, el sistema mirara la propiedad estrategia de la decisión_pedagógica y si se han cumplido los requerimientos explicados en el párrafo anterior pasara a evaluar esta propiedad.

5.2. Propiedad estrategia.

La propiedad estrategia tiene cuatro valores:

- Fin: El sistema terminara la sesión. Tenemos varias opciones diferentes: Se ha superado el área y no queda ningún área que superar.
- Comienzo: El planificador comienza a examinar el sistema desde el área inicio. Para realizar este examen el sistema evalúa y considera todas los preguntas y la fecha de la sesión. El sistema no hace el mismo recorrido del red si no es necesario.
- Continuar: El planificador continua con la evaluación del red de contenidos en profundidad y en anchura, tal y como lo efectuó antes de llegar a la decisión_pedagógica.
- Hacia arriba: Es la estrategia mas compleja de todas. El planificador sube por la misma rama del red por la que bajo volviendo a evaluar la propiedad condición_de_entrada de las áreas y evaluando otra vez las preguntas hasta que encuentra algún área que n esta superada o llega al objeto inicio. Si encuentra algún área que no esta superada el sistema vuelve a realizar la estrategia de la decisión_pedagógica y si llega al objeto inicio continua evaluando las ramas derechas del red que no habían sido examinadas antes.

5.3. Búsqueda de nuevas áreas y vuelta atrás.

Una vez que se entra en un área y los conocimientos han sido presentados y estudiados por el alumno no se considera finalizado el curso hasta que se hayan superado todas las áreas o todos los conceptos erróneos que se puedan haber adquirido durante la sesión. De cualquier forma, si después de llegar a un punto de la red el planificador sabe que hay mas áreas no superadas el sistema elegirá el próximo y volverá a la fase de diagnostico, decisión_pedagógica, etc. Solo cuando todos los conocimientos considerados relevantes han sido adquiridos y el alumno ha completado el curso se finalizara esta fase del sistema pasando a realizar la fase de informes.

6. Conclusiones.

La herramienta para la generación de Tutores Inteligentes Autorregulados propuesta es otra componente que puede ser incluida en los Sistemas Multimedia Inteligentes. Puede ser usado tanto para soporte básico para la enseñanza correcta o como un elemento adicional que haga mas potentes otras herramientas de enseñanza.

Actualmente se esta finalizando el desarrollo de la herramienta en Windows'95. La herramienta ha sido probada en varios sistemas reales con muy buenos resultados. El tiempo de adquisición de conocimiento de unas 200 horas en un entorno tradicional para el desarrollo de sistemas expertos a 10 horas con EDU-EX. La representación del conocimiento y el planificador utilizan un modelo análogo al de un experto humano. El estudio de la representación del modelo de conocimiento como una red de contenidos compartimentado en áreas y las estrategias de decisión_pedagógica son aspectos fundamentales de esta sistema.

Referencias.

[1] Adarraga, P. y Zaccagnini, J.L. (1994). "Psicología e Inteligencia Artificial". *Madrid: Trotta S.A.*

[2] García Crespo, A. (1994). "Estudio, desarrollo e implementación de una herramienta de generación de sistemas expertos basada en diagnosis diferencial y aplicación a un sistema experto en mecánica de la fractura". *Tesis presentada en la Universidad Politécnica de Madrid.*

[3] Maté Hernández, J.L. y Pazos Sierra, J. (1988). "Ingeniería del conocimiento: diseño y construcción de sistemas expertos". *Córdoba, Argentina: Sociedad para estudios pedagógicos*

[4] Waterman, D. A., (1986). "A guide to expert systems". *Addison-Wesley Publishing Company*

[5] Harmon P., King D. (1985). "Expert Systems". *John Wiley & Sons, Inc.*

[6] Hayes-Roth F. D., Waterman A., Lenat D. (1983). "Building Expert Systems". Addison-Wesley.

[7] Anderson, J. (1992) "Intelligent Tutoring and High School Mathematics". *Intelligent Tutoring Systems, 2nd International Conference.*

[8] Kaplan, R., H. Trenholm, D. Gitomer, and L. Steinberg. (1993). A Generalizate Architecture for Building Inteligent Tutoring System. *Proceeding of Applications of Artificial Intelligence Conference : Knowledge-Based Systems in Aerospace and Industry.*

[9] Linden, T. (1993). Tools for a New Generation of Educational Software. *Proceeding of the 13th International Conference on Critical thinking and Educational Reform. Upper Monclair, N.J., August.*

[10] Pstka, J., d. Massey, and S. Mutler.(1988). "Intelligent Tutoring Systems : Lessons Learned". *Hillsdale, N.J.*

[11] Wenger, E. (1987). "Artificial Intelligence and Tutoring Systems". *Los Altos, California.*

[12] Vanlehn, K., S. ohllson, and R. Nason. (1987). "Applications of Simulated Students : An Exploration". *Journal of Artificial Intelligence in Education 5 (2) : 135-175.*

Desarrollo de material docente con Java. Aplicación en la enseñanza en Ingeniería Telemática

Juan José Uncilla, Iñaki Goirizelaia, Eduardo Jacob, Jon Mikel Omagogeaskoa
Area de Ingeniería telemática
Departamento de Electrónica y Telecomunicaciones
ETSII y de IT, Bilbao (UPV/EHU)
Alda. Urquijo S/N - E 48013 Bilbao
Email: {jtpungaj, jtpgoori, jtpjatae, jtaominj}@bi.ehu.es
Tfno : 34 (9) 4 427 80 55 - Fax: 34 (9) 4 441 40 41

Abstract:

This paper presents several interactive applications with their correspondent management tools developed using JAVA applets. These applications are inserted in electronic books using Hyper-G an hypermedia information server. In this book users can run interactive lessons on several subjects as OSI transport level communication protocol and a simulation of hayes modem, and multimedia contents related with them.

1. Introducción

El libro de texto ha jugado y juega un papel esencial en el proceso educativo; en las Escuelas y Facultades, prácticamente todo sigue moviéndose a su alrededor. El profesor los utiliza como base para sus explicaciones, y son consultados por los alumnos para el seguimiento y profundización de las clases.

Pero siendo una herramienta clave para el almacenamiento y la divulgación del conocimiento, son demasiadas las veces en las que se escuchan razones como "Eso no se aprende en los libros" para hacer referencia a aquel conocimiento práctico inadquirible del texto escrito. Las nuevas herramientas telemáticas pueden ofrecernos una respuesta.

Observando el boom que ha supuesto Internet y aprovechando potentes herramientas desarrolladas a su alrededor como son la WWW, una primera aproximación hacia el nuevo soporte, podría ser el crear los nuevos libros de texto como un conjunto de páginas entrelazadas ofrecidas a través de un servidor Web [1]. De esta forma se heredarían las ventajas inherentes a la distribución Web: disponibilidad global, actualización "just in time" de los contenidos, independencia de la plataforma,...

Y además se aportaría una nueva ventaja al proceso de aprendizaje asociada al formato en el que se ofrece la información: el hipertexto. El hipertexto nació de la idea de imitar el funcionamiento del razonamiento humano [2]. Utilizándolo como base de los libros de texto electrónicos, se abandona la visión secuencial que ofrece la escritura clásica, para al igual que se unen unas con otras las ideas en el cerebro, ofrecer los contenidos en forma de unidades de información enlazadas entre sí mediante hiperenlaces. Ante un libro escrito así el lector puede lograr una dinámica de autoaprendizaje completamente personalizada, creando según sus propias exigencias caminos a través de los diferentes conceptos. Podrá quedarse con una primera lectura superficial para obtener una idea básica de lo tratado o cuando así lo desee, le será posible profundizar en los temas de interés siguiendo hiperenlaces a referencias y conceptos relacionados.

Pero para un sistema ambicioso que pretenda sustituir el modelo de enseñanza clásico y plantear una alternativa real a la clase magistral, esto no es suficiente. Sería necesario disponer de más capacidades:

- En cuanto al formato de los libros de texto electrónicos, hay que ofrecer algo más que un simple conjunto de páginas web entrelazadas. Para ello, en este artículo se presenta una solución basada en un servidor de segunda generación (Hyper-G) como alternativa al clásico servidor Web.

- En cuanto a los contenidos, el libro de texto electrónico debe ir más allá del hipertexto. Hay que incorporar contenidos multimedia apoyando las explicaciones textuales (voz del autor explicando diversos conceptos, vídeos de demostración, etc.), y como se describe en este artículo, hay que ofrecer posibilidades de interacción con los contenidos del libro mediante demos y simulaciones en tiempo real. Así el lector pasará de ser el consumidor de información pasivo que viene siendo en el modelo tradicional, para entrar a tomar parte activa en el proceso educativo.

- Es necesario disponer de una serie de herramientas de gestión integradas que faciliten la labor de creación, gestión y mantenimiento de los libros.

2. Nueva generación de servidores: Hyper-G

Como soporte de los libros de texto electrónicos se ha seleccionado el servidor *HyperWave Server* [3], desarrollado en la *Graz University of Technology*. La funcionalidad ofrecida por este nuevo servidor, permiten responder a los requerimientos de nuestro prototipo. Las características más destacadas para ofrecer el servicio de libros de texto electrónicos se describen a continuación.

2.1 Hipermedia estructurada

El servidor Hyper-G ofrece algo más que un conjunto de páginas enlazadas entre sí por medio de hiperenlaces. Añade una nueva entidad al modelo documento-enlace: ***la colección***. Una colección es un objeto compuesto (contenedor), que puede contener documentos u otras colecciones. Para aclarar este concepto puede ser útil pensar en los directorios de un sistema de archivos.

La característica de recursividad inherente en el concepto de colección nos lleva a estructuras jerárquicas como la de la figura 1.

Esto resulta muy útil en el caso de los libros de texto electrónicos ya que:

- Facilita la navegación del usuario. Ofrece un interfaz adecuado para la visualización de la jerarquía anterior. El usuario podría concebir un modelo mental del libro electrónico en su conjunto, algo imposible en el modelo documento-enlace.

- Permite saber en todo momento dónde se encuentra el usuario en el libro. Aunque no se visualice la jerarquía de colecciones, al usuario siempre se le indicará la posición del documento visualizado dentro de la jerarquía. Esta es una medida importante en la lucha contra el síndrome "*lost in hyperspace*".

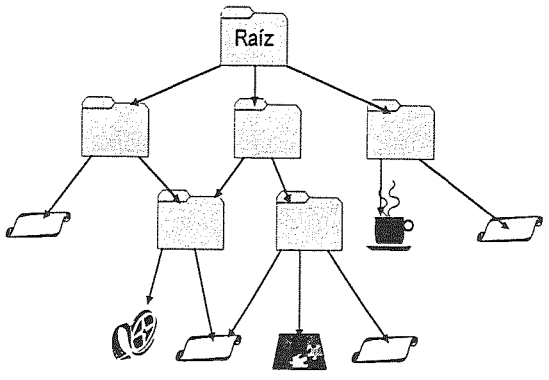


Figura 1: Jerarquía de colecciones

- Se produce un descenso importante en labores de mantenimiento. En el modelo documento-enlace, la incursión de un nuevo documento siempre conlleva la modificación de, por lo menos, otro para enlazar desde él el nuevo documento (sino resultaría inaccesible). Del mismo modo, retirar uno de los documentos insertados supone la eliminación de todos los enlaces hasta el mismo. En Hyper-G, como las colecciones son gestionadas por el sistema, un documento recién insertado se añade automáticamente a la colección que le corresponde sin necesidad de realizar ningún cambio adicional. A la vista del usuario, el nuevo documento aparecerá como un nuevo elemento en la lista de contenidos de la colección.

2.2 Herramientas de búsqueda

Un motor de búsqueda es una pieza esencial en un servidor Web en el que se van a almacenar grandes cantidades de información. Como el servidor Hyper-G esta construido sobre una base de datos orientada a objetos, todos los objetos almacenados (documentos, enlaces, colecciones) están desde su creación dotados de un conjunto de propiedades. Algunas de estas propiedades están indexadas para facilitar labores de búsqueda y localización. En este grupo de propiedades se encuentran: títulos, palabras clave, nombre de autor, fecha de creación,...

Además de esta búsqueda en torno a valores de propiedades, para documentos de tipo texto, el servidor Hyper-G ofrece una herramienta integrada para la búsqueda a lo largo de todo el documento.

2.3 Separación entre documento y enlace

A diferencia de los servidores Web clásicos, en Hyper-G los documentos y los enlaces se almacenan como objetos separados en la base de datos. Esto frente al almacenamiento de los enlaces en el propio documento, ofrece las siguientes ventajas:

- Los enlaces son de doble sentido. Esto quiere decir que desde el extremo final del enlace se puede alcanzar el origen, permitiéndose la navegación en sentido inverso.

- Esta característica del doble sentido garantiza la consistencia de los enlaces. Esto es: si un documento es eliminado, se buscan automáticamente todos los enlaces hasta él y son eliminados sin necesidad de intervención por parte del administrador.

- Como los enlaces son también objetos, pueden tener propiedades de acceso. Se puede así limitar ciertos caminos entre documentos a usuarios con derechos especiales, haciendo que enlaces que

para estos sean visibles, permanezcan ocultos para los usuarios corrientes.

Como principal inconveniente hay que citar la pérdida de control sobre la estructura de la información almacenada, ya que es el servidor el que gestiona de forma autónoma la parte del sistema de ficheros donde residen las colecciones.

2.4 Sistema multiadministrador y multiusuario

La mayoría de los sistemas hypermedia están diseñados como calles de único sentido. Los proveedores ofrecen la información y los usuarios la consumen utilizando cada uno de ellos herramientas diferentes para su labor. En el caso del servidor Hyper-G esto no funciona así. En Hyper-G la inserción de información puede realizarse desde el propio cliente. Para evitar la inserción y/o modificación no autorizada de contenidos, Hyper-G ofrece un sistema de identificación de usuarios basado en ***login*** y ***password***. A diferencia de los servidores clásicos, el servidor ofrece al usuario la posibilidad de identificarse de forma integrada. Esto facilita el control de la inserción y/o modificación de documentos de las colecciones que estén bajo su responsabilidad desde el mismo navegador.

3. Aplicaciones de valor añadido en un servicio de libros electrónicos

En este punto se describe la tecnología empleada para desarrollar las aplicaciones que posteriormente conformarán el servicio de libros electrónico implementado. Por una parte tenemos las aplicaciones que se ofrecen al lector (demos interactivas, información multimedia) y por otra las herramientas que se emplean para gestionar las mismas, que en general serán empleadas por el responsable del sistema.

3.1 Demos interactivas en tiempo real

El objetivo principal del servicio es transformar la figura del lector para que de ser una entidad pasiva, pase a ser un ente activo que interactua con los contenidos del libro. Si se le están ofreciendo explicaciones sobre el funcionamiento de algún dispositivo o sistema, mediante estas demos en tiempo real se le ofrecerá una simulación interactiva para que le sea posible obtener una experiencia "real" sobre el tema tratado.

Estas demos aparecerán insertadas entre las explicaciones como si de imágenes se tratara. Pero a diferencia de estas, no se limitarán a ofrecer información estática. El espacio que ocupen en la página del navegador se convertirá en un escenario interactivo y el lector tendrá la posibilidad de tomar parte en lo que en él se represente.

Como materias en las que se puede emplear esta tecnología podemos citar aquellos contenidos en los que la variable tiempo forma parte del proceso (por ejemplo un sistema de comunicaciones) y aplicaciones que simulan el comportamiento de equipos (por ejemplo un modem).

Para que todo esto sea posible, se utilizan applets de JAVA [4]. Los applets de JAVA son pequeños programas que pueden crear animaciones, contenidos multimedia, operaciones por la red, interactividad real, etc. Todo lo que un programa clásico pueda hacer, lo puede hacer un applet de JAVA y además:

- Tiene la capacidad de viajar a través de la red pegado a una página Web, como muestra la figura 2.
- Al llegar al cliente se ejecutará sobre la propia página visualizada por el navegador.

Como se puede ver, son una herramienta casi a medida para poder ofrecer el servicio de demos interactivas insertadas entre las explicaciones de los libros de texto electrónicos. Para ello habrá que crear los applets codificando el comportamiento requerido de las demos a desarrollar en el lenguaje de programación JAVA, y luego enlazar dichos applets con las páginas de los libros de texto electrónicos. Para que la sensación de interactividad sea aceptable, dichos applets no deben ser excesivamente grandes, por lo que este es un aspecto a considerar en su diseño.

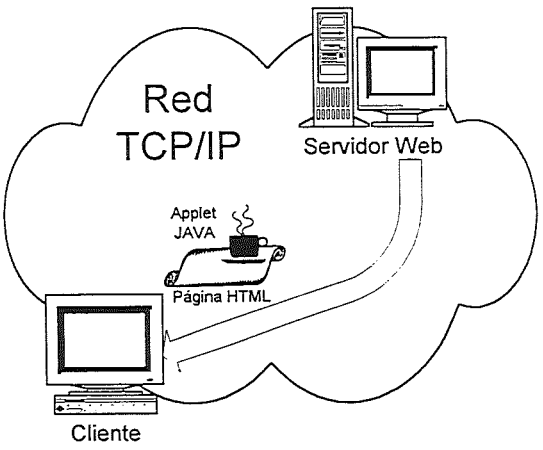


Figura 2: Applets a través de la red.

3.1.1 Creación de demos interactivas

Para dar el primero de los pasos, se describen a continuación las características que ofrece el

lenguaje JAVA a la hora de crear las demos interactivas.

3.1.1.1 Clase Applet. Interactividad con el navegador

Para que la aplicación pueda relacionarse con el navegador de una forma interactiva, JAVA ofrece la clase Applet. Cuando en vez de una aplicación convencional sea un applet lo que se desee crear, se utilizará una clase derivada de Applet como base del diseño.

Cuando el navegador muestra una página y se encuentra con que en ella llega un applet, busca dentro del código esta clase derivada de Applet y de encontrarla crea un objeto desde la misma. De ahí en adelante se dedica a enviarle distintos mensajes:

- Tan pronto como se crea el objeto, son tres los mensajes enviados: uno para su inicialización (*init*), el segundo para la puesta en marcha (*start*) y el último para que pueda actualizar el área que le corresponde dentro de la página del navegador pintando en ella lo que desee (*paint*).

- Una vez arrancado el applet, durante su ciclo de vida son estos los mensajes que le pueden llegar del navegador:

    - Si en el navegador se abandona la página que contiene el applet para ir a una nueva, el navegador envía un aviso en forma de mensaje (*stop*) para que el applet interrumpa su ejecución.
    - Si en el navegador se vuelve a visualizar la página que contiene el applet para volver a iniciar la ejecución interrumpida se recibe un mensaje del tipo *start*.
    - Los mensajes *paint* también se repiten durante el ciclo de vida del applet. Esto ocurre cada vez que hay que refrescar la zona que el applet ocupa dentro de la página del navegador (cuando otra ventana se superpone y luego desaparece, etc.)
    - Cuando se mueve el ratón o se pincha con él dentro del área correspondiente al applet en la página, el navegador también da cuenta de ello mediante los mensajes oportunos.

- Finalmente cuando se debe no interrumpir, sino detener indefinidamente la ejecución del applet, por ejemplo se cierra el navegador, éste enviará un mensaje para indicarlo (*destroy*).

Por lo tanto, diseñando como clase principal de la demo una derivada de la clase Applet ofrecida por JAVA, para lograr la interactividad deseada a través del navegador, solo habrá que recoger y atender adecuadamente los mensajes arriba analizados.

3.1.1.2 AWT (Abstract Window Toolkit)

Es una librería de clases orientada a objetos. Entre las clases que la componen, se puede hallar todo lo necesario para crear interfaces de usuario de calidad. Entre otras, hay clases para crear ventanas emergentes de la página en la que está confinado el applet, clases para crear diálogos de mensaje, clases para menús, botones, listas desplegables, campos de edición, áreas de texto,...

3.1.1.3 Gestión de eventos

Un programa clásico que quisiera saber de las acciones realizadas por parte del usuario, debía conseguir por sí mismo dicha información. Para esto, se construye un gran bucle y en él se testea una y otra vez si ha ocurrido algo (pulsación de una tecla, movimientos del ratón, etc.). Java abandona esta técnica conocida como *polling* para solucionar el problema empleando un nuevo paradigma: *la orientación a eventos.*

Todas las acciones realizables por parte del usuario, entran en un gran saco llamado *eventos*. Un evento puede describir completamente cualquier acción. Así, las acciones sobre los elementos del AWT, las acciones del ratón, teclado etc., todas poseen sus eventos asociados. Todos estos eventos creados por el usuario, en vez de tener que ser muestreados por el programa, son recogidos por el sistema JAVA y al recibir cualquier evento se lo hará saber al programa por medio de mensajes de aviso de eventos, indicando lo ocurrido.

3.1.2 Inserción de demos en las páginas de los libros electrónicos

Una vez creadas las demos mediante applets de JAVA, el siguiente paso será insertarlas entre los textos explicativos de las páginas de los libros de texto electrónicos, como aparece reflejado en la figura 3. Estos son los pasos necesarios para conseguirlo:

1. Crear los applets en un entorno de desarrollo JAVA (JDK).
2. Llevarlos a la máquina en la que se encuentra el servidor Hyper-G.
3. Crear los enlaces entre páginas de los libros de texto y los applets.

En el último de los pasos no se utilizará el procedimiento habitual que se emplea para crear enlaces entre documentos en el servidor Hyper-G. Toda la sencillez y automatización del proceso

normal se pierde para el caso de los applets. Para explicar cómo se deben crear estos enlaces se va a utilizar un ejemplo concreto:

- Supongamos que hemos creado un applet y lo hemos llamado *x.class*. El primer paso será depositarlo en el servidor en una localización conocida. Por ejemplo, en el directorio */home/HTTPd/java*.

- Hecho lo anterior, el segundo paso será dar parte de forma manual a la base de datos de objetos del servidor Hyper-G para insertar en ella como objeto de tipo "Program" el applet desarrollado. Esto no se podrá realizar desde el cliente Hyper-G. Será necesario utilizar la aplicación *hginsdoc* del servidor. Si *NombreCol* es el nombre de la colección donde queremos guardar el applet el comando a emplear para la inserción será:

  hginsdoc -type Program -name "java/x.class" -pname NombreCol -title "Applet X" -path /home/HTTPd/java/x.class

- Una vez insertados los applets en la base de datos de objetos, habrá que hacer referencia a ellos desde las páginas de los libros de texto. Para ello se procederá a editar las páginas y allí donde se desee insertar un applet habrá que utilizar las marcas habituales de <APPLET> utilizando como parámetros CODE y CODEBASE como se ve a continuación:

  <APPLET CODEBASE="/java" CODE="x.class"> </APPLET>

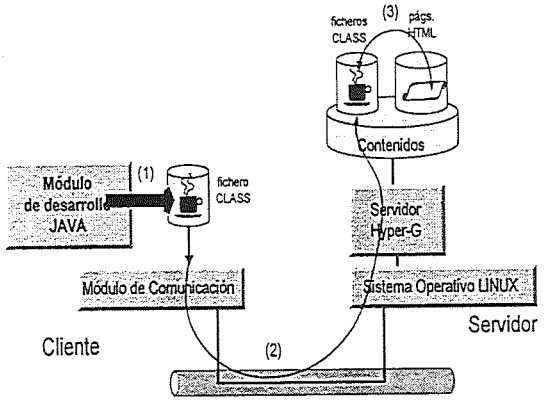


Figura 3: Inserción del applet

3.2 Información multimedia

Para apoyar las explicaciones y conceptos expuestos en el libro, es muy recomendable la inclusión de vídeos. En primer lugar será necesario obtener la información mediante una cámara de vídeo (vídeos), un micrófono (audio) o un escáner (imágenes fijas) y almacenarla en el formato correspondiente como puede verse en la figura 4.

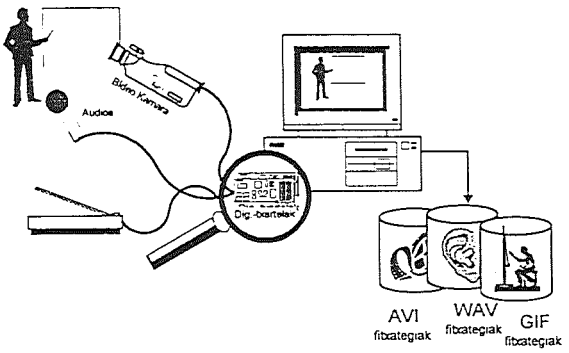


Figura 4: Generación de información multimedia

Tras crear la información, es necesario introducirla en el libro de texto electrónico, para lo cual tenemos varias opciones:

- Mediante un hiperenlace relacionar el texto con el fichero AVI. Así al pinchar se arranca el *plug-in* configurado en el navegador y se visualiza allí el vídeo.

- Si se va a trabajar con visualizadores Microsoft (a partir de IE2.0), se utilizará la marca <IMG DYNSRC="x.avi" CONTROL>.

- Utilizar la API del JMF (Java Media Framework) [5]. Se desarrollara un applet para visualizar el fichero AVI que se le pase como parámetro en la etiqueta APPLET del fichero HTML.

En la segunda opción su utilización es completamente directa (no hacen falta nuevas DLLs ni nada por el estilo) si se utiliza Internet Explorer como navegador, mientras que en la tercera es preciso instalar nuevas DLLs en el cliente. Esta última se usaría si además de reproducir un vídeo se quisieran hacer más cosas como, por ejemplo, sincronización con las acciones sobre otro applet de la pantalla (utilizando la comunicación entre applets que se ejecutan en la misma página del navegador).

3.3 Herramientas de gestión

Las herramientas de gestión son necesarias para facilitar la realización de tareas de creación de información y su mantenimiento, así como para disponer de sistemas de monitorización sobre el servidor. Para integrar las herramientas de gestión en el servicio de libros de texto electrónicos, son cuatro

los elementos que hay que poner a trabajar de forma cooperativa mediante la arquitectura adecuada:

- Los **programas de gestión** que son la base de las aplicaciones de gestión.
- La **base de datos** que se configura como soporte de la información utilizada.
- **El servidor Web** a través del que se debe ofrecer el servicio de gestión.
- **El cliente Web.** Que deberá ser el único software necesario para que el cliente pueda acceder a los servicios.

El camino clásico para dar solución al problema y relacionar los cuatro elementos ha sido el establecimiento de una arquitectura basada en el CGI (Common Gateway Interface) [6] como se puede ver en la figura 5.

En esta arquitectura a los programas de gestión se les conoce como programas de mas allá del servidor, ya que para el cliente estos programas están como ocultos detrás del servidor Web. Para llegar hasta ellos esto es lo que hay que hacer:

En el servidor Web se crean unas páginas especiales llamadas FORMS. Estas páginas pueden ser accedidas desde el cliente como cualquier otra utilizando su URL ((1) y (2)). Una vez en el cliente, es posible escribir en ellas la información que se quiere hacer llegar a los programas CGI y enviarla de vuelta hasta el servidor Web (3).

Cuando esta información llega al servidor Web, utilizando el interfaz CGI el servidor ejecuta el programa adecuado y le pasa la información recogida desde el cliente (4). Este programa realiza sus cálculos y operaciones con la información adquirida, que pueden conllevar operaciones sobre la base de datos ((5) y (6)), y los resultados obtenidos se vuelven a pasar al servidor Web en forma de página HTML utilizando de nuevo el interfaz CGI (7), y desde aquí, siguiendo el camino habitual la página creada dinámicamente es enviada al cliente (8).

Del uso de la arquitectura analizada se desprenden estos inconvenientes:

- El tipo de información requerible del usuario a través de los FORM está limitada.
- Hay limitación en cuanto al formato también en la información devuelta desde los programas de más allá del servidor (gráficas, etc.).
- Cada iteración tipo pregunta-respuesta entre el cliente y el programa de mas allá del servidor, hay que realizarlo vía protocolo HTTP con el importante gasto de recursos que conlleva el establecimiento y consiguiente liberación de la comunicación TCP en cada uno de estos pasos.

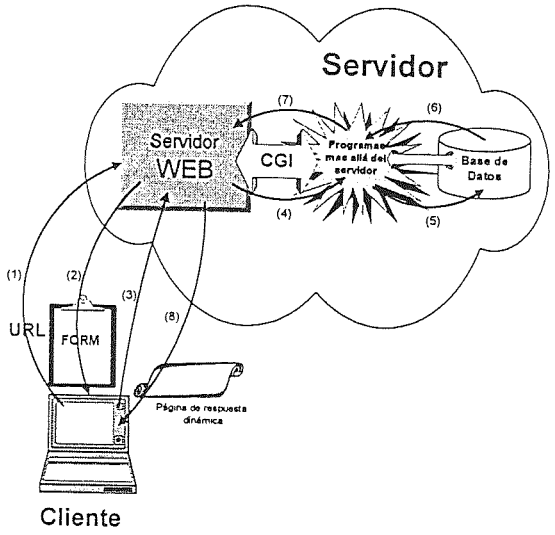


Figura 5: Esquema de funcionamiento usando CGIs

- Este bajo rendimiento degrada la sensación de interactividad.

Descartada la aproximación clásica, la solución adoptada para ofrecer los servicios de gestión se desarrolla en torno a estas dos decisiones:

- Los programas de gestión se traerán desde más allá del servidor hasta el propio cliente. Así:
  - Se libera al servidor de la ejecución de dichos programas.
  - Se consigue una interactividad real entre usuario y programa ya que ambos confluyen en la misma localización.
- Cuando los programas que ahora se ejecutan en el cliente necesiten trabajar contra la base de datos, se creará una conexión directa hasta la misma abriendo una conexión TCP hasta el puerto donde escuche el servidor de base de datos. Así:
  - El consumo de recursos es mucho menor comparándolo con el del método clásico vía CGI. Antes había que crear y destruir conexiones TCP (transporte vía HTTP). Ahora en cambio, la conexión TCP se realiza una sola vez (directa hasta el servidor de base de datos, sin pasar por el servidor Web) y se mantiene activa hasta que se deje de necesitar relación con la base de datos.

En la figura 6 se ofrece una visión de esta nueva arquitectura. En ella, los programas de gestión se crearán como applets de JAVA y para utilizarlos se seguirán los siguientes pasos:

- Se pedirá una página especial al servidor que hará de puerta hacia las aplicaciones de gestión (1) con acceso sólo para el administrador.
- El servidor servirá lo demandado (2).
- En esta página especial vendrá pegado el applet de gestión que una vez llegado al navegador del cliente se ejecutará en la misma página visualizada.

Con esto queda conseguida la primera premisa de diseño que trataba de traer los programas de gestión desde *más allá del servidor* hasta el cliente. Lo que queda ahora es crear el camino de comunicación entre la aplicación en el cliente y la base de datos en el servidor:

- Cuando el applet diseñado para trabajos de gestión necesite trabajar contra la base de datos del servidor, construirá una vía de comunicación TCP hasta el puerto donde escucha el servidor de base de datos. Para ello se utilizarán las facilidades que ofrece el lenguaje de programación JAVA para la gestión de comunicaciones a nivel de transporte utilizando sockets.
- Hecho esto será necesario incluir en los applets de gestión un módulo especial encargado de la comunicación con la base de datos hablando su propio protocolo a través del socket TCP. Así las consultas a la base de datos se podrán realizar de una manera directa ((3) y (4)).

Por lo tanto, a la hora de diseñar los applets para labores de gestión, habrá que concebirlos compuestos por dos módulos como se representa en la figura 7.

4. Prototipo realizdo

4.1 Demos interactivas

Como ya se ha comentado en el área de la enseñanza existen dos campos de aplicación inmediatos para el servicio de demos interactivas planteado:

- Por un lado como apoyo a explicaciones en las que intervenga el concepto de tiempo (protocolos de comunicación, algoritmos de búsqueda y ordenación, etc.). Se sustituyen las típicas gráficas en las que un eje representa el tiempo por una simulación en la que el tiempo es una magnitud real.

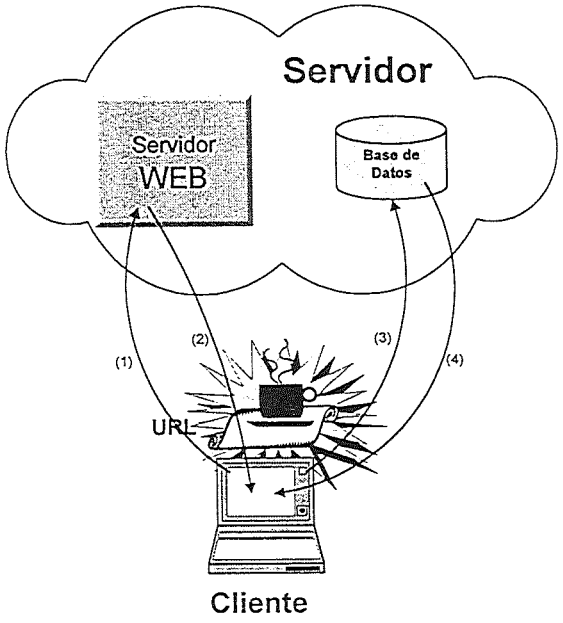


Figura 6: Esquema de funcionamiento usando JAVA. Conexión directa con la base de datos.

- Y por otro, para la simulación de comportamientos interactivos tipo acción reacción (consultas SQL a bases de datos, comandos HAYES para modems, etc.). Se sustituye un conjunto de ejemplos puntuales por la posibilidad de poder poner en práctica todas las posibilidades que al usuario se le ocurran.

A continuación se describen dos ejemplos de demos implementadas en el prototipo realizado.

4.1.1 Funcionamiento de la entidad de nivel de transporte OSI

Mediante esta demo, se ofrece al usuario la posibilidad de analizar una sesión de comunicación entre dos entidades de nivel de transporte tomando parte activa en la misma.

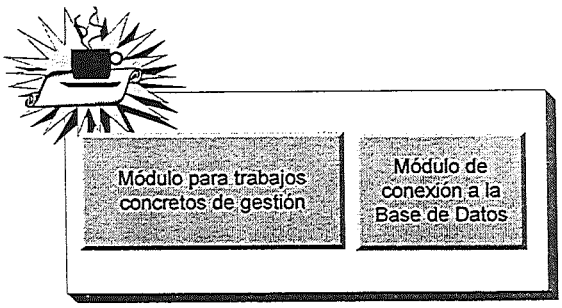


Figura 7: Diseño general de los applets para herramientas de gestión

En la misma página en donde se analizan los conceptos relativos a estas entidades de nivel de transporte, aparecerá incrustado un escenario interactivo, como puede verse en la figura 8, en el que se representarán dos entidades y el tráfico cursado entre ambas para poder ofrecer diferentes servicios. El alumno podrá:

- Lanzar primitivas sobre las entidades del nivel para observar los cambios de estado que se producen.

- O, analizar el tráfico creado por el lanzamiento de dichas primitivas interceptando los paquetes que viajan de una entidad a otra para visualizar su contenido.

4.1.2 Simulación de un modem Hayes

Cuando en el libro de texto se trate el tema de configuración y uso de MODEMs utilizando comandos HAYES, se ofrecerá una ventana emergente del texto, como recoge la figura 9, en la que el alumno podrá comprobar el funcionamiento de los distintos comandos que se le van presentando. Se le ofrecerá un área de texto donde componer los comandos HAYES y cuando así lo solicite la demo simulará el funcionamiento de un modem devolviendo las respuestas que ante estos comandos se obtendrían en una experiencia real.

Así el proceso de aprendizaje no quedará restringido a los ejemplos concretos que aparezcan en el libro, pudiendo el usuario probar en el simulador cuantos casos particulares desee.

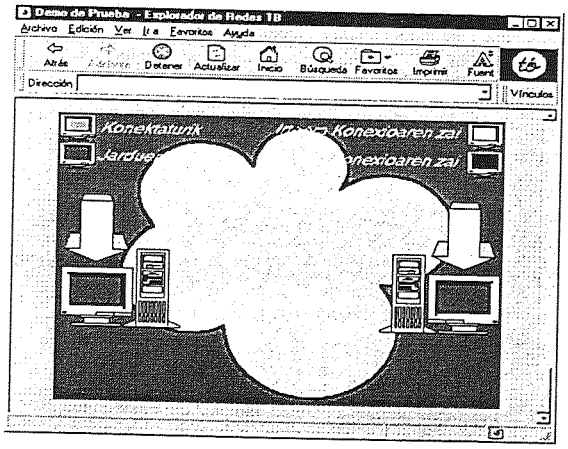


Figura 8: Demo interactiva del nivel transporte OSI

4.2 Vídeo de demostración

Como ejemplo de vídeo se ha grabado una secuencia en la que se muestra cómo configurar dos modems de un rack en línea dedicada. En el libro creado se ofrece la posibilidad de utilizar las diferentes opciones de visualización expuestas en el punto 3.2.

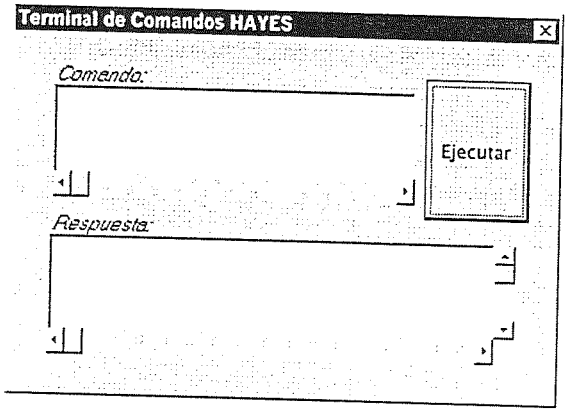


Figura 10: Pantalla para simulación de un modem Hayes

4.3 Herramientas de gestión

Las herramientas desarrolladas permiten facilitar la utilización del servicio de libros electrónicos desde el punto de vista del responsable del mismo, así como controlar el uso que se hace del laboratorio. Estas aplicaciones presentan ciertas características que las diferencian de las aplicaciones clásicas ofrecidas desde un web:

- Aumento del grado de interactividad debido a la conexión directa a la base de datos y abandono de la conexión vía HTTP.
- Los contenidos intercambiables con la aplicación ofrecida a través del servidor Web se amplían en cuanto a formato. Dejan de ser páginas HTML para poder usarse, gráficas, información animada, etc.
- Se abandona por completo el modelo basado CGI que obligaba a enviar datos en claro por la red. Se pueden intercambiar datos de forma segura implementando sobre la conexión directa hasta el servidor (socket) mecanismos de encriptación (clave pública - clave privada). En este caso en vez del modelo de la figura 6 será necesaria otra arquitectura. El applet del cliente no se conectará directamente contra la base de datos sino que lo hará contra una parte de la aplicación que permanecerá en el servidor (3), y es esta aplicación la que trabajará contra la base de datos ((4) y (5)), para devolver los datos requeridos al cliente (6).

El esquema de funcionamiento queda recogido en la figura 10.

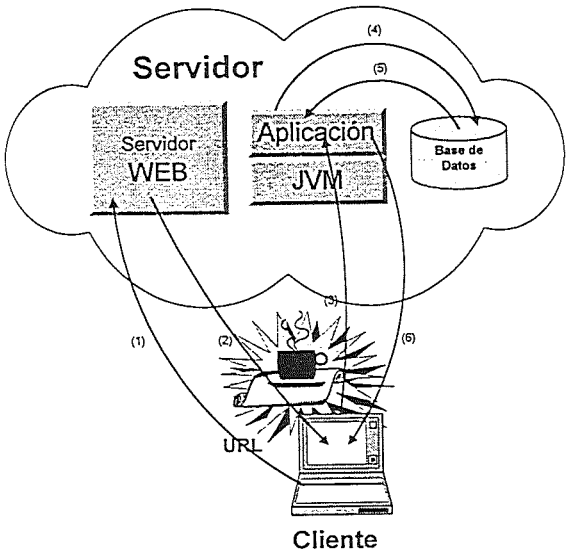


Figura 10: Arquitectura general de las aplicaciones de gestión

4.3.1 Control del laboratorio

Para controlar el uso del laboratorio de telemática y hacer un seguimiento del trabajo realizado por los alumnos en él, se ha desarrollado una aplicación dividida en 3 módulos:

- **El módulo a utilizar por los alumnos para crear la información.** Mediante este módulo al entrar en el laboratorio y encender uno de los equipos, se pedirán al usuario sus datos personales e información sobre el trabajo que va a desarrollar. Se utilizará para ello un applet que presente un formulario en una página Web.
- **El módulo de base de datos.** En el se almacenará toda la información utilizada por la aplicación. Se ha empleado la base de datos mSQL [7].
- **El módulo a utilizar por los profesores para analizar la información.** Un applet insertado en una página especial de los libros de texto será el que ofrezca una puerta para llegar hasta este módulo. Desde allí, se podrá analizar la información almacenada en la base de datos y pedir estadísticas como:
  - Utilización a lo largo de los meses del curso..
  - Porcentajes de las diferentes asignaturas en el uso total.
  - Uso soportado por los diferentes equipos,...

Además, si ocurriera una avería en algún equipo, cabría la oportunidad de realizar una consulta para obtener la lista de alumnos que utilizaron el equipo en cuestión en el intervalo de tiempo especificado.

Para este caso concreto la arquitectura de la figura 10 queda reflejada en la figura 11.

Tanto profesores como alumnos accederán al módulo que les corresponde en la herramienta de gestión desde una página del libro de texto electrónico ((1) y (2)). Después para realizar operaciones sobre la base de datos establecerán una conexión directa a nivel TCP contra el servidor mSQL. Los alumnos para insertar la información sobre el uso (4) y los profesores para analizar la información almacenada (5).

4.3.2 Gestión del servicio de tests de autoestudio

En la aplicación correspondiente a la gestión de test de autoestudio que se ofrecen junto con cada tema del libro, existen 3 módulos:

- **Módulo de base de datos**, es decir, la base de datos propiamente dicha con las tablas que almacenan los diferentes tests y sus respuestas.
- **Módulo de gestión**, que permite la inserción de tests y su mantenimiento (actualización).
- **Módulo de operación**, empleado por los alumnos para hacer los tests y obtener los resultados.

El módulo de base de datos se ha implementado utilizando la base de datos mSQL mediante 3 tablas, según el esquema relacional que aparece en la figura 13.

El módulo de gestión de los tests facilita el uso de la base de datos que da soporte al servicio:

- Ayudando a crear nuevos tests.
- Ofreciendo herramientas para modificar los existentes

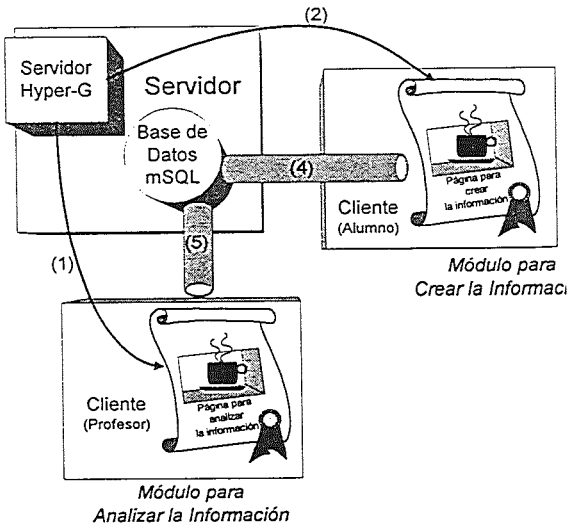


Figura 11: Visión general de la aplicación de control del laboratorio

- Permitiendo la eliminación de los que se desee descartar.

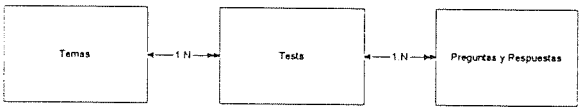


Figura 13: Modelo Entidad - Relación

La arquitectura de este módulo de gestión se muestra en la figura 13, y como puede verse está formada por tres elementos, uno para la comunicación con la base de datos, otro para la gestión de los test, y por último, otro de interfaz con el usuario.

El Módulo para la comunicación con la Base de Datos es el encargado de crear y mantener las relaciones con la base de datos del servidor. Construirá un enlace a nivel TCP y utilizando sobre él el protocolo propio de la base de datos, hará llegar hasta el servidor de base de datos las peticiones que realicen el resto de los módulos.

El Módulo para la gestión de los tests utiliza los servicios ofrecidos por el módulo de comunicación con la base de datos, y ofrece las siguientes funcionalidades al módulo de interacción con el usuario:

- Recoge toda la información sobre un test concreto distribuida en las diferentes tablas de la base de datos y presentar dicho test como un elemento con entidad propia.
- Ofrece posibilidades de edición para el test que se ofrece como objeto global.
- Posibilita la creación de un nuevo test.
- Posibilidad de volcar de nuevo a las distintas tablas de la base de datos la información del test manejado como entidad global.

Finalmente el Módulo de interacción con el usuario es el encargado de ofrecer y gestionar un interfaz de usuario sencillo e intuitivo detrás de la cual se oculte la complejidad de la herramienta de gestión, como se observa en la figura 14.

Figura 13: Módulos funcionales dela aplicación de gestión de tests.

Realiza esta tarea:

- Usando directamente las funcionalidades ofrecidas por el módulo de comunicación con la base de datos, presenta al usuario:
  - La lista de todos los temas sobre los que existen tests en la base de datos.
  - La lista de los tests que existen para un tema concreto.
- Y valiéndose del módulo de gestión de los tests ofrece:
  - La posibilidad de eliminar un test completo de la base de datos (haciendo las modificaciones pertinentes en cada una de las tablas).
  - La posibilidad de crear un nuevo test.
  - La funcionalidad de permitir al usuario visualizar y modificar toda la información de la que se compone un test.

El modo de operación es el siguiente:

- En la lista desplegable *Gaia* aparece una relación de todos los temas sobre los que existe algún test en la base de datos y al seleccionar alguno de sus elementos se actualizará la lista inferior mostrando los nombres de todos los tests existentes sobre el tema seleccionado.
- Al pulsar el botón *Ezabatu*, se eliminará de la base de datos toda la información sobre el test seleccionado en la lista de tests.
- Pulsar el botón *Aldatu* permitirá la edición de las preguntas y respuestas que forman el test seleccionado.

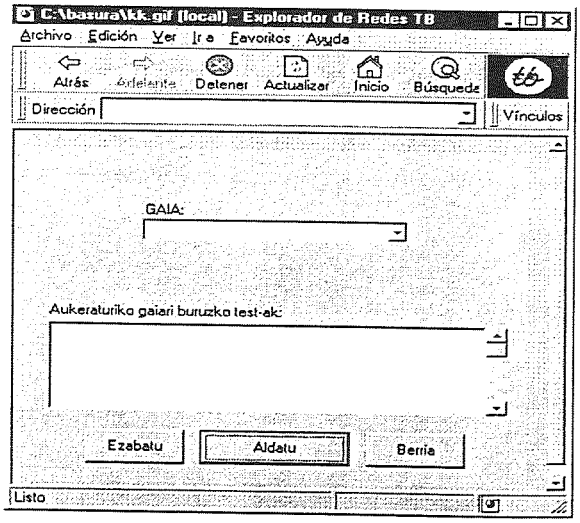


Figura 14: Pantalla para la gestión de tests

- Por otro lado, la pulsación del botón ***Berria***, creará un nuevo test sobre el tema seleccionado en la lista desplegable de temas.

En los dos últimos casos, es decir: al crear un nuevo test o al editar uno ya existente, aparecerá una ventana emergente del applet como la que se muestra en la figura 15.

En esta ventana se puede escribir o modificar el enunciado de la pregunta, sus cuatro respuestas, cuál de ellas es la correcta o incluso el nombre del propio test. Para poder navegar hacia delante o hacia atrás de una pregunta del test a otra se ofrecen dos botones con forma de flecha y una barra de deslizamiento para poder hacerlo como es habitual en las aplicaciones tipo Windows. Una vez terminado el trabajo de creación/edición será suficiente con pulsar el botón de finalizar para que todos los cambios realizados en el test tengan su reflejo en las distintas tablas de la base de datos.

Por último, para la posibilitar la realización de los tests al alumno se utilizará el módulo de operación, que es completamente paralelo al módulo de gestión excepto en que los campos de la figura 15 no serán editables y tras pulsar el botón BUKATU se procede a la corrección de las preguntas y a la visualización del resultado obtenido por el alumno.

5. Análisis de rendimiento y prestaciones

Un aspecto muy importante a considerar para evaluar la viabilidad de este tipo de libros es cuantificar por un lado los tiempos de respuesta y por otro la inversión necesaria para hacerlos operativos.

Para medir el rendimiento del sistema los tiempos que hay que considerar son los siguientes:

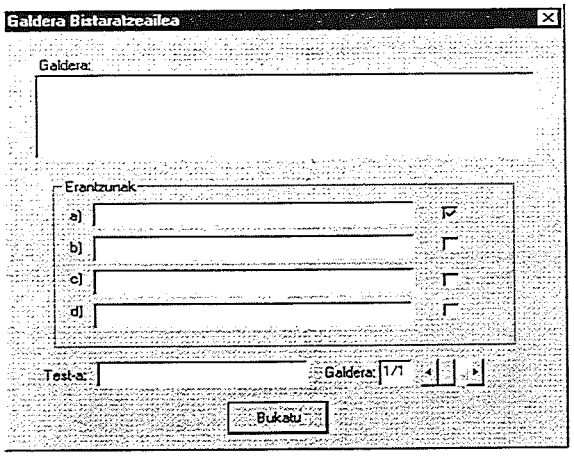


Figura 15: Pantalla para la creación/edición de preguntas de test.

- Tiempos en la carga de los applets:
  - Primero se trae la página Web en la que está insertado el applet (primera conexión HTTP)
  - Cuando se encuentra en el navegador la marca <APPLET> se trae en una segunda conexión HTTP el archivo .class indicado en el parámetro CODE de la marca.
  - Si el applet consta de varias clases, habrá que traer todos los ficheros .class desde el servidor (1 nueva conexión HTTP por cada uno).

- Cuando el applet está en el cliente se ejecuta en esta máquina, y una vez puesto en marcha hay que considerar:
  - Velocidad de ejecución de contenidos JAVA del navegador.
  - Si es necesario utilizar contenidos multimedia (sonidos, imágenes para animaciones, etc.), habrá que traerlos en tiempo de ejecución desde el servidor.
  - Interacciones con la base de datos. Tiempos de conexión, consultas, etc.

En mediciones realizadas en el laboratorio, utilizando una red Ethernet a 10 Mbps, los tiempos medios de carga de los applets de las aplicaciones descritas en el artículo es de 8 segundos. Realizando el acceso por red telefónica utilizando un modem de 28.800 bps y conexión PPP el valor medio se cuadruplica (33 segundos). En ambos casos se utilizó como cliente un PC Pentium 120 Mhz con 16 Mb de RAM, con Windows 95 y como navegador Netscape Navigator Gold 3.01.

Para medir la velocidad de ejecución de contenidos Java del navegador se ha utilizado el test ofrecido por la revista PC Magazine [8], con el que se han obtenido los resultados que se reflejan en la tabla 1, y de forma gráfica en la figura 16. Los valores reflejan la puntuación obtenida por cada navegador en la ejecución de cada tarea (mejor navegador cuanto más alta sea la puntuación) excepto en el último apartado que mide tiempos (mejor navegador cuanto menor tiempo).

De las pruebas realizadas se deduce que el Netscape Communicator 4.0 es el navegador que ofrece mejores prestaciones en la ejecución de applets JAVA.

En cuanto a la inversión en infraestructura para implementar este servicio, tenemos por un lado el coste de dos PCs (servidor y cliente) y el del equipamiento para generar el material multimedia

(cámara de vídeo, escáner, tarjetas de audio y vídeo), cuyo precio conjunto ronda el de un PC. El software necesario, exceptuando el sistema operativo del cliente (Windows 95), es gratuito para universidades (Hyper-G, linux, JDK, navegadores).

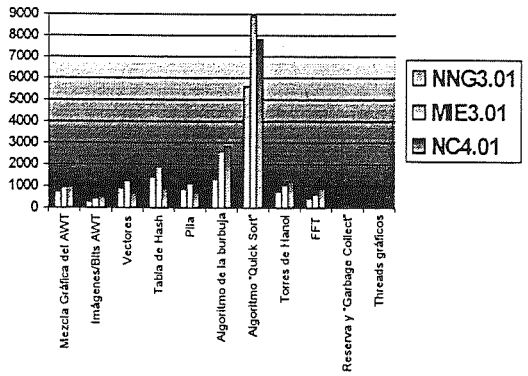
	NNG3.01	MIE3.01	NC4.01
Mezcla Gráfica del AWT	745	914	985
Imágenes/Blts AWT	278	438	515
Vectores	880	1237	688
Tabla de Hash	1405	1880	857
Pila	838	1122	716
Algoritmo de la burbuja	1313	2602	2817
Algoritmo "Quick Sort"	5604	8896	7822
Torres de Hanoi	760	1044	1187
FFT	441	610	8933
Reserva y "Garbage Collect"	70	103	86
Threads gráficos	56,03	41,74	24,23

Tabla 1: Comparación entre navegadores en ejecución de aplicaciones JAVA

6. Conclusiones

El desarrollo de materiales docentes con JAVA ofrece grandes posibilidades debido a la facilidad para crear y ofrecer información multimedia en la que se permite la interacción con el lector.

Para programadores familiarizados con C++ y la programación orientada a objetos la migración a esta nueva tecnología es casi inmediata, siendo además gratuito para universidades todo el software necesario tanto para desarrollo como para la explotación.



NNG3.01: Netscape Navigator Gold v3.01
MIE3.01: Microsoft Internet Explorer v3.01
NC4.01: Netscape Communicator v4.01

Figura 16: Cuadro comparativo entre navegadores en la ejecución de aplicaciones Java.

Finalmente hay que indicar es posible implementar estos libros con una infraestructura que se encuentra disponible en cualquier laboratorio universitario, con un grado de interactividad aceptable, empleando PCs conectados a redes LAN Ethernet

En este artículo se han presentado algunas de las muchas posibilidades que pueden realizarse con esta tecnología, tratando de cubrir con el prototipo desarrollado aquellos aspectos más relevantes desde el punto de vista de su aplicación a la enseñanza. Aún así no cabe duda de que queda mucho por hacer.

Referencias

[1] Goiricelaya, I., Uncilla, JJ., Jacob, E. "Using World Wide Web Servers as Electronic textbooks". *Proceedings of the EAEEIE '96*, 165-168 (1996).

[2] Bush, V "As we may think". *Atlantic Monthly* (1945).

[3] Maurer, H. "Hyper-G is now HyperWave. The Next Generation Web Solution.". *Addison-Wesley Publishing Company* (1996).

[4] Gosling, J, Joy, B, Steele, G. "The Java Language Specification". *Addison-Wesley Publishing Company* (1996).

[5] Sun Microsystems Inc, Silicon Graphics Inc,. Intel Corporation "Java Media Framework". *Sun Microsystems Inc* (1997).

[6] Rowe, J. "Building Internet database servers with CGI". *New Riders* (1996).

[7] Hughes, D. "mSQL database engine" ftp://Bond.edu.au/pub/Minerva/msql (1995).

[8] Seltzer, L. "Java Virtual Machines Perfomance Tests". *PC Magazine*, **16**, 10,111-115 (1997).

INDICE DE AUTORES

INDICE DE AUTORES