

# Capítulo 4

## Ubicación de recursos en redes ATM. Aplicación a fuentes de tráfico MPEG VBR

En este capítulo se estudia la transmisión de la información generada por las fuentes de vídeo MPEG VBR sobre redes que emplean ATM. Se desarrolla un análisis de la multiplexación estadística a través de simulación para una probabilidad de pérdida aceptable en este tipo de servicios y distintas restricciones de retardo máximo. Este análisis permite cuantificar la ganancia de multiplexación estadística para distintas combinaciones de fuentes de vídeo compartiendo los mismos recursos de red. A partir del estudio del tráfico agregado se propone un nuevo modelo ARMA capaz de capturar el comportamiento del tráfico entrante en un multiplexor estadístico. Para la síntesis de este modelo se propone un nuevo procedimiento basado en la estacionalidad presentada por la función de autocovarianza de la serie temporal del *número de celdas generadas por intervalo de slice*. Finalmente, se plantea un esquema para la transmisión de vídeo MPEG VBR sobre redes ATM que incorpora un dispositivo cuyas funciones son monitorizar el cumplimiento de contrato establecido con la red, detectar el nivel de actividad de la secuencia de vídeo digital y regular la tasa binaria generada por el codificador en situaciones de congestión.

### 4.1. Introducción

El principal desafío en el diseño de las redes ATM es la capacidad de soportar simultáneamente una amplia variedad de servicios con requerimientos y comportamientos muy dispares. Las redes ATM deben proporcionar diferentes grados de servicio, dependiendo del tipo de aplicación soportado a través de una adecuada técnica de ubicación de recursos.

El tráfico de vídeo es una clase de tráfico de banda ancha con características y requerimientos para la red muy diferentes a los del tráfico de datos. El vídeo requiere mucho más ancho de banda que los otros servicios, y por ello puede ser que fácilmente domine el tráfico de la red incluso con un número pequeño de fuentes. El tráfico de vídeo, aunque este no sea interactivo, está sumamente condicionado por la diferencia de *retardos* con los que llegan las celdas al receptor (CDV), salvo que se utilicen grandes buffers en recepción para absorber estas variaciones. Estas variaciones en el retardo son debidas a que las celdas, al atravesar la red, son almacenadas en los buffers intermedios en el caso de que éstos no estén vacíos. Para servicios interactivos, el retardo total también está limitado por el sistema visual humano a 100 ms [Gar93]. De forma que las celdas que sobrepasan este límite puede que no sean utilizadas por el receptor.

Los errores introducidos en las transmisiones de servicios con importantes requerimientos temporales no pueden ser solventados mediante técnicas de retransmisión. No obstante, para servicios de vídeo, el decodificador y el sistema visual no precisan tasas de error nulas. Una tasa de error en la región de  $10^{-3}$  a  $10^{-4}$  (y no de  $10^{-7}$  a  $10^{-9}$  como se consideraba primeramente) se tiene en cuenta actualmente como el umbral para la utilización útil de un codificador MPEG con una apropiada cancelación de errores y un buen soporte por parte del nivel de adaptación ATM (AAL) [Rei93]. Incluso, la tasa tolerable de pérdida de celdas puede incrementarse un orden de magnitud con las técnicas de transmisión binivel. Estas propiedades definen un problema de diseño para los algoritmos de control de tráfico y congestión de la red [UNI3.0].

Puesto que la ganancia de multiplexación depende del número de fuentes que compartan los mismos recursos de la red, la mejor manera de dimensionar las redes ATM y controlar su tráfico es tener una asignación de ancho de banda que cambia según el punto de la red. Es posible que cerca del centro de la red haya tantas fuentes compartiendo los enlaces, que el multiplexado estadístico reduzca sustancialmente el ancho de banda necesario. Por el contrario, para un grado de multiplexado bajo (p.e., cerca de los extremos de la red) puede que se deba de asignar para las fuentes de vídeo un ancho de banda cerca de la tasa de pico con tal de conseguir los requerimientos de QoS. Sin embargo, esto no presenta ningún tipo de problema para el vídeo VBR. El coste del ancho de banda en el bucle de abonado es menor que dentro de la red. Este ancho de banda en el bucle de abonado es mucho mayor que el que se necesita para un servicio, con lo que un incremento en el ancho de banda necesario es relativamente barato. Dentro de la red, a pesar de que hay economías de escala y por tanto el coste del ancho de banda por kilómetro es pequeño, al ser las distancias grandes el coste total aumenta. Pero, el número de conexiones es elevado, por lo que, se produce un elevado grado de multiplexación estadística. De esta manera se observa que donde el número de conexiones es mayor el coste es superior y es posible reducir la asignación de ancho de banda.

La compartición de los recursos de la red por parte de diferentes servicios se realiza de forma más eficiente cuando se emplea la multiplexación estadística. El uso de esta técnica motiva el desarrollo de modelos de tráfico, para las distintos tipo de fuentes, con los que se puedan evaluar las ventajas de la multiplexación estadística y permitan establecer criterios para la ubicación de recursos. Cada conexión queda caracterizada por un conjunto de parámetros descriptores del tráfico que permiten determinar los recursos necesarios y el grado de servicio alcanzable cuando se multiplexan estadísticamente un conjunto de conexiones. Una vez caracterizado el comportamiento de las distintas fuentes de tráfico, es necesario disponer de herramientas de análisis para evaluar los parámetros que fijarán la calidad ofrecida por la red a los diferentes servicios. Entre dichos parámetros estarán las probabilidades de pérdida o de bloqueo, el retardo medio (CDT), la variación entre llegadas (CDVT), etc.

Las técnicas de ubicación de recursos se apoyan en la caracterización de las fuentes individuales de tráfico y en la construcción de modelos de tráfico agregado homogéneo o heterogéneo. A través de los modelos agregados de tráfico se pueden evaluar los recursos necesarios del tráfico conjunto, permitiendo establecer criterios de control de admisión.

La naturaleza de las distintas fuentes de tráfico soportadas por las redes ATM obliga a establecer dos niveles de control de admisión. Para aquellas fuentes cuyo comportamiento es conocido y no presentan variaciones elevadas y sostenidas (p.e., los servicios CBR o los servicios VBR de videoconferencia y de videotelefonía) se puede emplear un control de admisión en el establecimiento de la conexión. Sin embargo, los servicios cuyo comportamiento es desconocido a priori o presenta amplias y sostenidas fluctuaciones en el transcurso de la conexión (p.e., los servicios VBR de vídeo bajo demanda y distribución de vídeo o los servicios de tráfico de datos ABR y VBR) es necesario aplicar un control de admisión a nivel de ráfaga. El control de admisión a nivel de ráfaga permite una ubicación de recursos dinámica durante el transcurso de la conexión con lo que se aumenta sustancialmente el grado de explotación de los recursos de la red manteniendo la misma calidad de servicio. La reserva dinámica de recursos durante el tiempo de conexión provoca un aumento de la complejidad en la gestión de la red y obliga a emplear técnicas de asignación de recursos simples para que el tiempo de respuesta de la red sea muy reducido.

El control de admisión a nivel de ráfaga o la renegociación de los recursos necesarios durante el transcurso de la conexión se ha planteado a través de dos protocolos de reserva rápida (Fast Reservation Protocol, FRP) que dependen de la tolerancia del servicio a la demora de la respuesta de la red en el proceso de renegociación [BoyTra92] [PanZar93]. Para los servicios que no admiten un excesivo retardo o para los servicios cuya duración de ráfaga es reducida respecto al tiempo de respuesta de la red, se ha planteado un protocolo con transmisión inmediata (FRP immediate transmission,

FRP/IT). Este protocolo se basa en el intento de transmitir, sin previa asignación, un bloque de información, el cual se precede con una celda de petición de reserva y finaliza con una celda de liberación. Los servicios donde se puede aplicar este protocolo se denominan fuentes Negotiated Burstwise Variable Bit Rate (NBVBR).

Para los servicios cuya actividad puede esperar a la aceptación o denegación de la nueva ubicación de recursos por parte de la red, se ha propuesto el protocolo FRP con transmisión demorada (FRP delayed transmission, FRP/DT). En este caso, se mantiene el comportamiento de la fuente de tráfico hasta que la red informa al servicio de la disponibilidad de los recursos demandados. Los servicios que pueden soportar este protocolo se denominan fuentes Negotiated Stepwise VBR (NSVBR). La estandarización de estos mecanismos de reserva rápida se está realizando actualmente por la ITU-T con la denominación ATM Block Transfer (ABT) con las variaciones ABT/DT y ABT/IT.

## 4.2. Multiplexación estadística de tráfico de vídeo MPEG VBR

Para investigar la viabilidad y la eficiencia de la transmisión de vídeo VBR en los futuros sistemas ATM, se debe determinar qué recursos de red son necesarios para el transporte de  $N$  señales de vídeo multiplexadas estadísticamente, ofreciendo a cada una de ellas la calidad de servicio (QoS) contratada. Esto permite hacer una comparación entre los diferentes tipos de servicios (CBR, VBR, o diferentes tipos de VBR, etc.), ya que cualquier comparación entre servicios debe incluir un intento de averiguar la asignación de recursos necesarios para cada uno de los servicios.

En esta sección se examina mediante simulaciones la relación entre la asignación de recursos de red y la calidad de servicio obtenida para un tráfico agregado de vídeo VBR MPEG. Las simulaciones realizadas miden el comportamiento del tráfico agregado de  $N$  fuentes de vídeo VBR MPEG a la entrada de un multiplexor modelado por una cola con un buffer de tamaño  $Q$  y con un tiempo de servicio determinístico  $1/C$ , donde  $C$  es la capacidad de canal, según se muestra en la figura 4.2.1. Cuando el buffer del multiplexor está lleno pueden producirse pérdidas de celdas, ya que las celdas de entrada que llegan en esta situación se descartan. Cabe destacar la relación existente entre el retardo máximo introducido por el buffer ( $T_{\max}$ ) que es uno de los requerimientos de la calidad de servicio del vídeo, y el tamaño del buffer ( $Q$ ) en celdas:

$$T_{\max} = \frac{Q}{C} \quad (4.2.1)$$

Las simulaciones realizadas evalúan el grado de servicio ofrecido, en términos de retardo y probabilidad de pérdida, para una ubicación de recursos determinada por la

dimensión del buffer ( $Q$ ) y la capacidad asignada ( $C$ ). El estudio de la influencia de estos dos parámetros en la calidad de servicio se ha llevado a cabo, habitualmente, a través del análisis de la probabilidad de pérdida en función del valor adoptado por uno de ellos cuando el otro se mantiene constante. En [Ens94] se calculan las gráficas de la evolución de la probabilidad de pérdida en un multiplexor para diferentes tamaños de buffer con una capacidad y una carga del canal determinada. En [Rei93] se presentan los resultados de simulaciones que calculan la probabilidad de pérdida en un multiplexor para diferentes retardos máximos introducidos por el buffer, con una capacidad del canal fija y un factor de utilización ( $\rho$ ) de dicho canal que varía entre 0'8 y 0'92, donde se considera el nivel de utilización del canal como el cociente entre la carga del canal, es decir, la media del tráfico total, y la capacidad del mismo. Este cociente es igual a  $1/C$ , donde  $C$  es la capacidad del canal normalizada respecto a la media del tráfico total.

$$\rho = \frac{\text{carga del canal}}{\text{capacidad del canal}} = \frac{1}{C(\text{normalizada})} \quad (4.2.2)$$

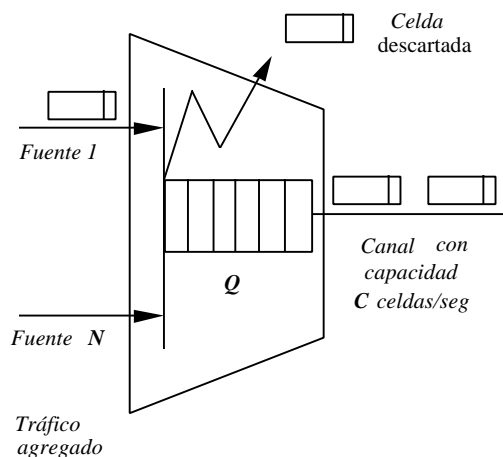
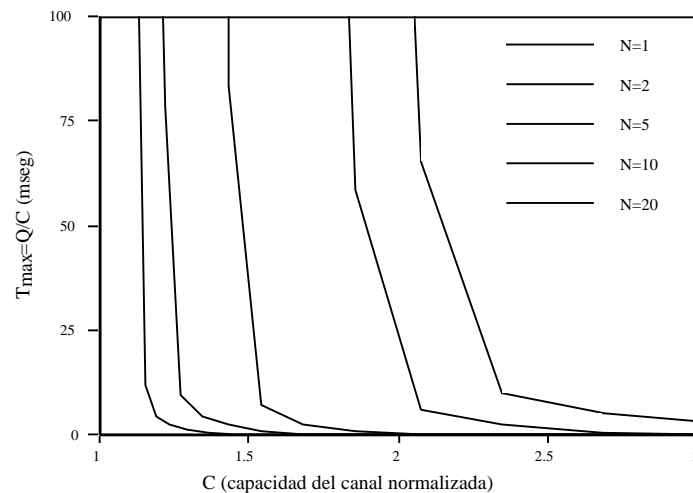


Figura 4.2.1. Esquema de multiplexor estadístico

En el estudio realizado, se fija un objetivo de probabilidad de pérdida y se calculan los recursos ( $Q, C$ ) necesarios. El objetivo es entender cómo los recursos están relacionados entre ellos. Para llevar a cabo este análisis se han empleado diferentes fragmentos de la codificación de la secuencia "Live in Central Park" cuya duración es de 41695 cuadros. Los parámetros de codificación MPEG VBR fijados han sido  $Q=6$ ,  $M=3$  y  $N=9$ . En este análisis se han contemplado las tasas binarias generadas en el intervalo de duración de un slice (1'8 ms). Cabe destacar que no se ha aplicado ningún tipo de conformación sobre el tráfico generado por el codificador, por lo que se mantienen las correlaciones temporales intrínsecas al algoritmo de codificación MPEG,

las cuales, provocan una variación periódica en las llegadas al multiplexor. Esta variación periódica reduce las ventajas de la multiplexación estadística. Esta consideración pretende establecer una cota inferior sobre los resultados obtenidos en la multiplexación estadística y, a su vez, evaluar la conformación de tráfico introducida en la propia red, debida al almacenamiento sufrido por las celdas en las colas de espera.

La Figura 4.2.2 muestra los resultados obtenidos en las simulaciones realizadas. En esta gráfica se visualiza la relación entre la capacidad del canal y el retardo máximo para satisfacer una probabilidad de pérdida de  $10^{-4}$ , cuando se multiplexan diferentes número de fuentes ( $N$ ). La capacidad del canal ( $C$ ) se presenta en el eje de abcisas, normalizada respecto a la media del tráfico total. El retardo máximo ( $T_{max}$ ) introducido por el buffer se especifica a través del eje de ordenadas en unidades de milisegundos. El retardo máximo del buffer, o tamaño del buffer en milisegundos, corresponde al *jitter* máximo introducido por el sistema de multiplexación, y forma parte de la calidad de servicio ofrecida. Este retardo es una medida con más información que  $Q$ , ya que el retardo condiciona más el diseño que el coste de los buffers, y por ello, ésta es la variable utilizada en este análisis. Los márgenes tolerados de retardo están en el orden de 10-50 ms [Rei93]. Para la construcción de las curvas se ha empleado un algoritmo de punto medio para el cálculo del retardo máximo introducido por el buffer para un valor de probabilidad de pérdida y capacidad del canal especificados.



*Figura 4.2.2 Relación entre el tamaño del buffer ( $Q$ ) y la capacidad normalizada ( $C$ ) asignada al un número ( $N$ ) de fuentes de tráfico de vídeo MPEG VBR para una probabilidad de pérdida de  $10^{-4}$*

Las curvas de la Figura 4.2.2 muestran un fuerte cambio de tendencia en su pendiente. Esta transición representa un punto natural de operación del sistema, ya que en los otros puntos, uno de los dos parámetros es muy sensible a ligeros cambios en el otro. Sin embargo, la elección última dependerá del coste relativo entre los recursos

(capacidad del canal y tamaño de los buffers), como también del tipo de servicio (interactivo o no) y de los objetivos de diseño. Obsérvese que la curva correspondiente a una fuente individual ( $N=1$ ) se corresponde con las posibles asignaciones de los parámetros descriptores de tráfico correspondientes al mecanismo de policía GCRA(I,L). En este caso el valor de  $Q$  se corresponde con el tamaño del bucket y  $C$  con la tasa de generación de créditos, normalizada respecto a la tasa media de la fuente, para una probabilidad de marcación o descarte de  $10^{-4}$ .

Los resultados muestran que, con una asignación de ancho de banda alrededor de 1'2 veces superior a la media del tráfico agregado, se pueden alcanzar los niveles de probabilidad de pérdida deseados y unos valores de retardo entre 10 y 50 ms para un número de 20 de fuentes multiplexadas. Esto supone que se pueden alcanzar unos niveles de utilización del canal en la región de 0'8 a 0'9.

### Evaluación de la ganancia de multiplexación estadística

Como una primera aproximación de la medida del comportamiento a ráfagas (*burstiness*) del vídeo agregado [NicAky92], en la figura 4.2.3 se muestra la relación entre la tasa de pico y la tasa media para diferente número de fuentes multiplexadas ( $N$ ). Considerando las tasas a partir del número de celdas generadas en un intervalo de slice, se puede observar el efecto de suavización que sufre el tráfico de vídeo de una fuente al combinarse con el de otras.

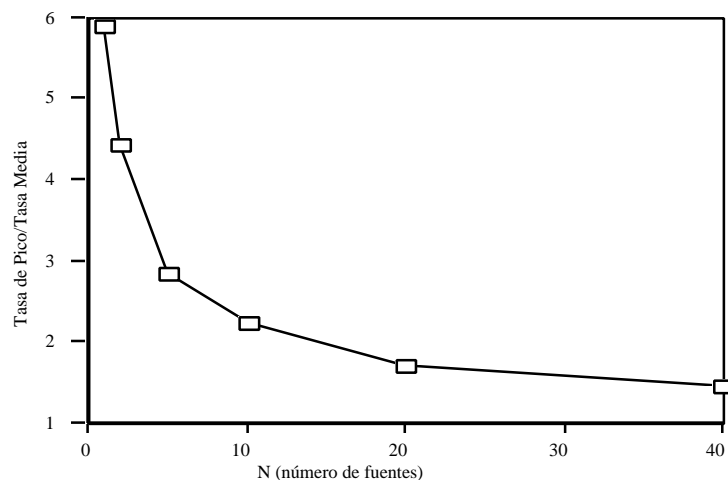


Figura 4.2.3. Relación entre tasa de pico y tasa media en intervalos de slice para diferente número de fuentes multiplexadas

Habitualmente se define la ganancia de multiplexación estadística (*Ganancia\_Mux*) como el cociente entre la suma de las tasas de pico de todas las fuentes y la capacidad del canal [Rei93]. Esta definición tiene el inconveniente de que depende de la capacidad del canal considerada y de la definición de la tasa de pico.

$$Ganancia\_Mux = \frac{\sum_{fuentes} tasa\_de\_pico}{capacidad\_canal} \quad (4.2.3)$$

En [Hey92] se define ganancia de multiplexación estadística, para fuentes de vídeo VBR, como el cociente entre el número de fuentes VBR y el número de fuentes de tasa binaria constante (CBR) que se pueden transmitir por un canal con una determinada capacidad, para una probabilidad de pérdida y retardo determinados para las fuentes VBR.

$$Ganancia\_Mux = \frac{\# fuentes\_VBR}{\# fuentes\_CBR} \quad (4.2.4)$$

Para determinar el número de fuentes CBR que se pueden transmitir por el canal se utiliza la tasa de pico, con lo que se está asumiendo la mejor calidad posible para el vídeo CBR. Por lo tanto, la ganancia calculada es mayor que la que realmente resultaría utilizando las tasas de las fuentes CBR que producen una calidad "semejante" a la obtenida en el modo de funcionamiento VBR. Esto se debe a que, para calidades de imagen comparables, las tasas de las fuentes CBR son menores que la tasa de pico en VBR y, por tanto, el número de fuentes CBR multiplexadas resulta menor.

Un problema fundamental en la comparación entre codificadores VBR y CBR es la definición de los niveles de calidad, ya que la degradación que se produce es diferente en cada caso. También es difícil comparar el ancho de banda consumido, ya que el ancho de banda VBR, evidentemente, no es constante. En este caso, cuando se habla del ancho de banda asignado por fuente VBR, se refiere al ancho de banda asignado al tráfico total dividido por el número de fuentes  $N$ . La figura 4.2.4 ilustra la idea de que una calidad "equivalente" en los codificadores VBR y CBR se obtendrá si la tasa binaria del codificador CBR se encuentra entre las tasa media y la tasa de pico del codificador VBR.

Otra manera de cuantificar la ganancia de multiplexado estadístico alcanzada para  $N$  fuentes, independientemente de la definición de la tasa de pico y la capacidad del canal, y que depende tan sólo del número de fuentes multiplexadas, es comparar el ancho de banda por fuente necesario cuando se transmite una fuente y cuando se multiplexan  $N$  fuentes [MatPag96]. Luego, podemos expresar la ganancia de multiplexación estadística como:



$$Ganancia\_Mux_N = \frac{C_1}{C_N} \quad (4.2.5)$$

donde  $C_k$ , con  $k \in \{1, 2, \dots, N\}$ , es el ancho de banda asignado a cada fuente cuando se multiplexan  $k$ . Esta capacidad se calcula dividiendo el ancho de banda asignado al tráfico total por el número de fuentes ( $k$ ).

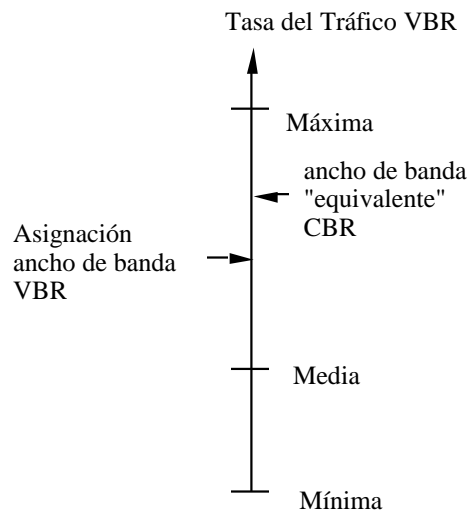


Figura 4.2.4. Comparación de la asignación de ancho de banda entre codificadores VBR y CBR para una calidad "equivalente".

La determinación del ancho de banda se realiza mediante las curvas  $(T_{max}, C)$  vistas anteriormente, fijando una calidad de servicio determinada (probabilidad de pérdida y retardo máximo del buffer). La Figura 4.2.5 muestra las ganancias de multiplexación alcanzadas con diferentes números de fuentes multiplexadas ( $N$ ) y definiendo el ancho de banda para un retardo máximo en el buffer de  $T_{max}=10$  y  $T_{max}=50$  ms. Se observa que es posible alcanzar ganancias del orden de 2 con un número suficientemente grande de fuentes ( $N=20$ ), para una probabilidad de pérdida de  $10^{-4}$ .

Se comprueba que la ganancia de multiplexación viene influenciada por el tamaño del buffer ( $T_{max}$ ) y el número de fuentes ( $N$ ). Esto se traduce en que servicios con la misma naturaleza de tráfico alcanzan mayor ganancia de multiplexación estadística cuanto menos restrictivo sea el retardo máximo de tránsito por la red, o de forma equivalente, necesitan una menor asignación de ancho de banda para la misma probabilidad de pérdida. En el límite, aumentando el buffer, el ancho de banda necesario tiende a la media a largo plazo de la fuente. Aplicando la ley de los grandes números se puede concluir igualmente que el límite inferior de la asignación de ancho de banda para el tráfico multiplexado es la tasa media total, es decir, la asignación de ancho de banda para una fuente está más próxima a su media cuanto mayor es el número de fuentes combinadas. No obstante, la tasa de pico establece la menor cantidad

que daría lugar a una ubicación de recursos con probabilidad de pérdida nula. Claramente, cualquier asignación por encima de la tasa de pico es innecesaria y derrochadora. Por lo tanto, los anchos de banda de pico y medio de las fuentes acotan el ancho de banda necesario en la red.

$$Ganancia\_Mux \leq \frac{tasa\_pico}{tasa\_media} \quad (4.2.6)$$

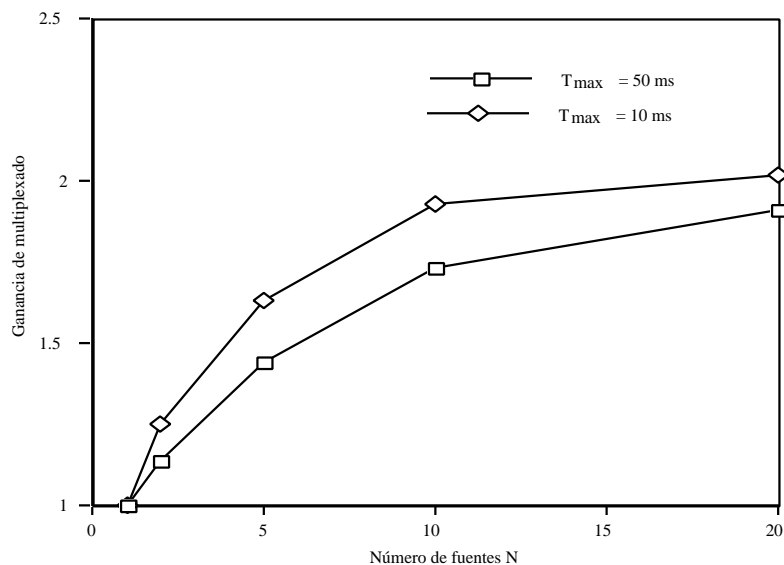


Figura 4.2.5 Ganancias de multiplexado alcanzadas para diferente número de fuentes multiplexadas con probabilidad de pérdida  $10^{-4}$  y retardos máximos 10 y 50 ms.

### 4.3. Caracterización del tráfico agregado de vídeo MPEG VBR

La aportación de nuevos modelos para la síntesis artificial de tráfico permite el desarrollo de simulaciones para el estudio de las prestaciones de los dispositivos de la red y de ésta misma. En particular, la caracterización del tráfico vídeo VBR es especialmente importante ya que este tráfico será el principal consumidor del ancho de banda de las futuras redes de banda ancha. Por lo tanto, es importante el modelado de fuentes de vídeo VBR para desarrollar estudios como el diseño de las redes ATM, el dimensionado de elementos de estas redes (conmutadores, multiplexores), la gestión del tráfico, la relación entre la asignación de recursos y la calidad de servicio obtenida, y otros temas relacionados.

El objetivo del modelado del tráfico es identificar y describir los procesos estocásticos asociados al tráfico soportado por las redes ATM. Esta caracterización

incluye tanto la descripción de fuentes individuales como la descripción de su tráfico agregado. Más aún, existe una necesidad de caracterizar agregaciones arbitrarias de fuentes heterogéneas, puesto que el mismo medio de transmisión será utilizado para multiplexar tráficos de datos, voz y vídeo.

Los modelos para tráfico de vídeo agregado deben capturar los efectos de la multiplexación estadística y deben predecir hasta qué punto se reduce su variabilidad. Estos modelos resultan útiles para dimensionar la red y para la gestión del tráfico. Dentro de los trabajos relacionados con este campo de investigación caben destacar los presentados en [LucNeu94], [NomFuj89], [MagAna88] y [HeyTab92].

En esta sección discutimos cuáles son las dificultades del modelado de la agregación del tráfico de varias fuentes de vídeo MPEG de tasa binaria variable (VBR). Para confeccionar un modelo de tráfico agregado se ha desarrollado previamente un análisis estadístico de su comportamiento. A partir de las dependencias temporales observadas en el tráfico agregado, se ha desarrollado un procedimiento genérico para la síntesis de tráfico artificial con estas características basado en procesos ARMA. Aunque esta técnica de modelado se ha aplicado sobre intervalos de duración de cuadro también sería válida para otros intervalos de duración como el de un slice, un SGoP o un GoP.

#### **4.3.1. Análisis estadístico del tráfico agregado MPEG VBR**

Para el análisis del comportamiento del tráfico de vídeo en un multiplexor ATM, y más en general en cualquier red ATM, se deben de construir modelos que capturen las propiedades esenciales del tráfico real para que sea posible generar una muestra artificial de tráfico equivalente que pueda ser utilizada en simulaciones. Realizar simulaciones con muestras de tráfico real tiene el inconveniente del gran coste computacional que supone la obtención del video codificado. Pero también, frecuentemente, se dispone de secuencias demasiado cortas con las que no se pueden calcular probabilidades de error suficientemente pequeñas.

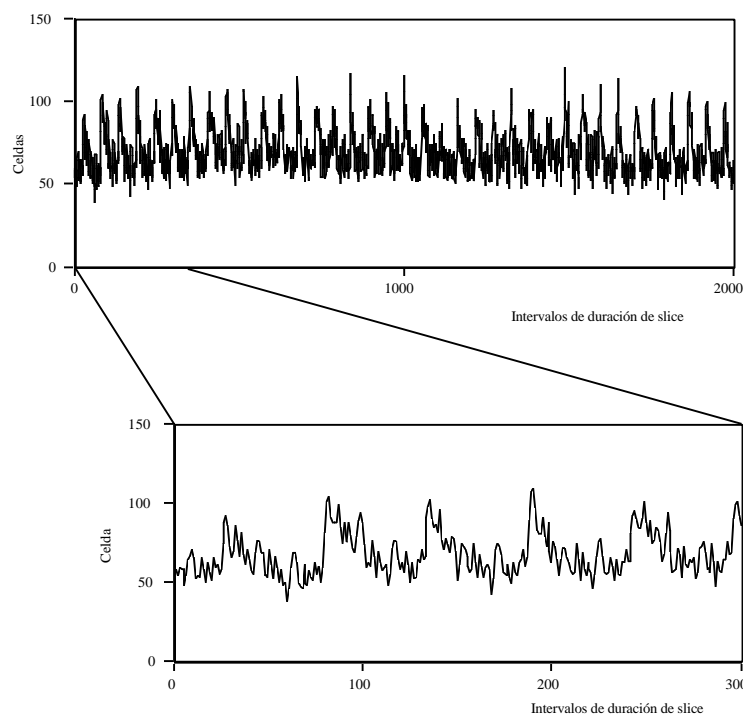
Además de los modelos de tráfico para simulaciones, también existen modelos analíticos como los modelos markovianos que permiten aplicar la teoría de colas para el cálculo de la probabilidad de pérdida en el multiplexor [MagAna88].

En este estudio del tráfico agregado se ha utilizado una superposición de diferentes fuentes aleatoriamente desfasadas, con un desfase que no se ha forzado a un número entero de tiempos de imagen como en estudios anteriores [Rei93], [Gar93], [LucNeu94]. Estas fuentes se han implementado mediante diferentes fragmentos de una secuencia de vídeo codificado obtenida a la salida de un codificador MPEG VBR con parámetros de codificación ( $Q=6$ ,  $M=3$ ,  $N=9$ ). Todas las fuentes, al llegar al final de la secuencia vuelven a empezar desde el principio, con lo que en la simulación se recorren

todos los 41695 cuadros que componen la secuencia muestra, obtenida de la grabación "Live in Central Park".

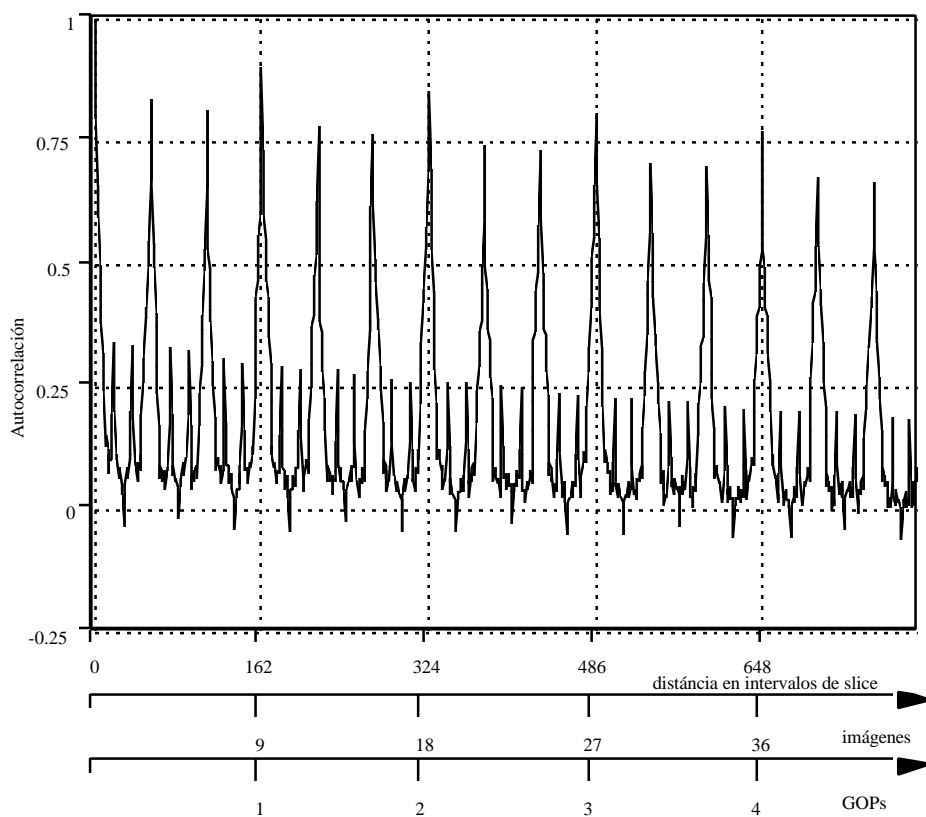
Para el análisis estadístico del tráfico se estudia la secuencia que contabiliza el número total de celdas generadas en intervalos de tiempo de duración igual al del slice. El estudio para intervalos menores no es necesario teniendo en cuenta el tamaño mínimo de los buffers de contención que se utilizarán en los enlaces capaces de transportar vídeo agregado.

La Figura 4.3.1 permite hacer un análisis del aspecto temporal del tráfico agregado. En esta figura se muestra la secuencia resultante de la agregación del tráfico de  $N$  fuentes, con  $N=10$ . Las unidades del tráfico medido están en celdas generadas por intervalo de duración de slice. Se observa una gran correlación para intervalos separados una distancia igual al tiempo correspondiente a 3 cuadros, o un SGoP (54 tiempos de slice). Esta alta correlación en el tráfico generado es debido al gran parecido en el tipo de codificación que tienen en el estándar MPEG dos slices separados un tiempo de SGoP.



*Figura 4.3.1. Serie temporal del número de celdas generadas por la agregación del tráfico de 10 fuentes de vídeo MPEG VBR por intervalo de duración de slice*

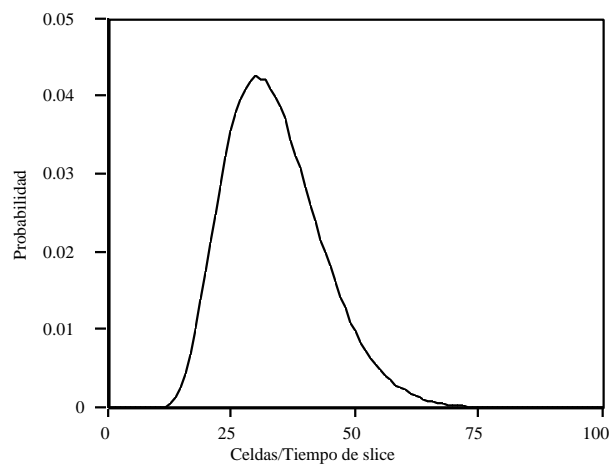
La variabilidad que se observa en el tráfico en un tiempo de SGoP sugiere la necesidad de buffers con tamaños capaces de, al menos, aceptar la variación del tráfico generado en este tiempo. El retardo que esos buffers introducen, como se observará, resulta inferior al introducido si la fuente promediara el tráfico generado durante ese intervalo. El retardo generado en tal caso sería igual a tres veces la duración de un cuadro (120 msec). Esta reducción del retardo se debe a la menor variabilidad del tráfico agregado respecto al tráfico simple. En la figura 4.3.1, en menor medida, también se observa un cierto comportamiento periódico de período un tiempo de GoP (162 tiempos de slice), este efecto se deriva de la estructuración de las imágenes en la codificación.



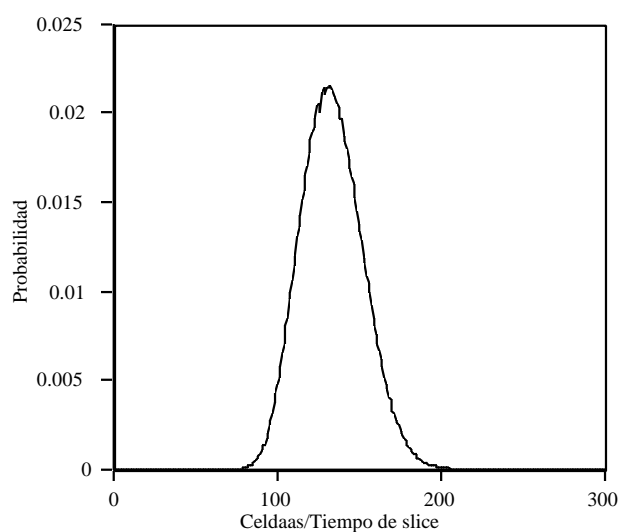
*Figura 4.3.2. Autocovarianza de la secuencia número de celdas en intervalos de slice para la agregación del tráfico de 10 fuentes MPEG VBR*

La observación de la autocorrelación mostrada en la Figura 4.3.2 de la secuencia estudiada permite hacer un análisis más detallado de la dependencia temporal que poseen las muestras de la secuencia de tráfico. En ella sigue apareciendo clara la correlación que sufre el tráfico separado un tiempo de SGoP. También se ve reproducida la estructura del GOP con los picos distanciados 162 tiempos de slice. Como se podía suponer a partir de la realización temporal, el tráfico se decorrela muy rápidamente dentro de un tiempo de imagen, entendiéndose por decorrelación el rápido decaimiento del índice de autocorrelación.

Las figuras 4.3.3 y 4.3.4 muestran las funciones de distribución de probabilidad de la variable que cuantifica el número de celdas generadas por intervalo de duración de slice, para un número de fuentes  $N$  igual a  $N=5$  y  $N=20$ . En ellas se puede ver la tendencia general: aumentando  $N$ , las distribuciones presentan un aspecto más parecido a una gaussiana y la probabilidad de los valores alejados de la media se reduce. No obstante, no es posible despreciar el decaimiento más lento que presenta la distribución para tasas de generación altas cuando se pretenden observar probabilidades de pérdida muy bajas ( $P_{\text{per}} < 10^{-4}$ ), ya que los intervalos con gran número de celdas influyen mucho en la probabilidad de pérdida, a pesar de ser poco probables [Gar93].



*Figura 4.3.3. Función de distribución de probabilidad de el número de celdas generadas por intervalo de duración de slice para un número de 5 fuentes*



*Figura 4.3.4. Función de distribución de probabilidad de el número de celdas generadas por intervalo de duración de slice para un número de 20 fuentes*

### 4.3.2. Modelado del tráfico de vídeo agregado MPEG VBR

En esta sección se propone un nuevo procedimiento, basado en un modelo autorregresivo de media móvil (ARMA), para el modelado del tráfico agregado de vídeo MPEG VBR. Este modelo intenta capturar el comportamiento del proceso estocástico  $Y[n]$  que contabiliza el número total de celdas generadas por diferentes fuentes de vídeo en un tiempo de cuadro. Para ello, se parte de las muestras  $\{y[0], y[1], \dots, y[N-1]\}$  de una realización de dicho proceso estocástico. El procedimiento presentado puede ser, igualmente, utilizado para la construcción de un modelo ARMA para cualquier proceso estocástico susceptible de ser modelado de esta manera. Para la validación del modelo se han contrastado los resultados obtenidos en la sección 4.2 con los del modelo sintetizado. Para llevar a cabo la comparación de resultados se ha empleado la técnica propuesta en [PanZar94] para la síntesis artificial de tráfico en intervalos de slice a partir de un modelo para intervalos de duración un tiempo de cuadro. Esta técnica se basa en que la tasa generada por el modelo es el valor medio de una variable aleatoria geométrica que determina la tasa generada para cada intervalo de duración un tiempo de slice. Así, cada tasa obtenida por el modelo se emplea en la síntesis de 18 intervalos de slice.

#### Estimación de parámetros del modelo ARMA

Para la selección de los órdenes y la forma del filtro ARMA debemos usar tanto los datos como cualquier información que dispongamos sobre la generación de estos datos. Observando tanto la serie temporal de la figura 4.3.1 como la función de autocorrelación en la figura 4.3.2, se comprueba la existencia de un cierto comportamiento periódico de período igual a la duración de un GoP (162 tiempos de slice). Esta periodicidad se deriva del tráfico de vídeo MPEG simple, dado que la separación entre imágenes  $I$  es un tiempo de GoP. Así, en general, el tráfico agregado mantiene la relación temporal presentada por el tráfico simple entre dos instantes separados un intervalo de duración de un GOP. Un modelo debe tener en consideración esta periodicidad, o *seasonality* [BoxJen76].

Para modelar esta relación temporal del tráfico agregado MPEG VBR, se utilizará la parte autorregresiva del modelo ARMA para la captura de los picos de la función de autocovarianza separados un intervalo de GoP y la parte MA para la síntesis de los valores del índice de autocorrelación que se encuentran entre los picos distanciados un tiempo de GoP. Esta última consideración es válida dado que la forma presentada por la autocovarianza es semejante entre cada par de picos separados por este intervalo.

Aplicando el método de mínimos cuadrados sobre las ecuaciones de Yule-Walker modificadas para la estimación de parámetros de un filtro AR se determina que el orden (p) de la componente autoregresiva es 1. La consideración de órdenes superiores queda descartada dado que el valor correspondiente de los coeficientes de orden superior a 1 es significativamente inferior al del primer coeficiente. Puesto que la estimación se ha realizado únicamente con los picos del índice de autocorrelación situados cada tiempo de GoP, la estacionalidad se traduce en el filtro autoregresivo en un orden 9 (p=9). A su vez la parte autoregresiva sólo dispondrá de un coeficiente no nulo, obviamente el mayor, que denominaremos a. Obsérvese que a es el parámetro de la atenuación de la recorrelación que ocurre al cabo de un intervalo de duración del GOP.

Para modelar los valores del índice de autocorrelación entre los diferentes picos sintetizados por la parte autoregresiva se utiliza la parte de media móvil (MA) del modelo ARMA. El orden del filtro MA (q) debe de ser menor que p, por construcción. La estructura resultante se presenta en la figura 4.3.5. Así, la colocación en cascada de los dos filtros, el autorregresivo y el de media móvil, produce el comportamiento periódico deseado de la autocorrelación. Esta consideración surge de la relación:

$$r_{ARMA}(n) = r_{AR}(n) * r_{MA}(n) \quad (4.3.1)$$

donde  $r_F(n)$  es el índice de autocorrelación de la respuesta impulsional del filtro F, o sea,

$$r_F(n) = h_F(n) * h_F(-n) \quad (4.3.2)$$

siendo  $h_F(n)$  dicha respuesta impulsional.

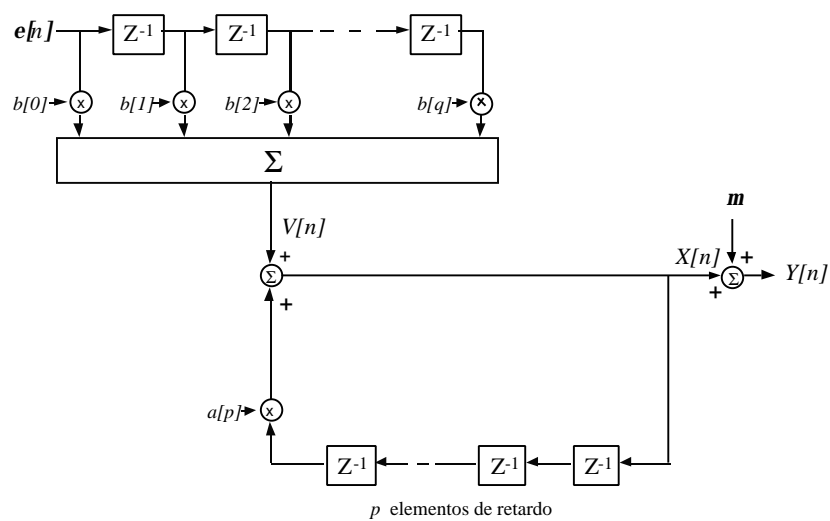


Figura 4.3.5. Esquema del modelo ARMA con la descomposición en una parte AR(9) estacional y una parte MA(q) con  $q < 9$



La obtención de los parámetros de la parte de media móvil mediante el método de mínimos cuadrados requiere la resolución de un sistema de ecuaciones no lineal. La resolución de este sistema de ecuaciones mediante un algoritmo iterativo ha presentado el inconveniente de la no convergencia del mismo. Para solventar dicho problema se ha acudido a la técnica de diseño de filtros digitales.

La figura 4.3.6 muestra la función de autocovarianza a modelar por el filtro de media móvil. Esta función se ha calculado a partir de la función de autocovarianza de las muestras resultantes de la extracción de la componente AR de las muestras iniciales. Este proceso se ha realizado, simplemente, aplicando sobre un filtro FIR, con coeficientes iguales a los de la parte autoregresiva, los valores de las muestra iniciales. Los coeficientes considerados en este caso son todos nulos salvo el mayor, cuyo valor es a.

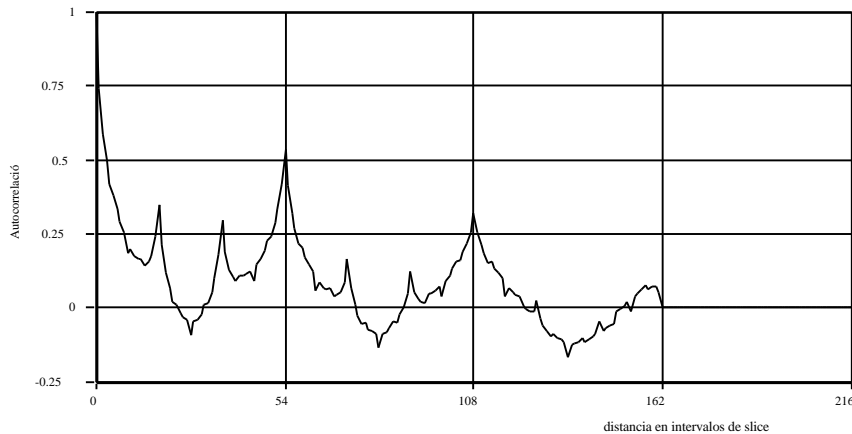


Figura 4.3.6. Índice de autocorrelación a la salida del filtro de media móvil

Observando que vuelve aparecer una semejanza en la función de autocovarianza cada tiempo de SGoP se plantea la descomposición del filtro MA en dos filtros, MA<sub>1</sub> y MA<sub>2</sub>, según se ilustra en la figura 4.3.7. El filtro MA<sub>2</sub> debe modelar las recorrelaciones que ocurren para intervalos de duración de 3 y 6 cuadros (54 y 108 slices), y el MA<sub>1</sub> modela los valores de la autocovarianza entre todos los picos. La estructura planteada se presenta en la figura 4.3.8. Esta consideración se basa de nuevo en el principio:

$$r_{MA}(n) = r_{MA_1}(n) * r_{MA_2}(n) \quad (4.3.3)$$

donde  $r_F(n)$  es el índice de autocorrelación de la respuesta impulsional del filtro F, o sea,

$$r_F(n) = h_F(n) * h_F(-n) \quad (4.3.4)$$

siendo  $h_F(n)$  dicha respuesta impulsional.

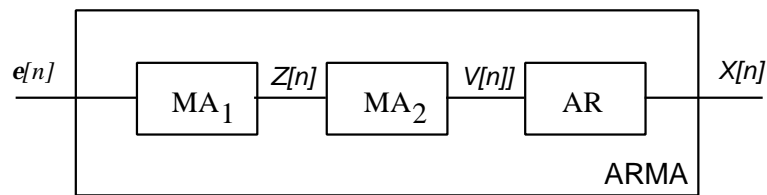


Figura 4.3.7. Descomposición del filtro ARMA en dos filtros MA y uno AR en cascada

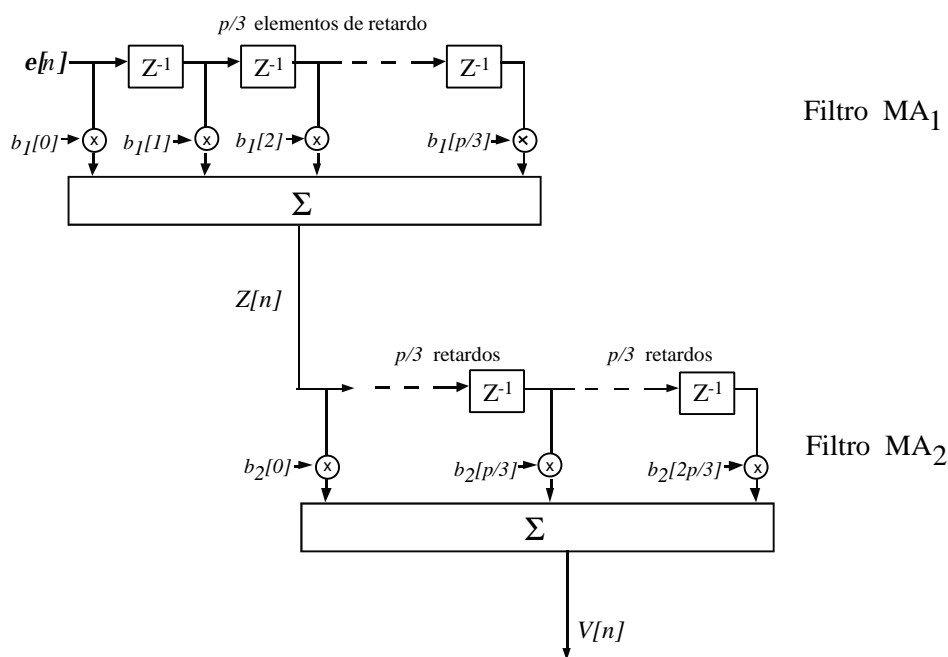


Figura 4.3.8 Filtro MA compuesto por el filtro  $MA_1$  y el filtro  $MA_2$  en cascada.

La función de autocorrelación a la salida del filtro  $MA_1(R_{ZZ})$ , se puede obtener directamente como sección de la función de autocorrelación de  $V[n]$  ( $R_{VV}$ ) y los parámetros del filtro se pueden calcular por el método del diseño de filtros digitales descrito a continuación, limitando el orden del filtro a  $p/3$ . Para el cálculo de los coeficientes del filtro  $MA_2$ , puesto que se quiere igualar perfectamente los picos de la función de autocorrelación, y el número de parámetros del filtro es pequeño, podemos utilizar simplemente identificación, considerando en este caso el orden de filtro 2. Para ello se resuelve el sistema de ecuaciones expresado en (4.3.5), con  $n=0, p/3, 2p/3$ , donde  $r_{MA_2}[n]$  en estos puntos se puede obtener a partir de los picos de  $R_{VV}$ .

$$r_{MA_2}[n] = \sum_{k=-\infty}^{\infty} h_{MA_2}[n+k]h_{MA_2}[k] \quad (4.3.5)$$

La obtención de los parámetros del filtro de media móvil  $MA_1$  se puede realizar utilizando la teoría del diseño de filtros digitales [OppSha89]. Por el teorema de Wiener-Kintchine, la densidad espectral de potencia de un proceso estocástico  $S(w)$  es la transformada de Fourier de la función de autocorrelación de dicho proceso de la siguiente manera,

$$S(w) = \mathfrak{F}\{R[k]\} = \sum_{k=-\infty}^{\infty} R[k] \cos(wk) \quad (4.3.6)$$

Sea  $H(w)$  la función de transferencia de un filtro lineal transversal. La densidad espectral de potencia  $S(w)$  de la señal de salida del filtro está relacionada con la densidad espectral de potencia a la entrada  $G(w)$

$$S(w) = |H(w)|^2 G(w) \quad (4.3.7)$$

Esto proporciona una nueva manera de encontrar los parámetros del filtro  $MA_1$ ,  $b[n] = h[n]$ . La señal de entrada del filtro es ruido blanco, por lo tanto, su densidad espectral de potencia es constante y vale  $G(w) = \sigma_e^2$ . La función de transferencia de un filtro lineal es la transformada de Fourier de su respuesta impulsional, luego, se pueden obtener los coeficientes de la respuesta impulsional mediante la expresión siguiente.

$$h[n] = \mathfrak{F}^{-1} \left\{ \frac{|S(w)|^{1/2}}{\sigma_e} \right\} \quad (4.3.8)$$

En teoría, la respuesta impulsional así obtenida tiene longitud infinita. El método más directo de obtener una aproximación FIR de un sistema igual es truncando la respuesta ideal. La expresión 4.3.8 se puede interpretar como la representación en serie de Fourier de la respuesta frecuencial periódica  $H(w)$ , con la secuencia  $h[n]$  jugando el papel de los coeficientes de Fourier. Por lo tanto, la aproximación de un filtro ideal mediante la truncación de la respuesta impulsional ideal es idéntico al problema de la convergencia de las series de Fourier.

Si se calcula  $h[n]$  como la transformada de Fourier inversa de una función real puesto que se dispone del módulo de la función de transferencia, se obtiene una  $h[n]$  simétrica. Teniendo en cuenta que el orden del filtro está limitado a  $p-1$ , el orden máximo que podemos utilizar es de  $q=8$ .

La manera más simple de obtener el filtro FIR a partir de  $h[n]$  es definiendo un nuevo sistema con la respuesta impulsional dada por:

$$h[n] = \begin{cases} h[n] & 0 \leq |n| \leq q/2 \\ 0 & \text{otros} \end{cases} \quad (4.3.9)$$

La Figura 4.3.9 presenta la función de autocorrelación del modelo obtenido junto a la autocorrelación de las muestras. En ella se observa que se han conseguido igualar los picos que presentaba la función de autocorrelación que se pretendía modelar.

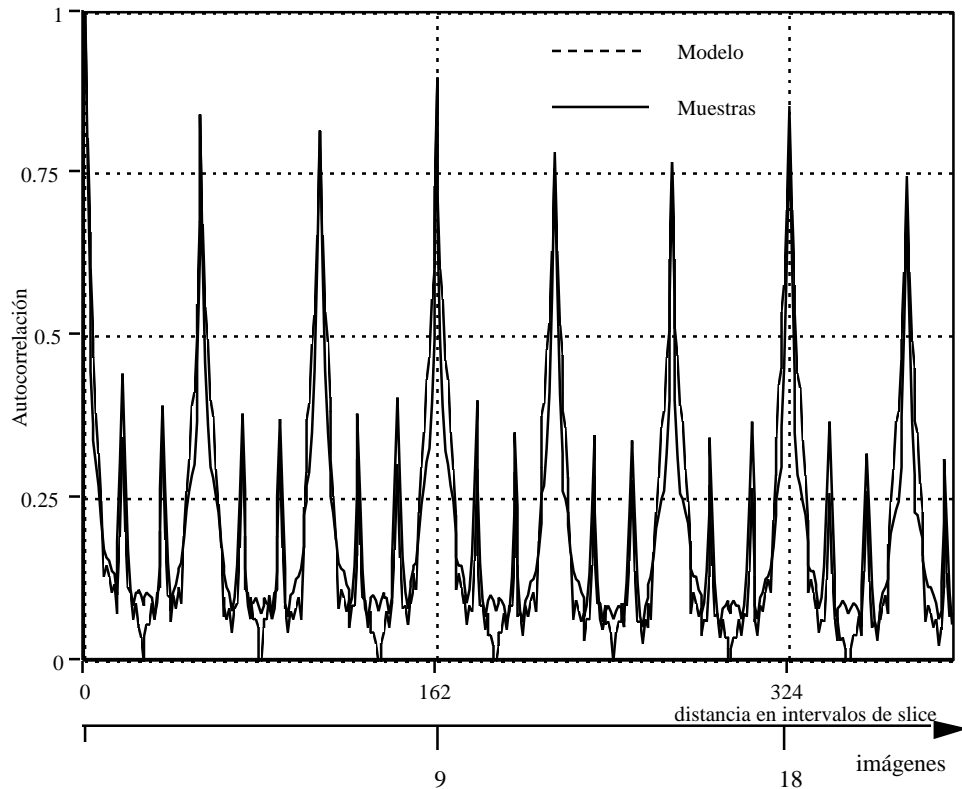


Figura 4.3.9. Índice de autocorrelación del modelo y de las muestras

Se vio que las curvas  $(T_{max}, C)$  permiten hacer una validación de los modelos de tráfico ya que capturan el comportamiento del tráfico en varias escalas de tiempo. En la Figura 4.3.10 se utilizan estas curvas para comparar el comportamiento del modelo ARMA propuesto con el tráfico real estudiado la sección 4.2, para una agregación de tráfico de 10 fuentes. Se puede observar que el modelo predice bastante fielmente el comportamiento del tráfico real, y resulta útil para determinar el punto natural de operación del sistema comentado anteriormente. Se comprueba pues, que la aproximación de la distribución de la variable número de celdas por intervalo de slice mediante una distribución gaussiana resulta ser eficaz para el cálculo de probabilidades de pérdida del orden de  $10^{-4}$ .

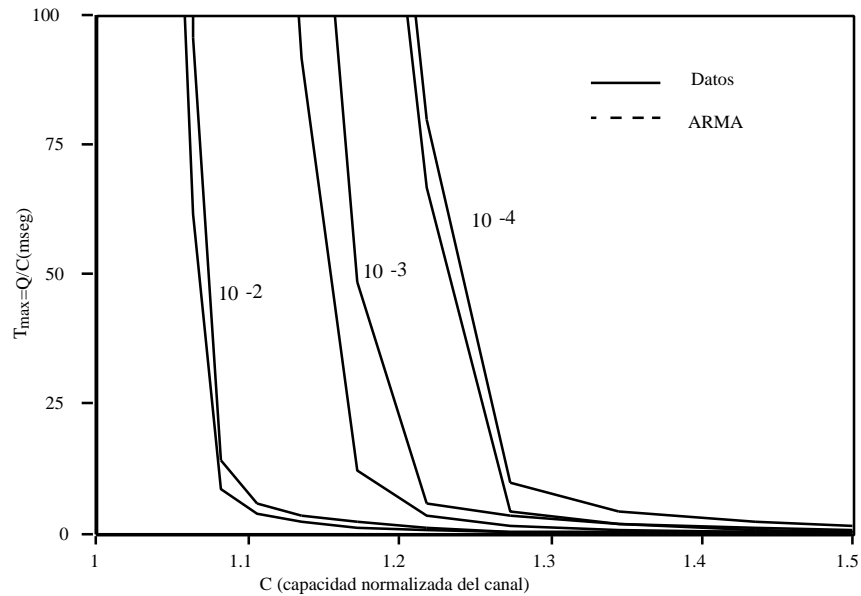


Figura 4.3.10. Comparación del comportamiento del modelo ARMA y el tráfico real , para una agregación de tráfico de 10 fuentes

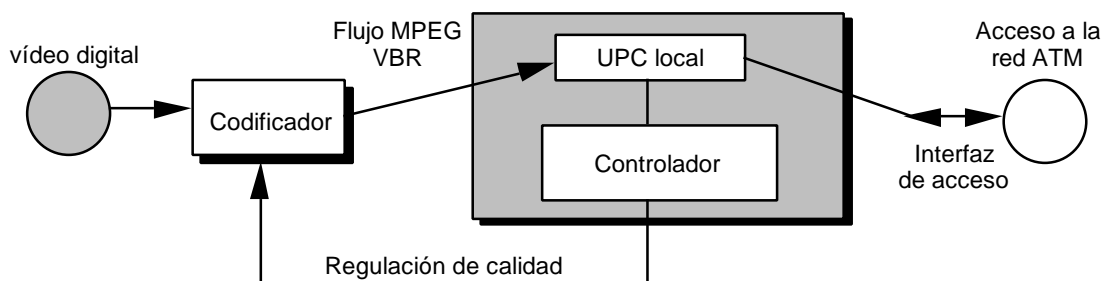
#### 4.4. Gestión dinámica de recursos en redes ATM para el tráfico de vídeo MPEG VBR

La adopción del modo de transferencia asíncrono para la RDSI de banda ancha si bien ha permitido maximizar la explotación de los recursos de la red también ha incrementado sustancialmente la complejidad del diseño de esta red y su gestión. La adecuada ubicación de recursos en las redes ATM exige la declaración de unos descriptores del tráfico generado por el servicio. No obstante, en muchos casos, el tráfico generado por los servicios es desconocido a priori, por lo que, se emplean unos descriptores característicos para el tipo de servicio. Ante las posibles desviaciones del comportamiento de tráfico generado por el servicio respecto al acordado con la red se está considerando la posibilidad de la renegociación de los descriptores de tráfico durante el transcurso de la conexión. Esta renegociación implica una nueva asignación de recursos al servicio, por lo que, el tiempo de respuesta de la red para esta asignación debe ser suficientemente rápida para que esta técnica pueda llevarse a la práctica. Asimismo, los servicios deben a su vez emplear la reserva dinámica de recursos de forma moderada para permitir a la red ofrecer un reducido tiempo de respuesta. La reserva dinámica de recursos tiene la ventaja de poder maximizar la explotación de los recursos y mejorar el grado de servicio ofrecido por la red.

La maximización de la explotación de los recursos provoca ocasionalmente situaciones de congestión en la red. Este efecto se produce cuando el conjunto de conexiones que comparten un recurso pretenden superar de forma sostenida su máximo nivel de utilización. Para controlar estas situaciones, las redes ATM pueden emplear el control reactivo EFCI. Este control se realiza a través de la marcación de las celdas que atraviesan un dispositivo congestionado que indican al receptor la situación de congestión. El dispositivo receptor notifica al emisor una disminución de la tasa generada para reducir el nivel de utilización del recurso hasta que desaparezca la situación de congestión.

Las variación del comportamiento del tráfico generado por el servicio y del nivel de utilización de los dispositivos de la red da lugar al empleo de un dispositivo que regule la tasa binaria generada por el servicio y realice las tareas necesarias para la renegociación dinámica de la asignación de recursos de la red al servicio. Por estas razones, en este trabajo se ha denominado a este dispositivo supervisor-controlador.

En la figura 4.4.1 se plantea el esquema de funcionamiento del supervisor-controlador para un codificador de vídeo MPEG VBR. Este dispositivo debe monitorizar la tasa binaria generada por el codificador para detectar las variaciones de la actividad de la secuencia de vídeo digital. La monitorización se puede llevar a cabo a partir de un mecanismo de policía local que permita observar la adecuación del tráfico ofrecido a los descriptores especificados en el contrato usuario-red.



*Figura 4.4.1. Esquema de regulación dinámica de ancho de banda para un codificador de vídeo MPEG VBR*

La detección de una desviación apreciable de la tasa binaria sobre la acordada provocará la actuación del supervisor-controlador. Cuando la tasa binaria rebasa un umbral se desencadenará el proceso de renegociación de los descriptores de tráfico. En el caso de ser aceptada la nueva ubicación de recursos se procederá a modificar los valores del UPC local empleado por el supervisor-controlador. Si por el contrario no se acepta la variación de los descriptores de tráfico, se actuará sobre el nivel de calidad de la codificación, hasta que se liberen suficientes recursos en la red y pueda ser restituida la

calidad del servicio. De igual forma, cuando se produce una notificación de congestión al supervisor-controlador se reducirá la tasa binaria generada por el codificador hasta que se supere el estado de congestión.

Para el tráfico de vídeo se han podido establecer tres niveles de actividad con los que se pueden definir los umbrales de actuación del mecanismo de reserva dinámica. Los niveles de actividad están relacionados con conjuntos de tasas binarias, cuyos valores dependen de la calidad fijada en el proceso de codificación. Los tasas binarias asociadas a los niveles de actividad pueden ser determinadas a través de la caracterización del tráfico generado por el codificador MPEG VBR. El modelo MMFP bidimensional, propuesto en la sección 3.4, permite acotar los valores de las tasas binarias correspondientes a cada nivel. Como ejemplo ilustrativo, en la tabla 4.4.1 se muestran los valores extremos de los niveles de actividad para la secuencia "Live in Central Park", codificada con los parámetros (Q=9,M=2,N=6). En la figura 4.4.2 se presentan los umbrales de actuación del mecanismo de reserva dinámica derivados de la caracterización del tráfico realizada. Obsérvese que el tiempo medio entre renegociaciones para este tipo de secuencias es del orden de minutos, lo cual hace factible el empleo de la renegociación durante el transcurso de la conexión

Estado de actividad	Tasa mínima (Kbps)	Tasa máxima (Kbps)
Actividad Regular	285	863
Actividad elevada	650	1229
Gran Actividad	1015	1594

Tabla 4.4.1. Valores extremos generados en los niveles de actividad de una fuente de vídeo MPEG VBR

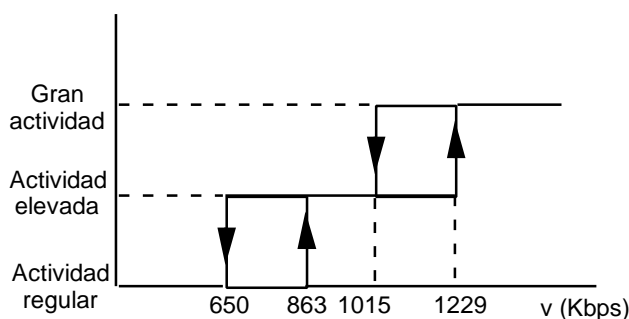


Figura 4.2.2. Umbrales de decisión de la renegociación del contrato usuario-red para el tráfico MPEG VBR

## 4.5. Conclusiones

A través del estudio realizado de la transmisión de vídeo MPEG VBR sobre redes ATM se han podido comprobar las ventajas que ofrece la multiplexación estadística en este tipo de fuentes de tráfico. Se han llevado a cabo un conjunto de simulaciones que han permitido cuantificar la reducción de los recursos necesarios cuando diferentes conjuntos de fuentes de tráfico de vídeo son multiplexados con unas restricciones de retardo y probabilidad de pérdida adecuadas para este tipo de servicios. En el caso particular de una única fuente de tráfico, los resultados obtenidos determinan los posibles parámetros descriptores de tráfico que se pueden emplear en el contrato usuario red y el dimensionado del mecanismo de policía GCRA.

Para cuantificar las ventajas de la multiplexación estadística se ha planteado una nueva definición de la ganancia de multiplexación estadística. Esta definición se ha revelado mucho más apropiada que las propuestas anteriores ya que permite una evaluación de las ventajas de la multiplexación estadística tanto para la reducción de ancho de banda como para el tamaño del buffer para un conjunto de fuentes de tráfico con las mismas restricciones de retardo y probabilidad de pérdida.

Con el estudio estadístico del tráfico agregado de vídeo MPEG VBR se ha llevado a cabo una caracterización del tráfico multiplexado mediante un modelo ARMA para la tasa binaria generada en intervalos de duración de un cuadro. La síntesis de este modelo se ha desarrollado con un nuevo procedimiento para determinar las componentes autoregresivas y de media móvil del modelo. La síntesis se ha llevado a término a partir de la observación de las relaciones temporales de los servicios de vídeo agregado. La función de autocovarianza del tráfico agregado pone de manifiesto que, en general, este tráfico mantiene las dependencias temporales del tráfico simple, de forma que aparecen máximos relativos en la función para valores de retardo en tiempos de cuadro, SGoP y GoP. Así, la función de autocovarianza presenta una estructura regular entre picos, situados en múltiplos de un intervalo de GoP.

El procedimiento propuesto para la síntesis artificial de tráfico agregado se basa en esta estructura de repetición de la función de autocovarianza. La parte autoregresiva se ha diseñado para la captura del comportamiento de los máximos situados en intervalos de GoP. De esta manera se ha conseguido aproximar la atenuación de la estructura con un filtro AR estacional. La parte MA se ha descompuesto en dos términos. Un término sintetiza los máximos relativos debidos a la relación temporal entre SGoPs. Este término da lugar a un modelo estacional de media móvil. El otro término captura el comportamiento de la secuencia entre intervalos de duración un tiempo de GoP una vez extraídas las dependencias en tiempos de SGoP. Este término se ha obtenido aplicando las técnicas de diseño de filtros digitales.



Finalmente, se ha propuesto un nuevo esquema de control del codificador de vídeo MPEG VBR. En este esquema se contempla la reserva dinámica de ancho de banda durante el transcurso de la conexión y el control reactivo EFCI.

Empleando un supervisor-controlador de la tasa binaria generada por el codificador MPEG VBR se actúa sobre la calidad de la codificación. Asimismo, el supervisor-controlador se encarga de la renegociación de los parámetros del servicio durante el transcurso de la comunicación y recibe las notificaciones del control reactivo de la red ATM.

A partir del modelo de fluidos bidimensional, desarrollado en el capítulo 3, se han presentado los umbrales de actuación del controlador que determinan el inicio de una renegociación de ancho de banda.

Según los estudios realizados en el capítulo 2 también se pueden establecer, a través de las curvas de distorsión, las correcciones de calidad necesarias para la reducción de la tasa binaria. Así, el controlador realizará un ajuste del nivel de calidad cuando se le notifique una situación de congestión o cuando el nivel de actividad aumente en la secuencia y se superen los descriptores de tráfico negociados.