

# Capítulo 3

## Caracterización del tráfico de vídeo MPEG-I VBR

En este capítulo se plantean las distintas técnicas de modelado de fuentes de tráfico para redes ATM y, en particular, el estudio se centra sobre el tráfico generado por los codificadores de vídeo MPEG VBR. Se discuten las distintas variables aleatorias que se pueden emplear para la caracterización de estas fuentes y cuáles son los parámetros que debería capturar un modelo.

Se propone una nueva técnica de modelado a nivel de cuadro basada en la caracterización individual de cada modo de funcionamiento de la codificación MPEG. Sobre este esquema de síntesis de tráfico se desarrollan cuatro modelos. Tras la evaluación de los modelos se concluye que los modelos de fluidos son los más adecuados para la caracterización del tráfico vídeo MPEG VBR y que los modelos autoregresivos son válidos para la predicción y suavizado de este tipo de tráfico. A partir de la consideración de la conformación del tráfico de vídeo MPEG VBR, dado su interés por la eliminación de las nocivas llegadas periódicas a los multiplexores y conmutadores de las redes ATM, propias de las variaciones intrínsecas de la tasa binaria generada por el codificador, se elabora un nuevo modelo de fluidos a nivel de GoP. Este modelo se muestra especialmente adecuado para la captura del comportamiento del tráfico de vídeo MPEG VBR entregado a la red. El modelo de fluidos a nivel de GoP presenta como mejora, respecto a los modelos de este género propuestos en la actualidad, un buen ajuste de los histogramas resultantes de la codificación real y una perfecta aproximación a corto y largo plazo de la función de autocovarianza.

Como contribución a la suavización del tráfico de vídeo MPEG VBR se confecciona un nuevo modelo ARIMA que, si bien no puede utilizarse para la síntesis de tráfico artificial de este tipo de fuentes, se muestra como un excelente predictor de tráfico, por lo que, puede ser empleado en la conformación predictiva. Cabe destacar que el valor de los parámetros de este predictor, obtenidos a partir del estudio de diferentes secuencias

de larga duración, se mantienen casi invariantes, incluso para distintos pasos de cuantificación del codificador.

### 3.1. Introducción

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: multimedia, 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 tráfico 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 ( $\lambda$ ) [And93]:

- i) Tasa media de generación ( $m$ )

$$m = E(\lambda) \quad (3.1.1)$$

- ii) Varianza de la tasa de generación ( $\sigma^2$ )

$$\sigma^2 = E((\lambda - m)^2) \quad (3.1.2)$$

iii) Coeficiente de variación de la tasa de generación (c). Este valor se obtiene como cociente de la desviación estándar y la media de la tasa generada.

$$c = \frac{\sigma}{m} \quad (3.1.3)$$

iv) Momentos de orden superior de la tasa de generación.

$$m^{(j)} = E[\lambda^j] \quad j = 3, 4, \dots \quad (3.1.4)$$

v) Función de distribución de probabilidad de la tasa de generación.

vi) Función de autocorrelación de la tasa de generación. Esta función permite ajustar la relación temporal de las tasas generadas por el modelo.

vii) Generación en avalancha cuando la tasa se mantiene a niveles significativamente superiores a la media de forma sostenida. Este parámetro se cuantifica a través del índice de dispersión de cuentas o mediante los valores que ajustan el algoritmo GCRA para una determinada probabilidad de descarte de celdas [AndGar92]. Está relacionado directamente con el Sustainable Cell Rate (SCR). El índice de dispersión de cuentas (IDC) se define a partir del número de celdas generadas en un intervalo de duración t ( $N_t$ ), de forma que [Gus91]:

$$IDC(t) = \frac{Var(N_t)}{E(N_t)} \quad (3.1.5)$$

viii) Relación de rafagueo (Burstiness, B). Es el cociente entre la tasa máxima ( $R_p$ ) de generación y la tasa media.

$$B = \frac{R_p}{m} \quad (3.1.6)$$

ix) Parámetro de Hurst (H) o relación de autosemejanza. Intenta relacionar la variación a largo y corto plazo de la tasa de generación.

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 [GihGia91]:

- 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 [Hui90]. 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 contínuos o discretos, atendiendo al tipo de síntesis realizada.

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 transferencias 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 contínuo 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).

### **Clasificación de modelos de tráfico**

La clasificación de modelos se puede realizar atendiendo a múltiples criterios; como modelos en tiempo discreto o contínuo, parámetros que sintetizan, modelos de tasa media en un intervalo o modelos de tasa de generación instantánea, etc... Seguidamente se presenta una clasificación en cuatro grandes categorías según las técnicas de génesis de eventos que emplean.

#### *Modelos markovianos de renovación.*

Los modelos de renovación sintetizan procesos de generación de tasa instantánea con distribución de probabilidad genérica. Los eventos generados son independientes e idénticamente distribuidos. El principal inconveniente presentado es la incorrelación de las generaciones, dado que, la captura de la relación temporal entre llegadas es

imprescindible para hallar los efectos provocados por la generación en avalancha en redes ATM.

Para capturar la correlación temporal de los eventos se controlan las generaciones a través de una cadena de Markov. Estos procesos reciben el nombre de procesos de renovación markovianos. Los procesos de renovación markovianos se caracterizan por un conjunto de estados, que forman una cadena de Markov, y los tiempos de transición entre estados. La función de distribución de probabilidad de los tiempos de transición entre estados es genérica y depende exclusivamente del estado previo a la transición a un nuevo estado de la cadena de Markov. Este tipo de modelos se emplea para determinar la tasa instantánea, o de forma equivalente, el tiempo entre llegadas consecutivas. Así, cada vez que se produce una transición entre estados se considera que se ha producido una nueva generación de un paquete. Este modelo puede ser ampliado con la consideración de llegadas en grupo, asociando un segundo proceso independiente del tiempo entre llegadas. El proceso de llegadas define el número de paquetes de una llegada. En general, este proceso de generación en grupo se liga a otra cadena de Markov con un número finito de estados, donde cada estado define el número de paquetes de la llegada.

Dentro de estos modelos de renovación markovianos los más empleados son los procesos de llegada markovianas (Markovian Arrival Process, MAP). En este caso particular se considera que el tiempo de permanencia en un estado está controlado por una cadena de Markov con el primer estado absorbente [Neu91]. Esta cadena se la suele denominar cadena de fase. La determinación del tiempo de permanencia en un estado, es decir el tiempo entre dos generaciones consecutivas, dependerá del tiempo necesario para que la cadena de fase alcance el estado absorbente. Puesto que la distribución del tiempo de permanencia depende del estado previo, en los procesos de renovación markovianos, la probabilidad de inicio de los estados de la cadena de fase dependerá del estado desde el cual se realiza la transición. La gran ventaja que presentan los modelos MAP es su tratamiento analítico, dado que, la superposición de procesos MAP es un proceso MAP.

Como ampliación de los procesos MAP se puede asociar al momento de la llegada el número de elementos de ésta. Los procesos con llegadas en grupo, con determinación del tiempo entre llegadas a través de un proceso MAP, son los procesos B-MAP (Batch Markovian Arrival Process) donde, la distribución del número de elementos de la llegada es independiente del proceso MAP y está regulada por una cadena de Markov con un número finito de estados [Gar92].

Como casos particulares de los procesos MAP cabe resaltar:

i) Procesos de Markov. En este caso las transiciones entre estados siguen una distribución exponencial, o lo que es equivalente, la cadena de fase está formada únicamente por dos estados: el estado inicial y el absorbente, con distribución dependiente del estado.

ii) Procesos de renovación de fase. Es un proceso MAP donde la cadena de fase empleada en la transición presenta para todos los estados la misma distribución y

siempre se inicia en el mismo estado con lo que la cadena de fase es la que únicamente regula las llegadas.

iii) Proceso de Poisson. Es simplemente el caso particular de un proceso de renovación de fase donde sólo existen dos estados, el inicial y el absorbente.

### *Modelos de tasa modulada por Markov.*

Estos modelos sintetizan la tasa de generación según una función de distribución de probabilidad que depende del estado en que se encuentra una cadena de Markov. Como caso particular, se emplea una distribución exponencial de generación de tasas, dando lugar a los procesos de Poisson modulados por Markov (Markov-Modulated Poisson Processes, MMPP). Estos procesos conjugan las características exponenciales del tiempo de permanencia en los estados y del tiempo entre generaciones. Así, en cada estado de la cadena de Markov se generarán tiempos entre llegadas exponenciales según una tasa de generación poissoniana que depende del estado en curso. Estos modelos permiten un tratamiento analítico y han sido ampliamente estudiados [FisMei93]. Los procesos MMPP también pueden ser contemplados como procesos de renovación markovianos a través de una transformación que relaciona el generador infinitesimal de la cadena de markov y la matriz diagonal de tasas generadas en cada estado con el generador infinitesimal de un MAP [Gar92]. Los casos más simples de MMPP son los de un estado, es decir, un proceso de Poisson, y los modelos ON/OFF, también denominados procesos de Poisson interrumpidos, (Interrupted Poisson Processes, IPP). Los IPP son procesos de dos estados donde la tasa de generación poissoniana en un estado es nula y en el otro es constante [Kuc73].

### *Aproximación de fluidos.*

Los modelos de fluidos podrían considerarse como un caso particular de los modelos de tasa modulados por Markov cuando la tasa generada está controlada por una cadena de Markov y la generación es determinista en cada estado. Estos procesos se denominan procesos de fluidos modulados por Markov (Markov Modulated Fluid Processes, MMFP). Estos procesos merecen una especial mención, puesto que, permiten un tratamiento analítico diferenciado respecto a los procesos B-MAP [ElwMit91]. La aproximación realizada en estos modelos es considerar la tasa de generación constante durante el intervalo de tiempo en que se encuentra la cadena de Markov en un estado. Por ello, también se denominan Markov Modulated Constant Rate (MMCR). La cadena de Markov discretiza las posibles tasas de generación en un número de niveles igual al número de estados, como se ilustra en la figura 3.1.1. Esta aproximación es válida en los casos donde el número de unidades básicas generadas durante un intervalo es muy elevado. En consecuencia, la casuística de una unidad generada carece de importancia y el análisis se debe centrar en el volumen de información transferido.

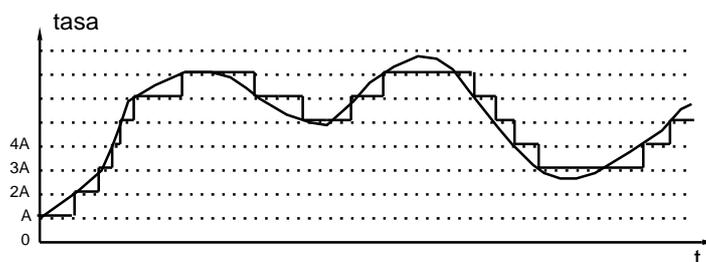


Figura 3.1.1. Aproximación de la tasa de generación mediante un proceso MMFP

Dentro de los MMFP los más frecuentemente empleados son:

i) MMFP binomial ( $M/M/\infty/S$ ). Estos modelos presentan una función de distribución de probabilidad binomial para la generación de tasas y una función de autocorrelación exponencial [MagAna88]. La cadena de Markov es una cadena de nacimiento y muerte que puede describir un sistema con un número infinito de servidores, con distribución de tiempo exponencial de parámetro  $\beta$ , donde las llegadas se producen desde una población finita de  $S$  elementos con tasa individual de generación poissoniana  $\alpha$ . Este sistema, según la notación de Kendall, sería un  $M/M/\infty/S$ . Según el estado de la cadena de Markov se genera a una tasa constante. Se puede considerar un valor mínimo de generación en el estado 0 del valor  $A_1$ . Las tasas de generación entre estados consecutivos se suelen fijar de forma que difieran en un valor constante  $A$ . Según esta definición, la aproximación se basa en discretizar uniformemente la tasa de generación en un conjunto de  $S+1$  estados, donde, en un estado  $i \in \{0, 1, \dots, S\}$  de la cadena de Markov el tiempo de permanencia está distribuido exponencialmente con tasa  $(S-i)\alpha + i\beta$  y la tasa binaria generada es de valor  $iA + A_1$ . De esta forma, obtendríamos una aproximación como la presentada en la figura 3.1.2.

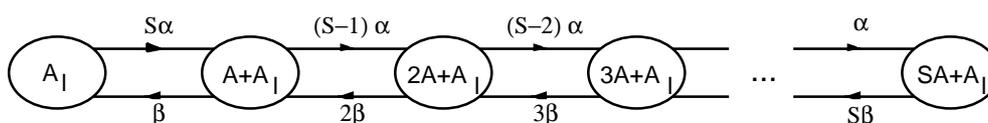


Figura 3.1.2. Estructura genérica del modelo de fluidos con distribución de probabilidad binomial

El caso más simple de los MMFP con distribución binomial son los denominados modelos MMCR ON/OFF, donde la cadena de Markov sólo dispone de dos estados con generación nula y generación a tasa  $A$ . Este caso queda ilustrado en la figura 3.1.3.

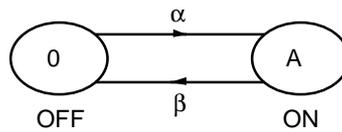


Figura 3.1.3. Miniflujo ON/OFF

Los procesos MMFP binomiales tienen la propiedad fundamental de poderse descomponer en la agregación de procesos elementales ON/OFF. Así, un proceso genérico como el presentado en la figura 3.1.2 se puede descomponer en S procesos simples, como se muestra en la figura 3.1.4.

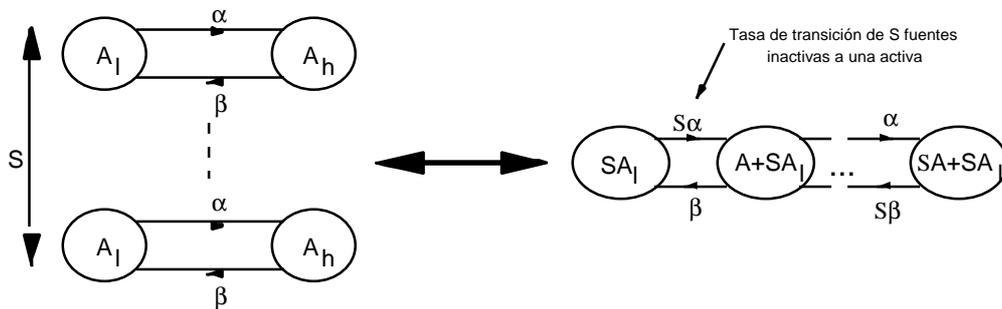


Figura 3.1.4. Descomposición de una fuente multiestado en minifuentes

donde,

$$A_{\min} = SA_1 \quad (3.1.7)$$

$$A_h = A + A_1 \quad (3.1.8)$$

ii) Procesos Autoregresivos de media móvil discretos (Discrete AutoRegressive Moving Average, DARMA). Estos procesos también generan una tasa constante según el estado de una cadena de Markov. Su tratamiento analítico es muy complejo, aunque, presentan como cualidad el ajuste de cualquier función de distribución de probabilidad y de la función de autocorrelación [JacLew83]. Para funciones de distribución simples y considerando una función de autocovarianza con decaimiento exponencial, estos modelos presentan una estructura regular fácilmente sintetizable, denominada DAR(1) [HeyTab92], [FraArn94]. La cadena de Markov, en general, permite las transiciones entre cada par de estados. La probabilidad de encontrar el proceso en un estado de la cadena viene fijada por la función de densidad de probabilidad sintetizada.

*Procesos autoregresivos.*

Estos procesos han sido ampliamente estudiados en la literatura y en su forma más general se denominan procesos autoregresivos, integrativos de media móvil (autoregressive integrative moving average, ARIMA) [BoxJen86], [AlpArg91]. 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 [GruCos91][YegJab93]. 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  $Y=(y_0,y_1,\dots,y_n)$  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  $W=(w_0,w_1,\dots,w_n)$ . 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

$$y_n = a_1 y_{n-1} + a_2 y_{n-2} + \dots + a_p y_{n-p} + w_n \quad (3.1.9)$$

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 :

$$x_n = b_0 w_n + b_1 w_{n-1} + b_2 w_{n-2} + \dots + b_q w_{n-q} \quad (3.1.10)$$

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 considerarse 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:

$$z_n = c_1 z_{n-1} + c_2 z_{n-2} + \dots + c_d z_{n-d} + w_n \quad (3.1.11)$$

donde

$$c_i = \binom{d}{i} (-1)^{i+1} \quad i \in \{1, 2, \dots, d\} \quad (3.1.12)$$

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.

Además, si existe una tendencia clara del comportamiento de la varianza en largos intervalos de tiempo es conveniente aplicar alguna transformación como la propuesta por Box-Cox [Uri92]. En este caso la generación puede expresarse por la relación:

$$y_n = Ke^{\gamma n + w_n} \quad (3.1.13)$$

donde K y  $\gamma$  son constantes por determinar.

La interpretación de un proceso ARIMA(p,d,q) puede ser realizada definiendo el operador de retardo  $z^{-1}$  [Pro83]. De forma que, la expresión general de un proceso ARIMA(p,d,q) quedaría expresada por su transformada  $z$  como:

$$Y(z) = \frac{B(z)}{A(z)C(z)} W(z) \quad (3.1.14)$$

donde,

$$Y(z) = \mathbf{Z}(y) = \sum_{n=0}^{\infty} y_n z^{-n}$$

$$W(z) = \mathbf{Z}(w) = \sum_{n=0}^{\infty} w_n z^{-n}$$

$$B(z) = \mathbf{Z}(b) = \sum_{n=0}^q b_n z^{-n}$$

$$A(z) = \mathbf{Z}(a) = \sum_{n=0}^p a_n z^{-n}$$

$$C(z) = \mathbf{Z}(c) = \sum_{n=0}^d c_n z^{-n}$$

Interpretando esta expresión como la relación entre entrada y salida de un filtro digital con excitación  $w_n$  y cuya salida es  $y_n$  en un instante dado, podríamos definir la función de transferencia del filtro  $H(z)$  como:

$$H(z) = \frac{Y(z)}{W(z)} = \frac{B(z)}{A(z)C(z)} \quad (3.1.15)$$

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  $c_j$  realizada en la expresión 3.1.12, 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  $k \in \{1, 2, \dots, p\}$ , 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 figura 3.1.5.

La serie temporal  $w$ , en general, se denomina serie residual. Se suele considerar como la parte impredecible de la siguiente generación a partir de los valores anteriores. Estos procesos estocásticos son incorrelados y su distribución suele ser gaussiana.

Dentro de los modelos autoregresivos se han introducido mejoras para ajustar, además de las funciones de autocorrelación, las funciones de distribución de probabilidad [GruCos91] [Ens94]. En particular, cabe resaltar los modelos Transform-Expand-Simple (TES) [FroMel94][LazPac94]. Los modelos TES realizan una transformación de una secuencia uniformemente distribuida para obtener la función de distribución deseada. La generación del proceso uniforme, además, captura las principales características de la función de autocorrelación.

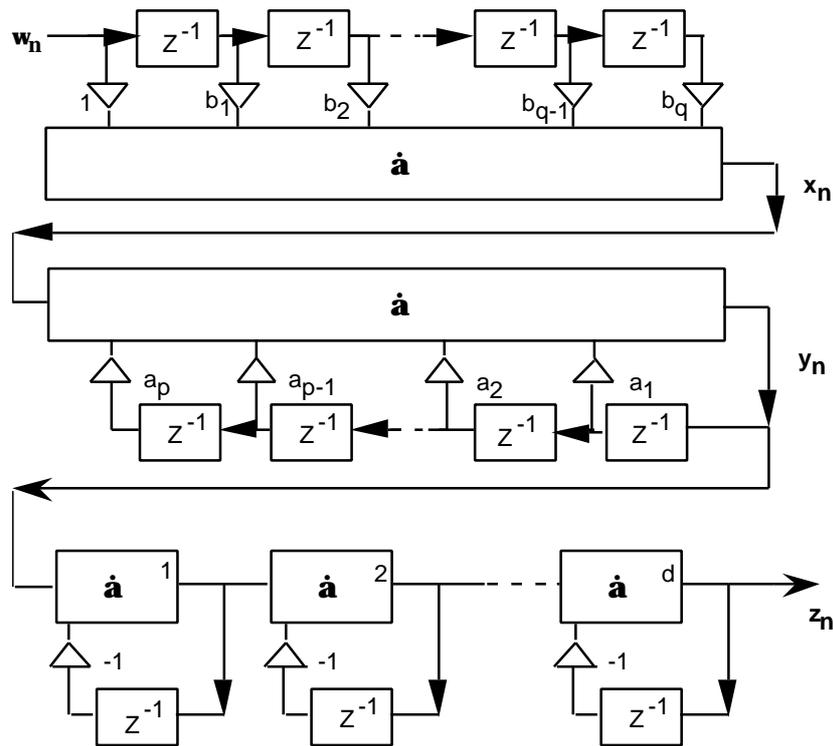


Figura 3.1.5. Esquema de un filtro ARIMA(p,d,q)

*Procesos autosemejantes.*

Los modelos autosemejantes sintetizan la tasa media generada en un intervalo fijo. A diferencia del resto de modelos presentados anteriormente intentan capturar el comportamiento que presentan las fuentes de tráfico a largo plazo, en vez de a corto plazo. En general, la mayoría de los procesos estocásticos reducen su relación temporal cuanto más distantes son los eventos, aunque, en particular, las fuentes de vídeo manifiestan una relación temporal que disminuye mucho más lentamente de lo habitual. Por ello, los modelos autosemejantes también reciben el nombre de persistentes, o bien modelos de efecto Hurst [Gar93].

Existen diferentes maneras de observar el efecto Hurst a través del análisis de las series temporales. La persistencia queda patente cuando la densidad espectral de potencia de las series tiene un decaimiento con relación a la frecuencia de la forma  $f^{-\alpha}$ , con  $\alpha$  un valor constante y positivo.

Los procesos persistentes también se caracterizan por disponer de una función de autocovarianza no sumable, lo cual implica que la autocorrelación decae de forma más lenta que una exponencial. Se ha comprobado que los procesos de Hurst presentan un decaimiento hiperbólico.

Otra manera de identificar los procesos persistentes es a través de las series temporales agregadas. Se considera una serie temporal agregada  $S^m$  como la serie obtenida a partir de una serie original  $S$ , donde el valor de las muestras es resultado de la adición de las muestras de la serie original que pertenecen a un mismo intervalo de duración  $a^m$ , donde  $a$  es una base con valor real, positivo y constante. En general, se emplean como intervalos de agregación  $\{10^1, 10^2, 10^3, 10^4, \dots\}$ . Se observa que las series temporales agregadas reducen su varianza según la relación:

$$\text{Var}(S^m) \approx \frac{\text{Var}(S)}{m^\beta} \quad 0 < \beta \leq 1 \quad (3.1.16)$$

Cuando los procesos estocásticos estacionarios no presentan una dependencia a largo término el valor de  $\beta$  es 1. Valores de  $\beta$  inferiores nos determinan que, para diferentes escalas temporales, los procesos mantienen un parecido en su comportamiento, razón por la cual se denominan autosemejantes. Este parecido se diluye mucho más lentamente que en los procesos sin relación a largo término, los cuales, reducen su varianza rápidamente cuando se considera la agregación en pocos órdenes de magnitud.

Los procesos autosemejantes se identifican a través del parámetro de Hurst ( $H$ ), el cual se relaciona con el valor de  $\beta$  según

$$H = 1 - \frac{\beta}{2} \quad (3.1.17)$$

La estimación más precisa de  $H$  se puede llevar a cabo con la técnica *reescalad adjusted range statistics* (R/S) presentada en [GarWil94].

Los efectos de dependencia a largo término pueden ser sintetizados por los procesos denominados procesos de diferenciación fraccional. Esos procesos se derivan de los procesos integrativos puros descritos en los modelos autoregresivos. Por ello, también se suelen denotar como procesos ARIMA(0,d,0) donde  $d$  no necesariamente debe ser un número entero, como en los procesos integrativos. Esta generalización de los procesos integrativos se hace en base a la expresión del polinomio en  $C(z)$  que caracterizaba la multiplicidad del polo  $z=1$  en el filtro asociado a los modelos ARIMA con parte integrativa no nula. Así,

$$C(z) = (1 - z^{-1})^d = \sum_{i=0}^{\infty} \binom{d}{i} (-1)^i z^{-i} \quad -\frac{1}{2} < d < \frac{1}{2} \quad (3.1.18)$$

donde

$$\binom{d}{i}(-1)^i = \frac{\Gamma(-d+i)}{\Gamma(-d)\Gamma(i+1)} \quad (3.1.19)$$

La relación entre la multiplicidad del polo (d) y el parámetro de Hurst es:

$$d = H - \frac{1}{2} \quad (3.1.20)$$

Los modelos autosemejantes además se conjugan con alguna transformación de la función de distribución gaussiana obtenida en el proceso ARIMA(0,d,0) para ajustar el comportamiento de las series temporales empíricas.

### 3.2. Modelado de las fuentes de tráfico de vídeo MPEG VBR en ATM.

La codificación de vídeo según el estándar MPEG, da lugar a una tasa binaria de generación de tráfico de gran variabilidad. La búsqueda de modelos que capturen el comportamiento de estos codificadores de calidad semiconstante es en la actualidad un problema no resuelto que se intenta abordar a través de una gran cantidad de vertientes diferentes.

La mayoría de modelos presentados en la literatura han conseguido ajustar un conjunto de los parámetros descriptores del tráfico MPEG VBR. Las técnicas empleadas son, en la mayoría de los casos, las expuestas en la sección 3.1. En algunos casos, se han complementado alguna técnica con otra para elaborar un modelo que ajustara un mayor número de características del tráfico real. En pocos estudios se han empleado un número elevado de imágenes codificadas, dado el coste computacional que entraña la codificación de largas secuencias. La mayoría de los análisis emplean secuencias de pocos minutos de duración, con lo que, si bien se pueden observar las características del tráfico real, no ofrecen suficiente precisión como para considerarlos aplicables con generalidad.

Los modelos de tráfico de vídeo se pueden desarrollar para que capturen, o bien la tasa instantánea de generación, o bien la tasa media generada durante un intervalo dado. Los modelos de tasa instantánea son extraordinariamente difíciles de ajustar, dadas las características de generación de los codificadores MPEG VBR [ChaLeo94]. La relación temporal y espacial de las imágenes, los diferentes modos de codificación de los macrobloques y las variaciones de la complejidad y actividad de las escenas establecen unas dependencias en la generación de celdas muy difíciles de sintetizar. Los modelos que mejor podrían capturar la tasa instantánea serían los MAP, aunque este tipo de modelos han sido poco estudiados para la síntesis de tráfico de vídeo MPEG VBR. En general, los codificadores no entregan directamente el tráfico al interfaz de usuario sino que, normalmente, se habilita algún sistema de suavización. La suavización se aplica para reducir la variabilidad del tráfico y su tasa de pico, con lo que se reduce sustancialmente la probabilidad de violación del contrato usuario-red. La suavización se

lleva a cabo a través de un pequeño buffer de almacenamiento, introduciéndose un retardo desde la generación de la celda hasta su entrega a la red. El empleo de la suavización permite mantener una tasa instantánea de generación de celdas entregadas a la red aproximadamente constante durante un intervalo de tiempo. Por este motivo, los esfuerzos se han centrado en la síntesis de modelos de tasa de generación media durante un intervalo dado. Estos modelos de generación media en intervalos de duración fijados también pueden ser empleados para la aproximación de generaciones de tasas instantáneas, considerando que la tasa instantánea generada sigue una función de distribución de probabilidad aproximable (geométrica, uniforme, exponencial...) cuya tasa media se obtiene a través del modelo. De esta manera se pretende simular el efecto de variación entre llegadas de celdas de forma continuada.

Los modelos de tasa media se pueden ajustar para distintos intervalos de duración. Estos intervalos pueden ser aleatorios, regidos por un proceso estocástico, o deterministas. En la mayoría de los casos se han intentado sintetizar modelos cuyos intervalos de duración eran deterministas y ligados a las características estructurales de la codificación MPEG. Así, se han desarrollado modelos asociados a la tasa generada en un macrobloque, una tira, una imagen, un SGoP y un GoP [PanZar94]. Conforme se aumenta la duración del intervalo empleado van desapareciendo las interrelaciones temporales de la codificación MPEG y de las propias imágenes.

El intervalo de duración de un macrobloque se ha considerado en muy pocos estudios porque la información generada es muy reducida, aproximadamente, un valor medio de dos celdas, por lo que no son suficientemente representativos cuando se inserta un buffer en el acceso a la red [Gha93].

Cuando se estudian sistemas con un retardo de almacenamiento reducido, aproximadamente 10 ms, el intervalo considerado es la duración de una tira. Cuando se contemplan retardos máximos tan reducidos, sólo se puede emplear modos de codificación I o P en el algoritmo MPEG. Esto es debido a que la introducción de las imágenes B producen un retardo en la codificación en tiempo real de tantos cuadros como imágenes B consecutivas se empleen en un SGoP, dado que, es necesario el conocimiento de la siguiente referencia temporal para poder codificarlas.

La mayoría de los modelos de generación de tasa para intervalo de tira se han elaborado solo empleando el modo intra. En [Gar93] se proponía un modelo autosemejante ARIMA(0,d,0) con parámetro de Hurst  $H = 0,8$ , obtenido sobre una secuencia de 170.000 imágenes con una corrección de la función de distribución gaussiana a una distribución híbrida Gamma-Pareto. En [PanZar94] se aproximaba la generación de la tasa generada en una tira mediante un modelo de tasa generada en una imagen, basado en el ajuste de la función de distribución de probabilidad. La tasa generada en la imagen se consideraba como el valor medio de una función de distribución de probabilidad exponencial que determinaba los valores de las tasas generadas en las tiras de la imagen. Esta misma técnica se propone para generar la tasa en tiempo de macrobloque. Los resultados obtenidos se basan en el análisis de un extenso conjunto de secuencias.

Los modelos para tiempos de tira son difíciles de sintetizar dadas las características de codificación del algoritmo MPEG. Las dependencias temporales periódicas hacen difícil conjugar la aleatoriedad del comportamiento con su relación temporal. Se

considera complejo abordar el problema, aunque, se empleen modelos híbridos que permitan capturar la periodicidad de los modos de codificación y la posible relación espacial y temporal de los cuadros.

La mayoría de los estudios se centran en el modelado de la tasa generada en un intervalo de cuadro. La razón fundamental es que el sistema perceptivo humano no aprecia un retardo inferior a 100 ms aunque es admisible hasta unos 200 ms [Gar93]. Puesto que el tiempo de duración de una imagen está alrededor de los 40 ms es perfectamente aceptable considerar una demora de este orden en el buffer de suavización. A su vez, se permite que el algoritmo de codificación MPEG pueda emplear el modo de codificación B con un SGoP de 2 imágenes en servicios interactivos. Los modelos de generación de tasa media en un intervalo de cuadro se han elaborado utilizando las técnicas descritas en la sección 3.1, así como, combinaciones de ellos.

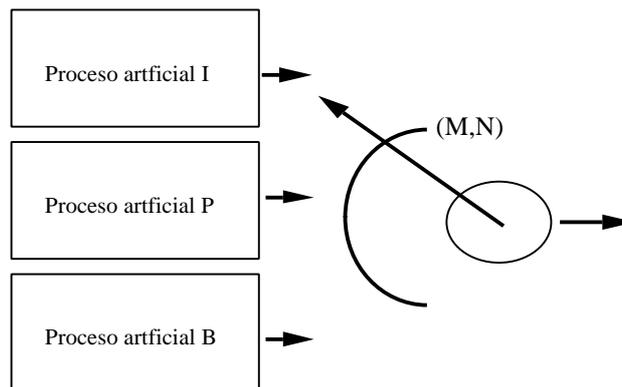
Los servicios de vídeo que consumirán mayor ancho de banda de las redes ATM serán los servicios de distribución y de vídeo bajo demanda. La información a transmitir de estos servicios suele estar almacenada en bases de datos de alta velocidad de acceso [ChaAna94]. Estos servicios no tienen una restricción fuerte en el tiempo de respuesta, ya que se considera admisible un retardo de hasta 4 segundos. Esto permite que se pueda realizar una suavización de hasta la unidad estructural de codificación de un GoP, cuya duración habitual oscila entre 100 ms y 375 ms. Cuando se sintetizan modelos para intervalos de duración de un GoP desaparecen las relaciones temporales introducidas por el algoritmo de codificación, de forma que, sólo se manifiestan las relaciones propias de la complejidad y actividad de las escenas.

### **3.3. Modelado del tráfico MPEG VBR a nivel de cuadro.**

Las características de la codificación de vídeo con el algoritmo MPEG VBR hacen difícil la elaboración de un modelo de tráfico que sintetice el comportamiento estadístico, así como, la relación temporal generada en el proceso de codificación. El modelado de la tasa generada en la codificación de un cuadro debe capturar la periodicidad de los modos de codificación seleccionada en el algoritmo de compresión MPEG. La periodicidad del empleo de los modos de codificación viene fijada por el número de imágenes que componen un SGoP (M) y un GoP (N). Así, la distancia entre dos imágenes intra será N y entre dos imágenes de referencia P, o una I y una P, será M. Con lo que el número resultante de cuadros consecutivos codificados en modo B es  $M-1$ .

Los modelos que sintetizan este comportamiento periódico presentan un elevado grado de complejidad en su estructura. La complejidad se cuantifica en los modelos markovianos con un número elevado de estados o en los modelos autoregresivos con órdenes de las componentes grandes. A su vez, estos modelos son difíciles de ajustar para que presenten una función de distribución de probabilidad semejante a la obtenida en las series temporales reales.

Para desarrollar un modelo de tráfico de vídeo MPEG VBR a nivel de cuadro, con la captura de las relaciones temporales de la codificación, se ha abordado su síntesis a través de la descomposición natural del tráfico en las tasas generadas en los modos I, P y B. De esta manera se han asociado tres variables aleatorias correspondientes a cada uno de ellos. Una vez capturado el comportamiento de cada modo de codificación por separado se podrá generar tráfico, de forma artificial, a través de las realizaciones de cada variable aleatoria en el mismo orden en que el codificador selecciona el modo de codificación de los cuadros. De forma gráfica se presenta en la figura 3.3.1 el mecanismo empleado en la generación de las tasas en cada cuadro.



*Figura 3.3.1. Modelado del tráfico de vídeo MPEG a nivel de cuadro a través de la síntesis individual de los distintos modos de codificación*

El proceso de transmisión seguido por el codificador de vídeo MPEG VBR puede ser presentado de forma esquemática según se muestra en la figura 3.3.2. El proceso se puede descomponer en dos niveles: nivel de ráfaga y nivel de celda. La intensidad de la ráfaga queda determinada por la tasa binaria del cuadro codificado y el nivel de celda por la fragmentación, empaquetado y entrega de los paquetes a la red. Una vez determinada la tasa de generación en un cuadro, se pueden considerar distintas técnicas de fragmentación y entrega de la información a la red. Las estrategias propuestas en diferentes estudios han sido:

- i) La distribución equiespaciada durante el intervalo de duración del cuadro.
- ii) La entrega de celdas de forma uniformemente distribuida en los slots del interfaz.
- iii) El volcado periódico de toda la información en slots consecutivos en intervalos de duración de un cuadro.

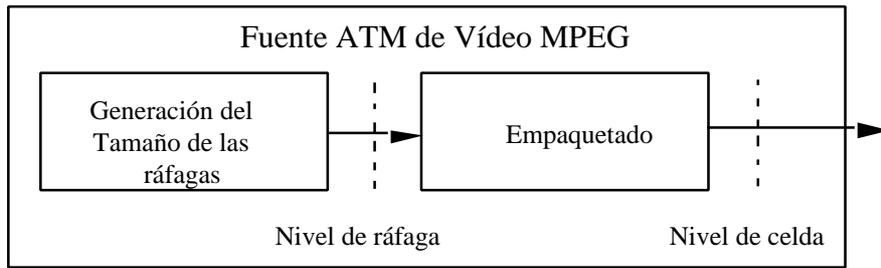


Figura 3.3.2. Esquema de transmisión de un codificador de vídeo MPEG

La consideración más realista es la distribución equiespaciada cuando el servicio admite el almacenamiento y posterior entrega de la información. En servicios de vídeo con fuertes restricciones temporales lo más adecuado es emplear una distribución uniforme, dado que, la información se fragmenta y entrega en cuanto está disponible. El estudio de la entrega periódica en slots consecutivos del interfaz tiene interés por establecer una cota superior en la evaluación de prestaciones.

En la figura 3.3.3 se ilustra que, una vez determinada información generada en un cuadro y tras su posterior empaquetamiento, las celdas son entregadas al interfaz con separación constante. Este enfoque del problema fue adelantado en los resultados presentados en [MatSal94].

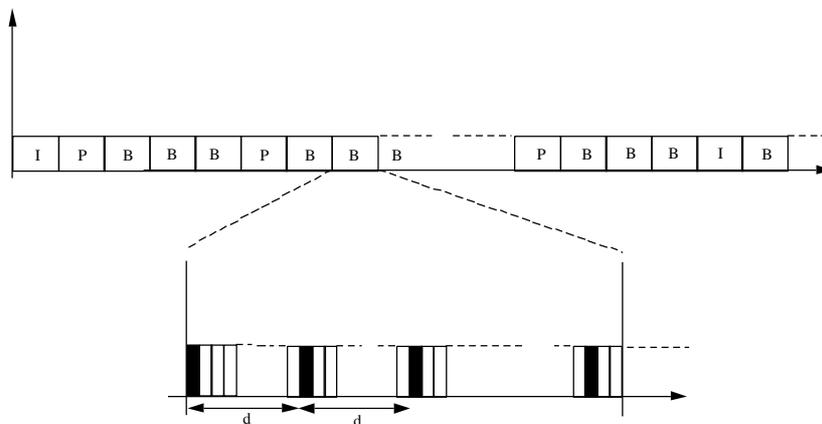


Figura 3.3.3. Empaquetamiento y entrega de la información generada en un cuadro de vídeo MPEG VBR

La elección de técnicas de modelado de los distintos modos de codificación del algoritmo MPEG se ha llevado a cabo a partir de las realizaciones de las series temporales reales y las características estadísticas de éstas. Los modelos de estas variables aleatorias deben ser modelos de generación de tasa media en el intervalo de duración de un cuadro. Puesto que el intervalo es determinista, los modelos B-MAP se reducirían a la consideración exclusiva de cómo se producirían las llegadas en grupo. Por consiguiente, este tipo de modelos serían equivalentes a los modelos de fluidos.

Otra formulación, de los modelos B-MAP sería la particularización MAP. En este caso, se pueden plantear las características entre generación de celdas, o tasa instantánea, donde este enfoque sería equivalente a los modelos de tasa modulada por Markov, como los MMPP.

Los modelos de tasa modulada por Markov y los modelos autoregresivos han sido los más estudiados para la captura del comportamiento del tráfico MPEG VBR. Dentro de los modelos de tasa modulada por Markov también se incluyen los modelos de fluidos, considerando que la generación en el estado es de tasa constante, y los modelos híbridos de tasa modulada por Markov y los autoregresivos. Los modelos autosemejantes no se han empleado en la génesis de tráfico MPEG VBR con diferentes modos de codificación, por lo que se considera una problemática abierta para posibles trabajos futuros.

Los modelos de tasa modulada por Markov admiten una interpretación ligada al comportamiento estadístico de la codificación. Se pueden asociar los estados de la cadena de Markov a diferentes niveles de actividad de la codificación, los cuales van relacionados con la actividad y complejidad de las imágenes. Así, se suelen agrupar diferentes valores de tasas generadas en un nivel de actividad. El estudio de la función de distribución de probabilidad de estas tasas generadas permite la construcción del modelo de tasa modulada por Markov, asociando a cada estado de actividad la función de distribución observada o una aproximación de ésta.

Un detenido análisis de la realización de las series temporales muestra que existe un elevado nivel de correlación en las generaciones consecutivas de cada modo de codificación. La variabilidad de la tasa generada en cuadros consecutivos es muy reducida. Además, el efecto de suavizado considerado en el interfaz provoca que la tasa instantánea de generación se mantenga constante durante todo el intervalo. Este comportamiento no se asemeja al presentado por los modelos de tasa modulada MMPP. Un modelo MMPP que capturara el comportamiento estadístico de las series temporales obtenidas, para los diferentes modos de codificación de los cuadros, dispondría de un número reducido de estados, asociados a diferentes niveles de actividad del tráfico generado. Las transiciones y tiempos de permanencia en los estados se podrían ajustar según el comportamiento asociado del tráfico real. La generación en un estado seguiría una distribución exponencial, lo cual provocaría una alta incorrelación entre generaciones consecutivas, efecto opuesto al observado en las realizaciones reales. La incorrelación presentada en la generación de tasas de cuadro en un estado hace poco adecuado este tipo de modelos para estudio del comportamiento del tráfico simple a nivel de cuadro MPEG VBR.

Para sintetizar la relación temporal en la generación dentro de un estado de un modelo de tasa modulada por Markov se ha complementado este tipo de modelos con los modelos autoregresivos [NomFuj89], [YegJab93]. El planteamiento híbrido se realiza considerando un modelo de tasa modulada por Markov donde en cada estado se sintetiza el comportamiento de cada modo de codificación, por separado, a través de un modelo autoregresivo o se emplea un modelo autoregresivo capaz de capturar la relación periódica temporal. Esta última propuesta no ha sido estudiada ante la dificultad de la síntesis de modelos autoregresivos capaces de capturar los efectos periódicos. Como se mostrará en la sección 3.5 del presente trabajo, se ha elaborado un

modelo autoregresivo de estas características, de forma que las técnicas empleadas y propuestas en su síntesis, podrían ser aplicadas a un modelo híbrido de tasa modulada por Markov y autoregresivo. Este estudio queda postergado a posibles trabajos futuros.

Los modelos híbridos de tasa modulada por Markov y autoregresivos tienen la capacidad de ajustar muy bien las características de generación de tráfico real. Este tipo de modelos pueden ser empleados en la síntesis artificial de tráfico y son especialmente útiles en la evaluación de dispositivos o ubicación de recursos cuando se realiza un análisis vía simulación. La desventaja que presenta es su desconocido tratamiento analítico y la dificultad para emplearlos para la predicción de tráfico, características que presentaban, respectivamente, los modelos de tasa modulada con función de distribución de probabilidad en cada estado obtenida empíricamente [SkeDix92] y los modelos autoregresivos.

Como consideración especial de los modelos de tasa modulada por Markov se pueden contemplar los modelos de fluidos. En este caso se pueden aplicar los mismos criterios de asociación de estados a niveles de actividad y emplear una cadena de Markov para la génesis de las tasas de generación en un nivel de actividad. La imbricación de cadenas de Markov en los estados de otra cadena de Markov se puede resolver como una única cadena de Markov [GihGia91]. Estas cadenas de Markov definirían un valor de tasa de generación durante el intervalo de un cuadro cuando se definieran como cadenas de Markov discretas. En los resultados avanzados de este trabajo [MatSal94] se modeló cada una de las tasas de generación de los modos de codificación con un modelo de fluidos binomial, donde se podían agrupar algunos estados según la distribución de los niveles de actividad de las series temporales.

El modelo de fluidos binomial admite otra interpretación que se ajusta al comportamiento de las series temporales. En este caso, para cada modo de generación, se discretizan los valores de tasa generados en un número uniforme de niveles, donde cada nivel o estado tiene un tiempo de permanencia distribuido exponencialmente. Este tipo de modelo es muy adecuado, dado que, si el número de niveles de discretización es elevado se capturan los efectos de correlación temporal de imágenes próximas. La correlación temporal se obtiene aproximando el valor de la tasa de cuadros consecutivos de un modo de cuantificación a un mismo valor. La función de distribución de probabilidad se ajusta través de la probabilidad de visita a los estados en régimen permanente, cuyo resultado se aproxima mediante el histograma de las series temporales. El grosor de las barras consideradas en el histograma es igual al paso de discretización y el valor central de la barra es el de la tasa constante generada. En el modelo de fluidos binomial es necesario que la función de distribución de probabilidad de la variable aleatoria sintetizada sea aproximable con una distribución binomial y que la función de autocorrelación presente un decaimiento exponencial.

A fin de aumentar las posibilidades de este tipo de modelado se han propuesto como modelos de fluidos los DARMA. Estos modelos mantienen la discretización de las tasas de generación pero la discretización no es necesariamente uniforme. Asimismo, la función de autocorrelación puede presentar comportamientos más generales. En los estudios realizados sobre este tipo de modelos [Rei93] [MatSal95] sólo se han considerado funciones de autocorrelación con decaimiento exponencial, por los que se

denominan DAR(1). Se plantean como posibles trabajos futuros el estudio de modelos de órdenes superiores, DARMA(p,q).

Otra familia de modelos son los ARIMA(p,d,q) que han sido ampliamente empleados en otros campos de investigación. Estos modelos ajustan perfectamente las funciones de autocorrelación del tráfico generado por la codificación MPEG VBR. Presentan como inconveniente que, sin ningún acondicionamiento, pueden generar eventos negativos. Asimismo, las funciones de distribución de probabilidad generadas, a veces, necesitan un reajuste mediante una función de transformación, a fin de aproximarse a la función de distribución de las series temporales. La mayoría de los estudios realizados en el modelado de tráfico de vídeo MPEG VBR sólo han contemplado el modo de codificación intra. El modelo de tráfico MPEG VBR, considerando los tres tipos de codificación con generación separada en tres variables aleatorias, ha sido estudiado en [MatSal94][Ens94] [MatSal95].

Para estudiar el comportamiento y la bondad del ajuste de las diferentes alternativas analizadas se ha llevado a cabo la síntesis de cuatro modelos de tráfico de vídeo MPEG VBR con generación separada de tasas de cuadro según el modo de codificación. En la sección 3.5 se presentará un modelo ARIMA(p,d,q) de tráfico de vídeo MPEG VBR a nivel de cuadro que, como se analizará, presenta el inconveniente de disponer de una parte integrativa no nula. Esto da lugar a que este modelo tenga una generación no acotada de tráfico, por lo que no es adecuado para la síntesis artificial. Sin embargo, como se comprobará, presenta unas extraordinarias características como predictor.

Los modelos estudiados para la síntesis artificial de tráfico MPEG VBR a nivel de cuadro se presentan en las próximas subsecciones, y en la última se establece un análisis comparativo de los mismos. Parte de estos resultados fueron avanzados en [MatSal95]. Estos resultados se han obtenido a través de la codificación de una secuencia de 4200 imágenes representativa del comportamiento del tráfico MPEG VBR. Esta secuencia se ha obtenido de la grabación del concierto "Live in Central Park" del grupo América. Los parámetros seleccionados en la codificación fueron  $Q = 9$ ,  $M = 2$  y  $N = 4$ .

### **3.3.1. Modelo ARMA para fuentes de tráfico MPEG VBR a nivel de cuadro.**

La identificación de los modelos ARMA asociados a los series temporales de cada modo de generación se han desarrollado aplicando las técnicas descritas en [BoxJen94], [Key88], [Mar87] y [Wei93]. En esta bibliografía se describen diferentes algoritmos para la estimación de modelos ARMA. En este estudio se ha empleado el algoritmo de Durbin para la estimación por mínimos cuadrados de la parte MA. Para la parte AR se han utilizado dos posibles métodos. El primer método empleado busca el ajuste de la función de autocorrelación a través de las ecuaciones modificadas de Yule-Walker que dan lugar a una matriz de Toeplitz, no necesariamente definida positiva. El segundo método de estimación de la parte AR es el denominado de mínimos cuadrados aplicado sobre las ecuaciones modificadas de Yule-Walker, considerando un número de valores de autocorrelación mayor que el orden de la parte AR. También se ha empleado el algoritmo de Marquardt para la estimación por mínimos cuadrados de un modelo

ARMA general, a través de un procedimiento no lineal para la minimización de la suma cuadrática de los errores.

Para determinar los ordenes  $p$  y  $q$  de los modelos ARMA se han comparado los valores de los coeficientes obtenidos a través de las distintas técnicas. También se ha aplicado un análisis de los residuos a través de su función de autocorrelación y el test de falta de ajuste de Portmanteau modificado [BoxJen94].

Los modelos ajustados para las tasas por cuadro generadas en cada modo de codificación se detallan en la tabla 3.3.1. Se especifica el orden de la parte MA y AR con los coeficientes obtenidos por el algoritmo de Marquardt. Cabe destacar que el test de Portmanteau da un valor de falta de ajuste del 70%, en el peor de los casos. Este valor es muy bajo para un modelo ajustado a los valores reales, por lo que, se puede considerar un modelo muy preciso dado que captura muy bien el comportamiento de la serie real. Para la generación de tráfico de forma sintética se deben hallar también la media y varianza del ruido blanco que debe excitar al filtro ARMA para obtener un ajuste del valor medio y varianza de la serie temporal. En la síntesis de tráfico se ha forzado a valor cero aquellos eventos obtenidos negativos que, a pesar de ser poco frecuentes, carecen de sentido físico. Los valores obtenidos de media y varianza en el proceso generado por el modelo ARMA, prácticamente, no se ven modificados al establecer esta restricción.

### 3.3.2 Modelo DAR(1) para fuentes de tráfico MPEG VBR a nivel de cuadro.

Los modelos DAR(1) ajustados al comportamiento de las tasas generadas para cada modo de codificación MPEG se han desarrollado a partir de una aproximación exponencial de la función de autocorrelación y la aproximación de la función de densidad de probabilidad a través de una función de densidad binomial negativa [HeyTab92]. La función de autocorrelación exponencial se ha obtenido a través de regresión log-lineal de las funciones de autocorrelación derivadas de las series temporales de cada modo de codificación. En la sección 2.4 se mostró que las tasas de los modos de codificación presentaban un decaimiento aproximable exponencialmente. En la tabla 3.3.1 se muestran los índices de autocorrelación obtenidos a partir de los valores reales.

La aproximación de la función de densidad de probabilidad por una función de densidad binomial negativa es muy adecuada según el comportamiento observado en las figuras 2.4.4 a 2.4.6. Los histogramas correspondientes a las tasas generadas en cada modo de codificación presentan una forma de campana con una larga cola de decrecimiento lento para valores altos. Este comportamiento de la función de densidad de probabilidad puede ajustarse bien con la función de densidad de probabilidad binomial negativa, la cual podría ser expresada como:

$$f_k = \binom{k+r-1}{k} p^r (1-p)^k \quad \begin{matrix} 0 \leq p \leq 1 \\ r > 0 \end{matrix} \quad (3.3.1)$$

La interpretación de los parámetros de la función de densidad binomial negativa se puede realizar en base a la agregación de un conjunto de  $r$  fuentes geométricas idénticas, e independientes, con probabilidad de generación  $p$ . Obsérvese que la función de densidad binomial negativa admite como expresión continua la función de densidad gamma dada por:

$$f(t) = \frac{\lambda(\lambda t)^{s-1}}{\Gamma(s)} e^{-\lambda t} \quad (3.3.2)$$

donde,

$$\Gamma(s) = \int_0^{\infty} t^{s-1} e^{-t} dt \quad (3.3.3)$$

La distribución gamma se caracteriza por el parámetro de forma ( $s$ ), y el parámetro de escala ( $\lambda$ ). En este caso se puede interpretar la distribución a partir de la agregación de las tasas de un conjunto de  $s$  fuentes exponenciales de tasa  $\lambda$ .

Los momentos de la distribución binomial negativa se pueden derivar de la expresión 3.3.1. En el ajuste de los modelos se han empleado el valor medio ( $m$ ) y la varianza ( $\sigma^2$ ) que se relacionan con los parámetros de la distribución binomial negativa según:

$$\begin{aligned} m &= \frac{r(1-p)}{p} \\ \sigma^2 &= \frac{r(1-p)}{p^2} \end{aligned} \quad (3.3.4)$$

A partir de los valores de las series se han determinado los parámetros a través de las relaciones [CohHey93]

$$\begin{aligned} p &= \frac{m}{\sigma^2} \\ r &= \frac{mp}{1-p} \end{aligned} \quad (3.3.5)$$

La matriz de transición de la cadena de Markov viene dada por la relación [JacLew83]:

$$P = \rho I + (1-\rho)Q \quad (3.3.6)$$

con  $\rho$  el índice de autocorrelación,  $I$  la matriz identidad y  $Q$  una matriz cuadrada de filas iguales, cuyos valores se obtienen a partir de la función de densidad binomial negativa. Cada elemento de la fila se obtiene por identificación con el valor la función de densidad binomial negativa hasta el valor máximo alcanzado por la serie temporal a modelar. El último elemento de la fila se fija para que la suma del resto de elementos sea 1. Así, siendo  $K_{\max}$  el valor máximo de la serie, la expresión de  $P$  sería

$$P = \rho \begin{pmatrix} 1 & 0 & \dots & 0 \\ 0 & 1 & 0 & \dots & 0 \\ & 0 & 1 & & \cdot \\ \cdot & \cdot & & & 0 \\ 0 & 0 & \dots & & 1 \end{pmatrix} + (1-\rho) \begin{pmatrix} f_0 & f_1 & \dots & f_{K_{\max}} & 1 - \sum_{i=0}^{K_{\max}} f_i \\ f_0 & f_1 & \dots & f_{K_{\max}} & 1 - \sum_{i=0}^{K_{\max}} f_i \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ f_0 & f_1 & \dots & f_{K_{\max}} & 1 - \sum_{i=0}^{K_{\max}} f_i \end{pmatrix} \quad (3.3.7)$$

Obsérvese que en este caso la discretización realizada en las tasas es la unidad, es decir, se aplica la máxima resolución posible. Con esta discretización se obtienen un número muy elevado de estados en la cadena de Markov, tantos como el valor de pico de las tasas generadas. Como aproximación, y a fin de reducir el orden de estados de los modelos de cada modo de codificación, se ha realizado una agrupación de estados, o lo que es equivalente, se ha reducido la resolución de la discretización a un paso de 2000. Con esta consideración se hallan los valores de los parámetros del modelo DAR(1) con los mismos coeficientes de autocorrelación  $\rho$ , dividiendo el valor medio y la varianza de las series temporales por el paso de cuantificación y por su cuadrado, respectivamente. La síntesis de tráfico se obtendrá a través de la visita de cada uno de los modelos de cada modo de codificación de la misma forma que lo realiza el algoritmo MPEG, y multiplicando las tasas obtenidas en los modelos por el paso de discretización empleado. Los parámetros que caracterizan los modelos DAR(1) ajustados se presentan en la tabla 3.3.1.

### 3.3.3. Modelo de fluidos binomial para fuentes de tráfico MPEG VBR a nivel de cuadro

Los modelos MMFP M/M/∞/S se caracterizan por presentar una función de autocorrelación exponencial y una función de distribución de probabilidad binomial. El estudio y ajuste de este modelo se puede realizar a través del análisis del sistema de cola M/M/∞/S que presenta una distribución binomial de parámetro  $p$ . Definiendo la variable aleatoria  $Z$  asociada al número de elementos en el sistema M/M/∞/S y  $X$  otra variable aleatoria binomial relacionada con  $Z$  a través de

$$X = A \cdot Z \quad (3.3.8)$$

Se pueden interpretar los estados del sistema como el número de minifuentes ON/OFF en actividad que generan en estado ON una tasa constante  $A$  y en estado OFF una generación nula. Puesto que las series temporales no alcanzan valores nulos de generación, se ha propuesto una mejora en el modelo, de forma que su comportamiento sea más realista. En este caso se considera que existe un estado de generación mínimo. Para ello se ha definido una nueva variable binomial relacionada con  $X$  de la forma:

$$Y = X + k \quad (3.3.9)$$

donde  $k$  es el valor mínimo generado por el modelo. Se puede interpretar  $Y$  a partir de la agregación de un número de  $S$  minifuentes ON/OFF con tasa de generación constante en cada estado. Como generalización a los resultados presentados en [MagAna88] se ha considerado que cada minifuerte dispone de una tasa de generación  $A_I$  no nula, en estado OFF y una tasa  $A_m$  en estado ON. La diferencia entre estas tasas es constante y de valor  $A$ . De esta manera podemos describir una minifuerte como:

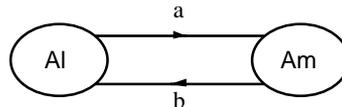


Figura 3.3.4. Modelo de minifuerte de dos estados

con

$$A = A_m - A_I > 0 \quad (3.3.10)$$

Para determinar la función de autocorrelación de estas fuentes es necesario recurrir a la evolución temporal de la probabilidad de encontrarse en un estado cuando se parte de uno fijo. El régimen transitorio puede ser fácilmente calculado a través del análisis de balance de flujos de la cadena de Markov de dos estados según las ecuaciones:

$$\begin{aligned} \frac{dP_{OFF}(t)}{dt} &= -aP_{OFF}(t) + bP_{ON}(t) \\ \frac{dP_{ON}(t)}{dt} &= aP_{OFF}(t) - bP_{ON}(t) \end{aligned} \quad (3.3.11)$$

cuya expresión matricial sería:

$$\frac{dP(t)}{dt} = QP(t) \quad (3.3.12)$$

con

$$Q = \begin{pmatrix} -a & b \\ a & -b \end{pmatrix} \quad \text{y} \quad P(t) = \begin{bmatrix} P_{OFF}(t) \\ P_{ON}(t) \end{bmatrix} \quad (3.3.13)$$

Este sistema de ecuaciones diferenciales ordinarias admite una solución de la forma vectorial.

$$P(t) = \sum_{k=0}^1 c_k e^{w_k t} v_k \quad (3.3.14)$$

donde  $w_k$  son los autovalores del generador infinitesimal ( $Q$ ),  $v_k$  sus autovectores por la derecha y  $c_k$  constantes que dependen de las condiciones de contorno. El polinomio característico en  $w$  asociado a  $Q$  sería:

$$C(w) = w[w + (a + b)] \quad (3.3.15)$$

por lo que los autovalores son:

$$\begin{aligned} w_0 &= 0 \\ w_1 &= -(a + b) \end{aligned} \quad (3.3.16)$$

Los autovectores asociados son:

$$v_0 = \begin{bmatrix} \pi_{\text{OFF}} \\ \pi_{\text{ON}} \end{bmatrix} \quad v_1 = \begin{bmatrix} 1 \\ -1 \end{bmatrix} \quad (3.3.17)$$

con

$$\begin{aligned} \pi_{\text{OFF}} &= \frac{b}{a + b} \\ \pi_{\text{ON}} &= \frac{a}{a + b} \end{aligned} \quad (3.3.18)$$

Nótese que  $\pi_{\text{OFF}}$  y  $\pi_{\text{ON}}$  son las probabilidades de encontrar la fuente en estado OFF y ON en régimen permanente. Las constantes  $c_0$  y  $c_1$  se obtienen directamente a través de las condiciones iniciales y en régimen permanente, de forma que

$$\begin{aligned} c_0 &= 1 \\ c_1 &= P_{\text{OFF}}(0) - \pi_{\text{OFF}} = -P_{\text{ON}}(0) + \pi_{\text{ON}} \end{aligned} \quad (3.3.19)$$

Con lo que:

$$\begin{aligned} P_{\text{OFF}}(t) &= \pi_{\text{OFF}} + (P_{\text{OFF}}(0) - \pi_{\text{OFF}})e^{-(a+b)t} \\ P_{\text{ON}}(t) &= \pi_{\text{ON}} + (P_{\text{ON}}(0) - \pi_{\text{ON}})e^{-(a+b)t} \end{aligned} \quad (3.3.20)$$

De esta manera hallamos:

$$\begin{aligned}
 p_{00}(\tau) &= P(\text{estado OFF / OFF en } \tau = 0) = \pi_{\text{OFF}} + \pi_{\text{ON}} e^{-(a+b)\tau} \\
 p_{01}(\tau) &= P(\text{estado ON / OFF en } \tau = 0) = 1 - p_{00}(\tau) = \pi_{\text{ON}} - \pi_{\text{ON}} e^{-(a+b)\tau} \\
 p_{11}(\tau) &= P(\text{estado ON / ON en } \tau = 0) = \pi_{\text{ON}} + \pi_{\text{OFF}} e^{-(a+b)\tau} \\
 p_{10}(\tau) &= P(\text{estado OFF / ON en } \tau = 0) = 1 - p_{11}(\tau) = \pi_{\text{OFF}} - \pi_{\text{OFF}} e^{-(a+b)\tau} \quad (3.3.21)
 \end{aligned}$$

La autocorrelación de las tasas generadas por la minifuerza se obtiene por la definición de la función para la variable aleatoria Y

$$\begin{aligned}
 r_{YY}(\tau) = E[Y(t)Y(t+\tau)] &= A_1^2 \pi_{\text{OFF}} p_{00}(\tau) + A_m A_1 \pi_{\text{OFF}} p_{01}(\tau) + \\
 &A_1 A_m \pi_{\text{ON}} p_{10}(\tau) + A_m^2 \pi_{\text{ON}} p_{11}(\tau) \quad (3.3.22)
 \end{aligned}$$

Teniendo en cuenta que el valor medio generado por la minifuerza es:

$$\eta = E[Y(t)] = A_1 \pi_{\text{OFF}} + A_m \pi_{\text{ON}} \quad (3.3.23)$$

Se puede expresar la autocorrelación como:

$$r_{YY}(\tau) = \eta^2 + (A_m - A_1)^2 \pi_{\text{ON}} \pi_{\text{OFF}} e^{-(a+b)\tau} \quad (3.3.24)$$

Por lo que la autocovarianza  $c_{YY}(\tau)$  será:

$$c_{YY}(\tau) = r_{YY}(\tau) - \eta^2 = (A_m - A_1)^2 \pi_{\text{ON}} \pi_{\text{OFF}} e^{-(a+b)\tau} \quad (3.3.25)$$

y la varianza se obtendrá a través de la autocovarianza como

$$\sigma^2 = c_{YY}(0) = (A_m - A_1)^2 \pi_{\text{ON}} \pi_{\text{OFF}} \quad (3.3.26)$$

El comportamiento de la fuente M/M/ $\infty$ /S se obtendrá a partir de la adición de variables aleatorias independientes asociados a las S minifuerzas, por lo que, la tasa media generada  $m_s$ , la varianza  $\sigma_s^2$  y la covarianza  $c_{ss}$  se obtienen directamente:

$$\begin{aligned}
 m_s &= S\eta = S(A_1 \pi_{\text{OFF}} + A_m \pi_{\text{ON}}) \\
 \sigma_s^2 &= S(A_m - A_1)^2 \pi_{\text{ON}} \pi_{\text{OFF}} = S A^2 \pi_{\text{ON}} \pi_{\text{OFF}} \quad (3.3.27) \\
 c_{ss}(\tau) &= S(A_m - A_1)^2 \pi_{\text{ON}} \pi_{\text{OFF}} e^{-(a+b)\tau} = S A^2 \pi_{\text{ON}} \pi_{\text{OFF}} e^{-(a+b)\tau}
 \end{aligned}$$

Estas expresiones se pueden obtener de forma equivalente a partir de las expresiones de una fuente M/M/ $\infty$ /S compuesta de S minifuerzas con las mismas tasas de transición entre estados, a y b, pero con generación nula en estado OFF y generación A en estado ON. La varianza y función de autocovarianza serían iguales pero el valor medio debería ser corregido en SA<sub>1</sub> unidades.

El procedimiento seguido para el ajuste de los modelos binomiales al comportamiento de las series temporales de tasas generadas en los modos de codificación se basa en la consideración de una variable aleatoria Y asociada al proceso estocástico real y una variable X con las mismas características que Y pero con valor medio SA<sub>l</sub> unidades inferior. De esta manera su relación queda expresada como:

$$Y = X + SA_l \quad (3.3.28)$$

Considerando que el modelo ajusta los valores máximos y mínimos de serie temporal tenemos que:

$$\begin{aligned} \text{Max}(X) &= \text{Max}(Y) - \text{Min}(Y) = SA_m - SA_l = SA \\ E[X] &= E[Y] - SA_l \end{aligned} \quad (3.3.29)$$

La probabilidad de estado ON de las fuentes:

$$p = \pi_{ON} = \frac{a}{a + b} \quad (3.3.30)$$

se puede obtener a través de Y como:

$$p = \frac{E[Y] - \text{Min}(Y)}{\text{Max}(Y) - \text{Min}(Y)} \quad (3.3.31)$$

y considerando la relación:

$$A = \frac{\sigma_Y^2}{\text{Max}(X)p(1 - p)} \quad (3.3.32)$$

Con lo que se obtiene un valor de S que se ajusta al entorno superior más próximo:

$$S = \left\lceil \frac{\text{Max}(X)}{A} \right\rceil \quad (3.3.33)$$

Finalmente, se recalcula A para el valor obtenido de N de forma que:

$$A = \frac{E[X]}{Sp} \quad (3.3.34)$$

Para obtener un valor preciso de la varianza es necesario variar el valor máximo o mínimo de la serie temporal. En este caso hemos considerado que el ajuste mejoraba al modificar el valor mínimo, de forma que, tras algunas operaciones algebraicas que se derivan de la expresión 3.3.32,

$$\sigma_Y^2 = \frac{(E[Y] - \text{Min}(Y))(\text{Max}(Y) - E[Y])}{S} \quad (3.3.35)$$

llegamos a :

$$\text{Min}(Y) = E[Y] - \frac{\sigma_Y^2 S}{\text{Max}(Y) - E[Y]} \quad (3.3.36)$$

Finalmente, a través de la regresión log-lineal de la autocorrelación de los datos obtenemos el coeficiente de autocorrelación  $\rho$ , de forma que, definiendo

$$\gamma = -\ln \rho = a + b \quad (3.3.37)$$

hallamos las tasas de transición a través de la relación 3.3.30:

$$\begin{aligned} a &= p\gamma \\ b &= \gamma - a \end{aligned} \quad (3.3.38)$$

Estas tasas de transición están expresadas por tiempo de duración de un cuadro, por lo que es necesario multiplicarlos por la duración de un cuadro si se considera un modelo continuo.

Los parámetros característicos de los modelos binomiales ajustados, para los modos de codificación se encuentran especificados en la tabla 3.3.1.

### 3.3.4. Modelo de fluidos binomial condicionado para fuentes de tráfico MPEG VBR a nivel de cuadro

El modelo binomial presentado en la sección 3.3.3. capturaba el comportamiento individual de cada uno de los modos de codificación del algoritmo MPEG. La síntesis de tráfico se obtenía a través de la simulación del comportamiento físico del codificador MPEG, visitando periódicamente cada variable aleatoria, acorde con la elección de parámetros M y N, de igual forma al funcionamiento de la codificación. Para introducir un comportamiento más similar al presentado por el codificador MPEG, se ha introducido una relación entre el valor adoptado por las diferentes variable aleatorias que caracterizan cada modo de funcionamiento.

En la evolución de las series temporales de las tasas generadas para los modos I, P y B, presentados en la figura 2.3.1, se observa que, cuando se producen cambios notables en la variación de la tasa generada en los cuadros I, las imágenes P y B muestran una correlación en su variación de tasa. Sin embargo, si el nivel de actividad de las imágenes I se mantiene, se aprecia una rápida disminución en las tasas generadas por los cuadros P y B, de forma que, en un intervalo de actividad regular, las tasas P y B presentan una relación proporcional y también regular.

Para simular la relación temporal en los cambios de actividad en las tasas de generación de cada modo de codificación, se ha realizado una mejora en el modelo MMFP binomial. A partir de las series temporales de las tasas de generación se ha estudiado la probabilidad del valor de la tasa generada en los modos de codificación P y B cuando del modo I está en un nivel de actividad. Este estudio se ha llevado a cabo a partir de los intervalos de discretización de cada variable aleatoria. Para realizar el análisis, definimos:

$\Delta\lambda_m^k$  = al intervalo de discretización k asociado al modo de codificación m, donde, el intervalo está centrado en la tasa generada por el estado k-ésimo del modelo MMFP binomial del modo de codificación m y la anchura del intervalo es el paso A de una minifunte.

$p_P^{ij}$  = probabilidad de que la tasa generada por el modo P ( $\lambda_P$ ) esté en el intervalo  $\Delta\lambda_P^j$  cuando la tasa generada por el modo I ( $\lambda_I$ ) está en el intervalo  $\Delta\lambda_I^i$

$$p_P^{ij} = P \left[ \frac{\lambda_P \in \Delta\lambda_P^j}{\lambda_I \in \Delta\lambda_I^i} \right] \quad (3.3.39)$$

$p_B^{ik}$  = probabilidad de que la tasa generada por el modo B ( $\lambda_B$ ) esté en el intervalo  $\Delta\lambda_B^k$  cuando la tasa generada por el modo I ( $\lambda_I$ ) está en el intervalo  $\Delta\lambda_I^i$

$$p_B^{ik} = P \left[ \frac{\lambda_B \in \Delta\lambda_B^k}{\lambda_I \in \Delta\lambda_I^i} \right] \quad (3.3.40)$$

$S_m$  = al número de minifuentes del modelo MMFP binomial asociado a cada modo de codificación.

$\pi_P$  y  $\pi_B$  = matrices cuyos elementos se obtienen a partir de las  $p_P^{ij}$  y  $p_B^{ik}$  definidas anteriormente. Las matrices serán, con  $m = \{P, B\}$ , de la forma:

$$\pi_m = \begin{pmatrix} p_m^{00} & p_m^{01} & \dots & p_m^{0S_m} \\ p_m^{10} & p_m^{11} & & \cdot \\ \cdot & & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \\ p_m^{S_1 0} & & \dots & p_m^{S_1 S_m} \end{pmatrix} \quad (3.3.41)$$

Estas matrices estarán compuestas por  $S_m+1$  columnas y  $S_i+1$  filas. Obviamente, la suma de cualquiera de sus filas será 1 por la definición de sus elementos.

A través de las matrices  $\pi_P$  y  $\pi_B$  se establece la relación entre las variables aleatorias, asociadas a los modos de codificación P y B, cuando se producen cambios en la actividad generada en los cuadros I. El mecanismo empleado para capturar la relación de las tasas generadas en las transiciones entre niveles de actividad se basa en forzar el cambio de estado en los modelos M/M/ $\infty$ //S de los modos P y B cuando se produce una transición de estado en el modelo M/M/ $\infty$ //S del modo I. El cambio de estado en los modelos de P y B se calcula según las probabilidades  $P_P^{ij}$  y  $P_B^{ik}$ . De forma que, si el cambio de estado lleva al modelo de las I al estado k se emplearán las filas k-ésimas de  $\pi_P$  y  $\pi_B$  para determinar el nuevo estado de generación de los modelos de P y B. Obsérvese que, mientras el modelo del modo I no cambie de estado, los modelos P y B funcionan libremente, de forma que, tenderán probablemente hacia sus valores medios, de la misma forma que lo hacen las series temporales observadas en la figura 2.3.1.

Modelo ARMA	I	P	B	Modelo Binomial	I	P	B
p	1	1	1	Amin	2307.07	1090.3	1110.4
q	0	2	3	Paso A	7780.93	10036	6742.3
b(0)	0.9368	0.9400	0.9634	Número de minifuentes	11	7	8
a(1)	x	-0.392	-.2785	a	0.6019	1.0518	0.4816
a(2)	x	-.0155	-.1059	b	1.0069	2.56	2.37
a(3)	x	x	-.0905	Escala Temporal	0.04	0.04	0.04
media gaussiana	3606.0	2798.8	1248.8				
dev. std. gaussiana	4430.8	6586.2	3400.4				
				Modelo DAR(1)	I	P	B
				Probabilidad (p)	0.73618	0.386	0.705
<b>Resultados</b>	<b>Máx.</b>	<b>Media</b>	<b>F. Ráf</b>	N. Fuentes Geometricas	80	9	21
ARMA	120911	30187	4.00	Distancia Max (Kmax)	58	38	30
Binomial	110968	30359	3.65	Indice autocorrelacion	0.93767	0.8655	0.8922
Codificación	110968	30370	3.65	Proporcionalidad	2000.0	2000.0	2000.0
DAR(1)	118000	30280	3.89				

Tabla 3.3.1. Parámetros característicos de los modelos de vídeo MPEG VBR a nivel de cuadro para la secuencia "Live in Central Park" codificada con (Q=9, M=2, N=4)

### 3.3.5. Comparación de modelos de vídeo MPEG VBR a nivel de cuadro.

El estudio de la bondad del ajuste de los modelos al tráfico real se ha llevado a cabo a partir del análisis de la función de autocovarianza y la función de distribución de probabilidad. Las funciones de autocorrelación para las tasas generadas por el codificador y para los cuatro modelos se presentan en la figura 3.3.5. Se observa que los cuatro modelos capturan la relación temporal de forma precisa, aunque, cabe destacar que, el modelo binomial condicional se aproxima más a la función de autocorrelación de los datos reales.

La función de distribución de probabilidad se presenta en la figura 3.3.6 para los modelos y para los datos reales. Se observa que, en todos los casos, los modelos ajustan con precisión la función de distribución de probabilidad de los datos reales. Los modelos binomiales revelan su factor de discretización dado que presenta transiciones

escalonadas, mientras que este efecto pasa inadvertido para el modelo DAR(1), ya que dispone de un paso de discretización mucho más fino.

Además del análisis del ajuste del comportamiento estadístico es necesario contemplar si los modelos capturan las características de transmisión del codificador. Para este estudio se puede observar en la tabla 3.3.1 que todos los modelos presentan, aproximadamente, las mismas tasas medias y de pico, con lo que aparentemente dispondrían de los mismos descriptores de tráfico que el generado por el codificador MPEG. Un análisis más detallado, basado en los parámetros descriptores de tráfico [LucNet94], puede mostrar la bondad del ajuste de los modelos desde el punto de vista de transmisión.

A partir del algoritmo GCRA (I, L) se puede establecer si un modelo se comporta de forma parecida al tráfico real. En este caso, se ha recurrido a las curvas denominadas "Leaky Bucket Contours". Estas curvas relacionan, para una probabilidad de descarte del mecanismo de policía, cuál es el valor de la longitud del bucket cuando se fija la tasa de llegada de créditos al bucket. La tasa de llegada de créditos se expresa de forma normalizada respecto a la tasa media del tráfico generado.

Para llevar a cabo este análisis se han desarrollado las "Leaky Bucket Contours" con una probabilidad de pérdida de  $10^{-2}$ . Los resultados obtenidos quedan reflejados en la figura 3.3.7, donde, la longitud del bucket se ha expresado logarítmicamente, dada la fuerte variación que sufre cuando se incrementa la tasa de llegada de créditos del mecanismo de policía. Se puede observar que, para valores de la tasa normalizada de generación de créditos superiores a 1/3 todos los modelos se ajustan bien a los datos reales, siendo el binomial condicionado una aproximación pesimista. Para valores de generación de créditos inferiores a 1/3 los modelos binomiales capturan mucho mejor el comportamiento del tráfico real. Esto se debe a que los modelos binomiales se mantienen en estados altos de generación tanto tiempo como lo hace el tráfico real. Es decir, los modelos binomiales capturan el efecto de ráfaga de la fuente mucho mejor que los modelos DAR(1) y ARMA. Estos, si bien alcanzan las mismas tasas de generación máximas, no las mantienen de forma suficientemente sostenida.

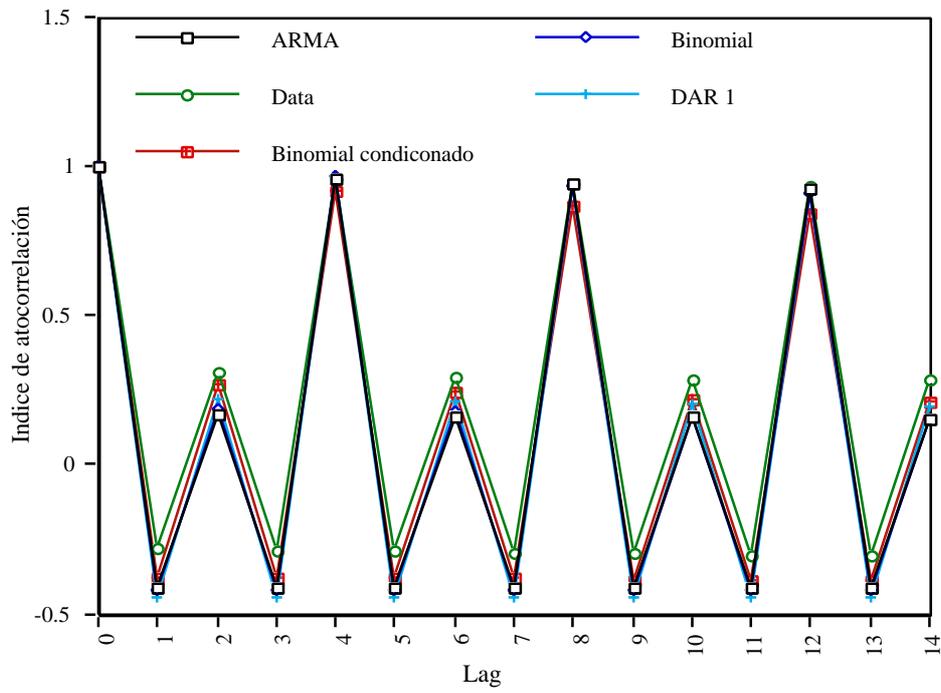


Figura 3.3.5. Comparación de la función de autocovarianza del tráfico real y los modelos

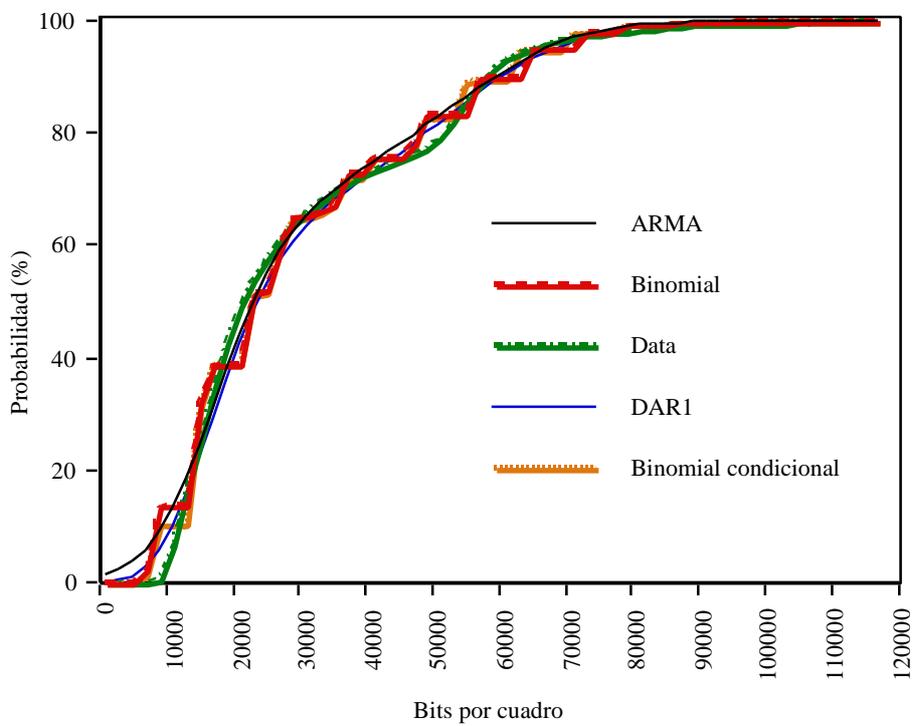


Figura 3.3.6. Comparación de la función de distribución del tráfico real y los modelos

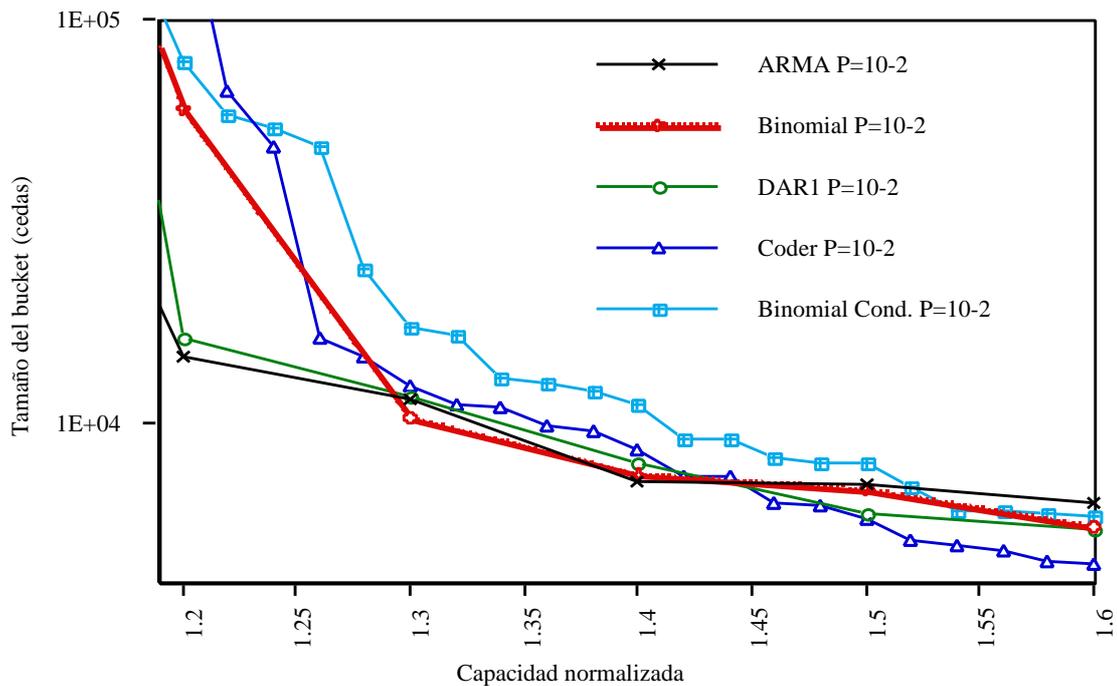


Figura 3.3.7. Comparación del tráfico generado por la codificación real y los modelos mediante los "GCRA contours"

### 3.4. Modelo de tráfico de vídeo MPEG VBR a nivel de GoP

En la transmisión de vídeo MPEG VBR es muy aconsejable el empleo de técnicas de suavizado que reduzcan el efecto de rafagueo de la estructura de codificación del algoritmo MPEG. El suavizado del tráfico se puede realizar a través del almacenamiento de la información en un buffer, intercalado entre el codificador y el interfaz de usuario [RosFra94]. Dependiendo de los requerimientos temporales del servicio, se puede llegar a almacenar toda la información correspondiente a un GoP, entregándola posteriormente a la red a una tasa constante durante un intervalo de duración igual al del GoP. Para servicios con requerimientos temporales más restrictivos, se hace imprescindible la reducción del tiempo de almacenamiento, en el orden de 80 ms, por lo que es necesario el empleo de técnicas de predicción. Las técnicas de predicción hacen compatible la reducción del factor de ráfaga de la fuente de tráfico y el cumplimiento de las restricciones temporales.

Puesto que las fuentes de tráfico emplearán técnicas de conformación, desde el punto de vista de la red, parece muy adecuado obtener un modelo que capture las características del tráfico suavizado, ya que, este tráfico será realmente el tráfico entregado a la red por gran cantidad de servicios de vídeo.

Por la estructura de codificación MPEG, el intervalo más adecuado para suavizar el tráfico es la duración de un GoP, cuando se emplea una codificación intra no distribuida. En los modos de funcionamiento con codificación intra distribuida basta

con realizar un suavizado durante un intervalo SGoP, con el coste de una mayor variabilidad, aunque, con la ventaja de una sustancial reducción del retardo.

Para el modelado del tráfico suavizado se ha desarrollado un modelo MMFP válido para los mecanismos de funcionamiento del codificador con I distribuidas o no, ya que, como se observará, el paso de discretización es mayor que las posibles variaciones de tasa que podrían surgir cuando se realiza una conformación en tiempo de SGoP respecto a la suavización en intervalos de GoP. Se ha elegido un modelo del tipo MMFP dada su tratabilidad analítica y por los resultados comparativos obtenidos en la sección anterior.

Los modelos clásicos MMFP binomiales presentan una función de autocovarianza con decaimiento exponencial y una función de distribución binomial. Tras el análisis de la serie temporal del número de bits generados por GoP, para la secuencia "Live in Central Park" de aproximadamente 42000 imágenes (7000 GoPs) de duración, se observa que, la función de autocovarianza presenta un decaimiento aproximadamente exponencial sólo hasta decalados de unas pocas decenas de cuadros, mientras que, para valores superiores, el decaimiento es mucho más lento que el exponencial. La función de densidad de probabilidad muestra a su vez una forma de campana aproximable con una distribución binomial.

Como primera aproximación al modelado del tráfico generado a nivel de GoP, se ha propuesto un modelo MMFP binomial, ajustado de la misma forma que se presentó en la sección 3.3. Como se puede observar la figura 3.4.1, la autocovarianza, si bien queda ajustada perfectamente para decalados de algunos tiempos de GoP, la dependencia temporal a largo plazo del tráfico real no es capturada por el modelo. Asimismo, la función de densidad de probabilidad, ilustrada a través del histograma de la figura 3.4.2, captura el comportamiento para los valores más probables, aunque, la probabilidad de los valores más altos es mucho más reducida en el modelo. Los parámetros que caracterizan el modelo  $M/M/\infty/S$  se presenta en la tabla 3.4.1.

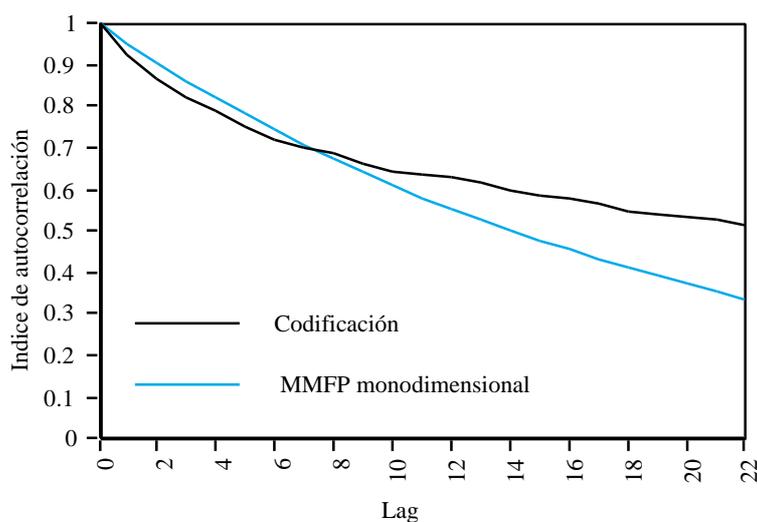


Figura 3.4.1. Función de autocovarianza de los datos y el modelo  $M/M/\infty/S$

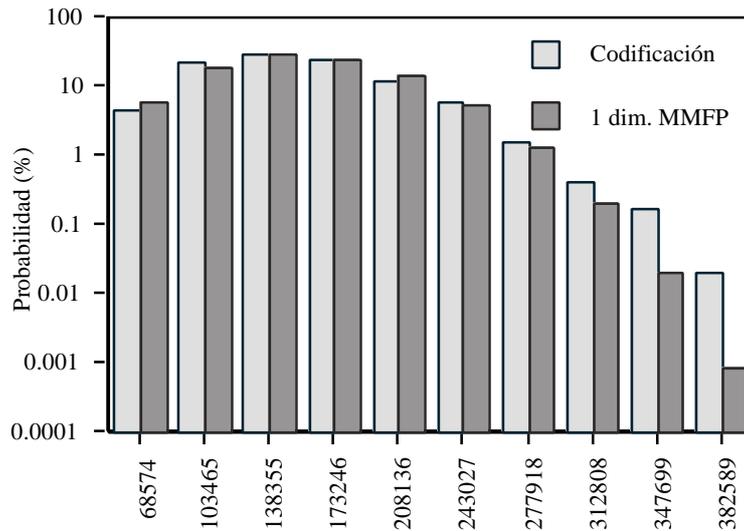


Figura 3.4.2. Histograma en escala logarítmica de las realizaciones del tráfico real y del modelo  $M/M/\infty/S$

S	A1	A	$\alpha$	$\beta$
9	7619.38	34890.51	0.0338	0.0903

Tabla 3.4.1. Parámetros característicos de las minifuentes del modelo MMFP unidimensional, siendo  $a$  y  $b$  las tasas de nacimiento y muerte por segundo

Con el fin de mejorar el modelo MMFP propuesto se introduce una generalización del modelo  $M/M/\infty/S$ . La estructura del modelo se presenta en la figura 3.4.3. Esta estructura bidimensional es interpretable como la agregación de dos tipos básicos de minifuentes, cuyas características se muestran en la figura 3.4.4. Por construcción, el número de minifuentes de cada tipo, que componen el modelo bidimensional, serían  $S_1$  y  $S_2$ , respectivamente.

El modelo MMFP bidimensional presentado, se ha denotado modelo  $M/M/\infty/S_1+S_2$ . Este modelo fue propuesto en [SenMag89], aunque, los resultados obtenidos no son aplicables en el presente trabajo, dado que, el enfoque realizado fue distinto al presente y, por consiguiente, tanto las aproximaciones planteadas como el mecanismo de ajuste expuesto no son válidos.

Para caracterizar el modelo  $M/M/\infty/S_1+S_2$  se ha considerado que las minifuentes de tipo 2 tienen una tasa de generación en estado OFF nula, es decir  $J_1=0$ , con lo que

$$J = J_m - J_1 = J_m \tag{3.4.1}$$

Esta consideración no produce ninguna pérdida de generalidad, ya que, el valor constante mínimo puede concentrarse en un tipo de minifuentes o distribuirse entre los dos tipos, obteniendo, en ambos casos, el mismo modelo MMFP exactamente. El análisis del modelo  $M/M/\infty/S_1+S_2$  se puede llevar a cabo en base a los resultados

obtenidos para el modelo binomial  $M/M/\infty/S$ . Si se considera que cada conjunto de minifuentes del tipo 1 o 2 genera una variable aleatoria binomial  $M/M/\infty/S_1$  y  $M/M/\infty/S_2$ , y ambas variables aleatorias son independientes, las características estadísticas del modelo  $M/M/\infty/S_1+S_2$  se obtendrán de la variable aleatoria obtenida por la adición de las variables correspondientes a cada conjunto de minifuentes.

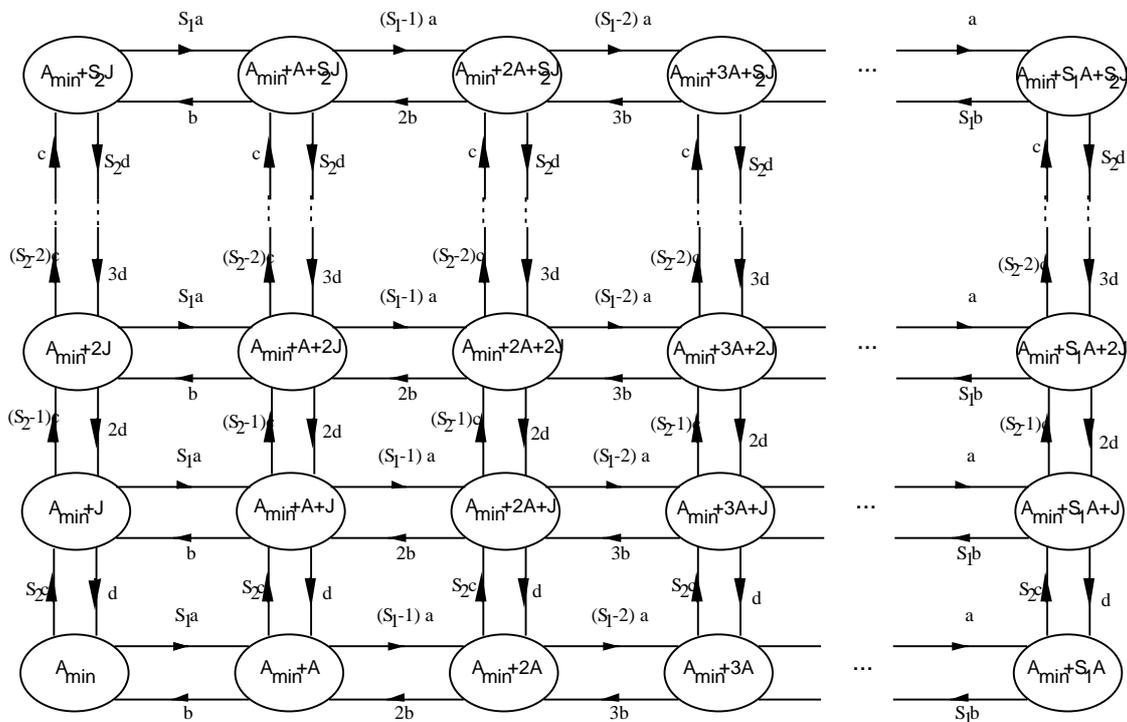


Figura 3.4.3. Modelo MMFP bidimensional

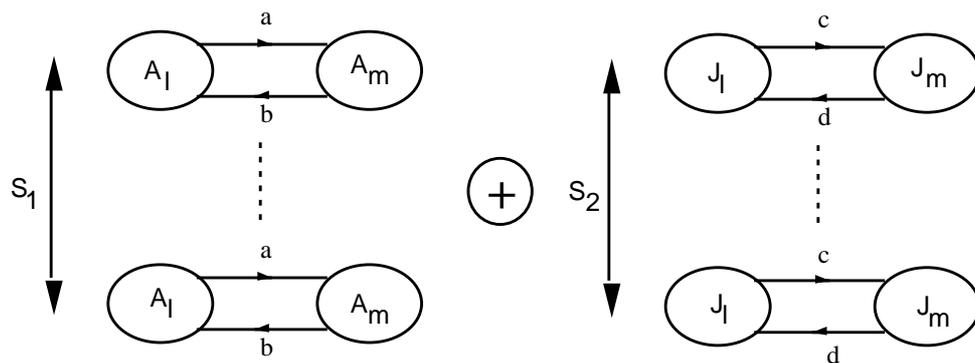


Figura 3.4.4. Descomposición del modelo MMFP bidimensional en fuentes simples

Si definimos:

$X$  = variable aleatoria asociada al modelo  $M/M/\infty/S_1+S_2$

$Y$  = variable aleatoria asociada al modelo  $M/M/\infty/S_1$

$Z$  = variable aleatoria asociada al modelo  $M/M/\infty/S_2$

entonces:

$$\begin{aligned}
 X &= Y + Z \\
 m_X &= m_Y + m_Z \\
 \sigma_X^2 &= \sigma_Y^2 + \sigma_Z^2 \\
 c_{XX}(\tau) &= c_{YY}(\tau) + c_{ZZ}(\tau)
 \end{aligned}
 \tag{3.4.2}$$

A partir de los resultados obtenidos para los modelos M/M/∞//S podemos expresar:

$$\begin{aligned}
 m_X &= S_1 p A + S_2 q J \\
 \sigma_X^2 &= S_1 p (1-p) A^2 + S_2 q (1-q) J^2 \\
 c_{XX}(\tau) &= S_1 p (1-p) A^2 e^{-\gamma_1 \tau} + S_2 q (1-q) J^2 e^{-\gamma_2 \tau}
 \end{aligned}
 \tag{3.4.3}$$

Dado que los parámetros observables son los correspondientes a la variable aleatoria X, se debe proceder, en primera instancia, al análisis de la función de autocovarianza para obtener los valores de  $\gamma_1$  y  $\gamma_2$ . De esta manera, se podrá determinar la relación de varianzas de las variables aleatorias Y y Z que dan lugar a la varianza de X.

Como se observó en el ajuste del modelo binomial M/M/∞//S para el tráfico a nivel de GoP, la captura de la autocovarianza para los primeros decalados era correcta pero la relación a largo término, prácticamente, era ignorada. En base a este comportamiento del modelo M/M/∞//S, se ha considerado separar las relaciones temporales a corto plazo y a largo plazo. Con este planteamiento se pueden sintetizar, individualmente, las relaciones temporales para cada uno de los tipos de fuentes. Así, las fuentes de tipo 2 se emplearán para la captura del comportamiento a largo plazo y las fuentes de tipo 1 para el comportamiento corto plazo, por lo que, se deberá cumplir la relación de los coeficientes de correlación

$$\gamma_1 \gg \gamma_2 > 0
 \tag{3.4.4}$$

Puesto que, para decalados de más de 100 intervalos de duración de un GoP, la influencia de la relación temporal a corto plazo es despreciable, se puede considerar que:

$$c_{XX}(\tau) \cong \sigma_Z^2 e^{-\gamma_2 \tau} \quad \tau > 100
 \tag{3.4.5}$$

Empleando la técnica de mínimos cuadrados sobre las ecuaciones de Yule-Walker modificadas para estimación de modelos autoregresivos, aplicada sobre la función de autocovarianza para valores de retardo entre 100 y 250 intervalos de duración un tiempo de GoP se obtiene, como se muestra la figura 3.4.5, un adecuado ajuste para el valor:

$$\gamma_2 = 0.02541 \text{ rad/sg}
 \tag{3.4.6}$$

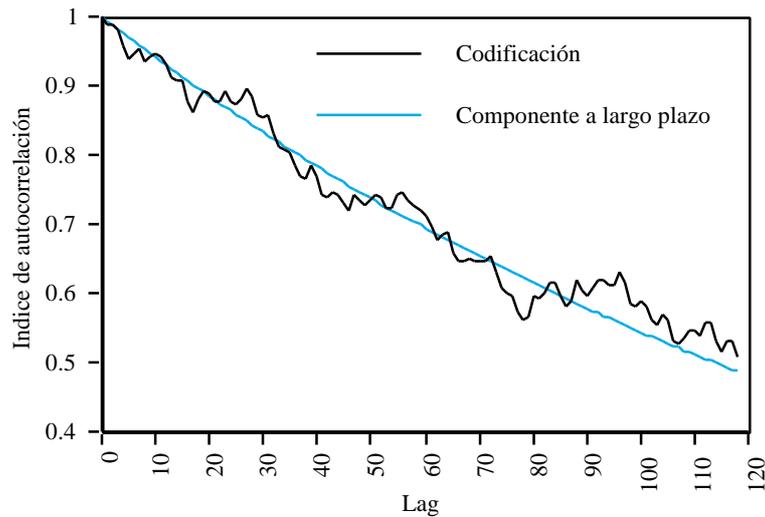


Figura 3.4.5. Autocovarianza de la serie temporal de los datos y de la componente a largo plazo del modelo  $M/M/\infty/S_1+S_2$

Una vez se ha determinado la componente de relación a largo término, se debe proceder a la obtención de la componente a corto plazo. En este caso, se debe realizar previamente, la extracción de la contribución a largo término de la función de autocovarianza de la serie temporal real. Debe tenerse en cuenta que se ha realizado la aproximación:

$$\sigma_Z^2 \cong \left. \frac{c_{XX}(\tau)}{e^{-\gamma_2 \tau}} \right|_{\tau=100} \quad (3.4.7)$$

con lo que

$$c_{YY}(\tau) \cong c_{XX}(\tau) - \sigma_Z^2 e^{-\gamma_2 \tau} \quad (3.4.8)$$

Aplicando la misma técnica de estimación que para la relación a largo plazo, se obtiene:

$$\gamma_1 = 0.5929 \text{ rad/sg} \quad \text{con } \sigma_z^2 = 1.217 \cdot 10^9 \quad (3.4.9)$$

con lo cual, se comprueba que la hipótesis  $\gamma_1 \gg \gamma_2 > 0$  era cierta. La bondad del ajuste se ilustra en la figura 3.4.6. Finalmente, la varianza de Y se halla directamente:

$$\sigma_Y^2 = \sigma_X^2 - \sigma_Z^2 = 9.63 \cdot 10^8 \quad (3.4.10)$$

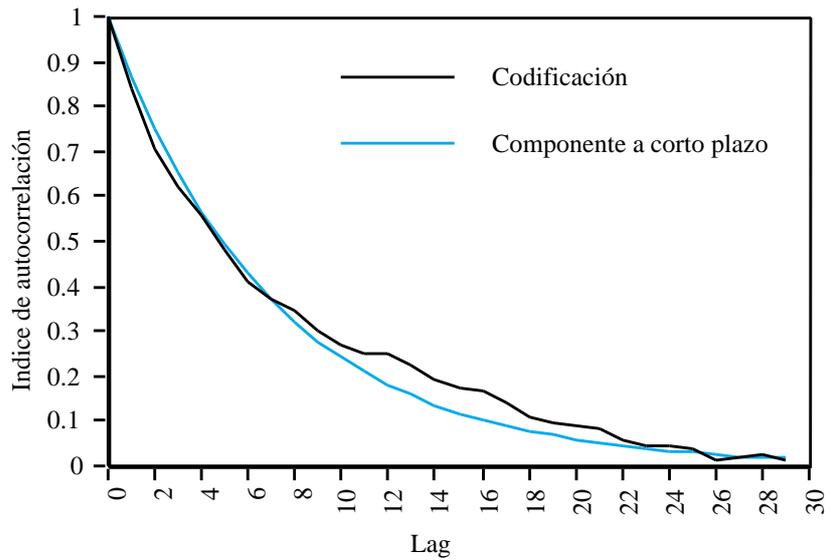


Figura 3.4.6. Autocovarianza de la serie temporal de los datos y de la componente a corto plazo del modelo  $M/M/\infty/S_1+S_2$

En la figura 3.4.7 se muestra el resultado final obtenido mediante la descomposición desarrollada. Se comprueba que el modelo  $M/M/\infty/S_1+S_2$  captura perfectamente la relación temporal que presentan los datos para las distintas escalas de tiempo, denominada persistencia de la generación [WilGar94], y que los modelos MMFP unidimensionales son incapaces de capturar dicho comportamiento.

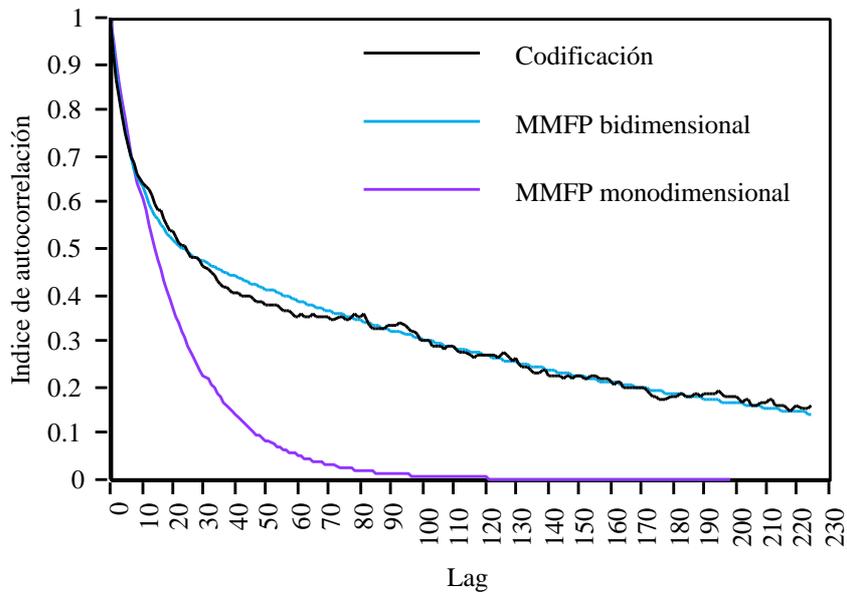


Figura 3.4.7. Autocovarianza de la serie temporal de los datos y de los modelos  $M/M/\infty/S$  y  $M/M/\infty/S_1+S_2$

El análisis de la serie temporal proporciona las restricciones necesarias para el planteamiento de un conjunto de ecuaciones. Estas ecuaciones determinan un número finito de soluciones de los parámetros que caracterizan el modelo. En este caso, a partir de los valores obtenidos de la serie temporal,

$$\begin{aligned} \min(X) &= 67527 \\ \max(X) &= 382589 \\ m_x = E[X] &= 154072.07 \\ \sigma_X^2 &= 2.17 \cdot 10^9 \end{aligned} \quad (3.4.11)$$

y los valores obtenidos a través de la descomposición de la función de autocovarianza:

$$\begin{aligned} \sigma_Y^2 &= 9.63 \cdot 10^8 \\ \sigma_Z^2 &= 1.217 \cdot 10^9 \end{aligned} \quad (3.4.12)$$

podemos determinar los parámetros del modelo MMFP bidimensional.

Puesto que el modelo MMFP sintetizado pretende capturar el comportamiento genérico del tráfico suavizado de las fuentes de vídeo MPEG VBR, se ha considerado que la tasa de pico obtenida puede ser superable con una probabilidad muy reducida por otras series temporales. Así, sobre el modelo resultante se pueden ampliar el conjunto de tasas binarias aumentando el número de minifuentes  $S_1$  o, en la propia síntesis del modelo, se puede forzar a que la tasa máxima generada no se corresponda con el estado de mayor tasa binaria. En este caso, las ecuaciones planteadas fuerzan el estado de máxima generación al valor máximo alcanzado por la serie temporal. Por tanto, el conjunto de ecuaciones que determinan los parámetros del modelo será:

$$\begin{aligned} J &= \frac{A_{\max} - S_1 A - A_{\min}}{S_2} \\ S_1 p A + S_2 q J &= m - A_{\min} \\ S_1 p (1 - p) A^2 &= \sigma_y^2 \\ S_2 q (1 - q) J^2 &= \sigma_z^2 \end{aligned} \quad (3.4.14)$$

Obsérvese que es un sistema indeterminado. En este caso se han hallado todas las soluciones para los distintos enteros que cumplen valores  $S_1 > 0$  y  $S_2 > 0$ . Aplicando una resolución por métodos numéricos, el sistema sólo presenta solución para valores que cumplen  $S_1 + S_2 \leq 13$ . Analizando los valores obtenidos de  $p$  y  $q$  (con  $p > q$ ), presentados en la tabla 3.4.2, se observa que estos no varían sustancialmente cuando se incrementa el valor de  $S_2$ , lo cual revela que el aumento de  $S_2$  sólo lleva a realizar una ampliación de los niveles alcanzables. Estos niveles se visitan con probabilidades extraordinariamente pequeñas ( $\approx 10^{-10}$ ) cuando  $S_2 = 3$ .

La síntesis de un modelo realista, no sobredimensionado, nos queda fijada a la consideración de los casos  $S_2 = 1$  y  $S_2 = 2$ . Las soluciones obtenidas para el caso  $S_2 = 1$

tampoco son válidas, dado que, para el conjunto de tasas sintetizadas existe una fuerte transición entre la máxima tasa generada, cuando la minifuerza de tipo 2 está inactiva, y la mínima generada, cuando la minifuerza está activa. Este modelo provocaría la no consideración de un conjunto de tasas de generación intermedias, por lo que, el modelo presentaría una deficiencia en la síntesis de estas tasas. Este efecto se podría interpretar, en algunos casos, como una probabilidad de generación nula de tasas intermedias, lo cual se aleja de comportamiento real. En la figura 3.4.8 se pone de manifiesto el mínimo relativo aparecido en la función de distribución del modelo MMFP bidimensional con  $S_2=1$  y  $S_1=7$ .

$S_1$	$S_2$	p	q	A	J
4	1	0.61	0.03	31912	186366
5	1	0.56	0.04	27953	174246
6	1	0.51	0.05	25341	161963
7	1	0.46	0.06	23508	149457
8	1	0.42	0.07	22173	136628
9	1	0.39	0.09	21191	123288
10	1	0.35	0.11	20501	109000
11	1	0.31	0.17	20171	92128
3	2	0.64	0.064	37463	100812
4	2	0.57	0.073	31346	94314
5	2	0.50	0.09	27757	87624
6	2	0.45	0.10	25468	80601
7	2	0.39	0.13	23994	73028
8	2	0.33	0.18	23246	64021

*Tabla 3.4.2. Parámetros característicos de las minifuerzas del los modelos MMFP bidimensionales*

El valor más apropiado de  $S_2$  que captura bien las características de generación del tráfico real, es  $S_2=2$ . Una vez determinado el número de minifuerzas de tipo 2, se debe estudiar el valor de  $S_1$ , de forma que, no se produzca sobreestimación de estados, es decir, no se tengan en cuenta estados con probabilidad de visita prácticamente despreciable. En este caso se obtiene que, para valores de  $S_1 > 6$ , las probabilidades de generación sintetizadas para los estados superiores son muy reducidas sin disminuir apreciablemente el paso de discretización, por lo que, es suficiente con considerar  $S_1=5$  o  $S_1=6$ .

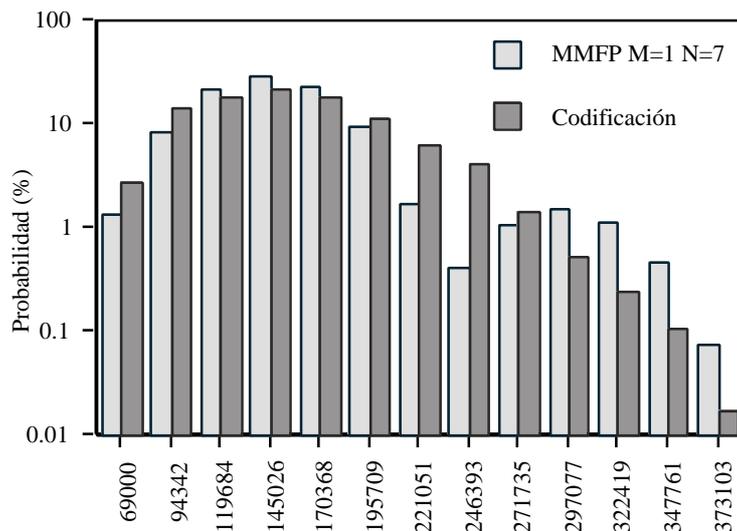


Figura 3.4.8. Histograma en escala logarítmica de las realizaciones del tráfico real y del modelo  $M/M/\infty/S_1+S_2$

Para el ajuste final del modelo MMFP conviene estudiar cuál es el valor mínimo sintetizable por el modelo, de forma que, la función de densidad de probabilidad quede bien ajustada en sus valores bajos. El resultado obtenido finalmente se presenta en la tabla 3.4.3.

Modelo	Tipo	núm.	$A_1$	A	$\alpha$	$\beta$
$S_1=5$ $S_2=2$	1	5	13710	27757	0.3	0.2929
	2	2	0	87614	0.0022	0.02321
$S_1=6$ $S_2=2$	1	6	11500	25468.6	0.2663	0.3266
	2	2	0	80601.47	0.00266	0.002275

Tabla 3.4.3. Parámetros característicos de las minifuentes del modelo MMFP bidimensional

Obsérvese que, por la propia estructura del modelo bidimensional sintetizado, los estados de generación del conjunto de minifuentes de tipo 1 para un mismo nivel de actividad de las fuentes de tipo 2, definen, de forma natural, tres niveles de actividad de la fuente de vídeo. Este hecho queda ilustrado en la figura 3.4.9. El tiempo de permanencia en estos niveles captura la relación temporal a largo término de la generación y viene condicionado por las tasa de transición de las minifuentes de tipo 2. A su vez, la transición entre niveles de actividad simula los posibles cambios significativos de actividad que se producen en una secuencia de vídeo de este tipo. Asimismo, con la estructura del modelo planteada, se comprueba que la función de densidad de probabilidad del modelo se ajusta de forma más precisa al tráfico real. En este caso, los valores altos se visitan mucho más frecuentemente que en el modelo

binomial  $M/M/\infty/S$ . Por construcción, la función de densidad de probabilidad binomial de las fuentes tipo 1 se adiciona ponderada y desplazadamente para cada nivel de actividad, como se ilustra en la figura 3.4.10, para los distintos niveles de actividad, denotados como:

- i) Nivel 0 o nivel de gran actividad.
- ii) Nivel 1 o nivel de actividad elevada.
- iii) Nivel 2 o nivel de actividad regular.

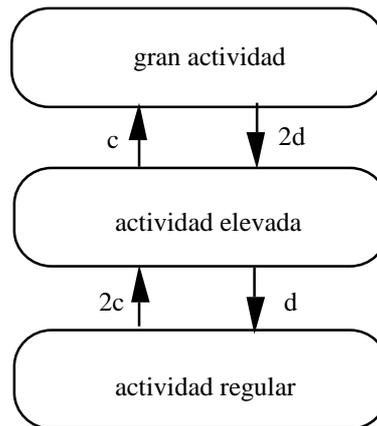


Figura 3.4.9. Niveles de actividad asociados a la generación del tráfico de vídeo MPEG VBR, derivados del modelo MMFP bidimensional

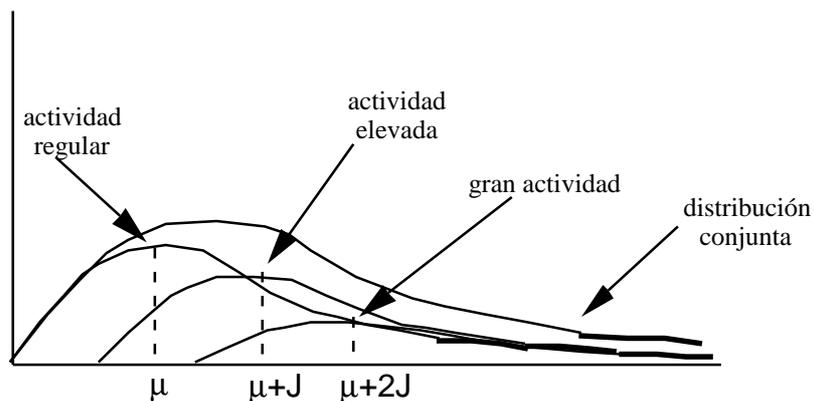


Figura 3.4.10. Obtención de la función de densidad del modelo  $M/M/\infty/S_1+S_2$  por adición de funciones de densidad de cada nivel de actividad

En las figuras 3.4.11 y 3.4.12 se muestran en escala lineal y logarítmica los histogramas de la serie temporal y del modelo. Para establecer una comparación representativa, se ha confeccionado el histograma de forma que las barras tienen la anchura igual al paso de discretización  $A$  y están centradas a partir de los valores de generación de los estados del nivel de actividad regular.

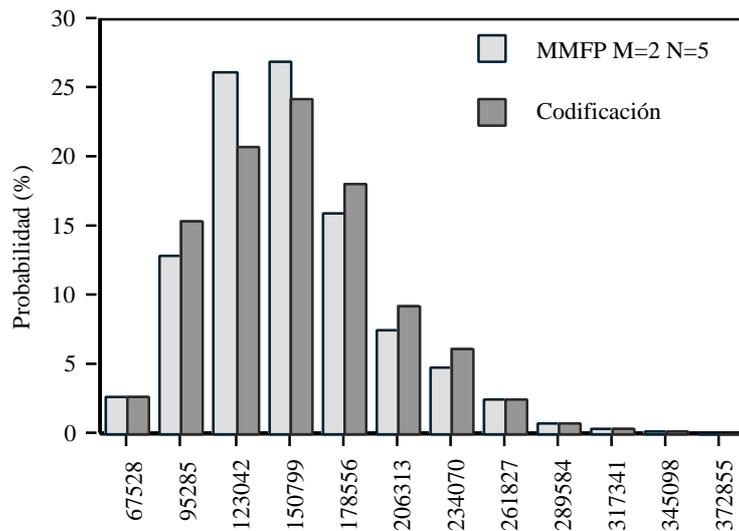


Figura 3.4.11. Histograma en escala lineal de las realizaciones del tráfico real y del modelo  $M/M/\infty/S_1+S_2$

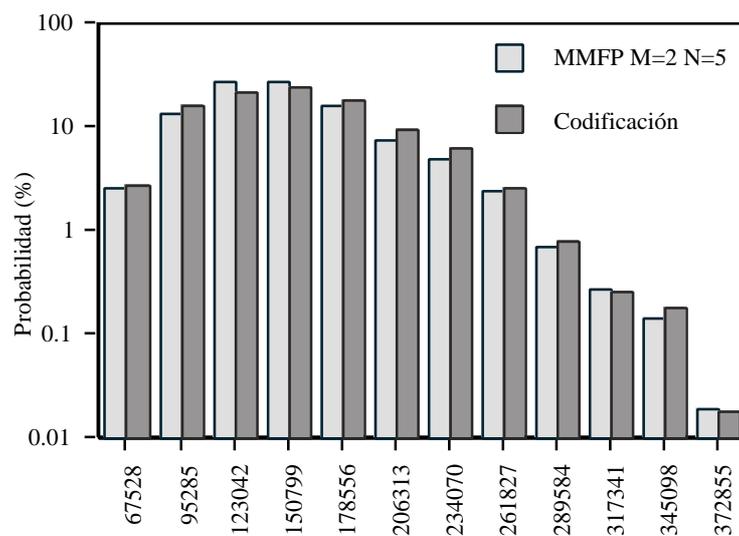


Figura 3.4.12. Histograma en escala logarítmica de las realizaciones del tráfico real y del modelo  $M/M/\infty/S_1+S_2$

Finalmente, para mostrar la generalidad del procedimiento aplicado en la caracterización del tráfico de vídeo MPEG VBR, en las figuras 4.3.13 y 4.3.14 se muestran los resultados obtenidos con esta técnica para la serie temporal "Jurassic Park". Esta serie temporal ha sido codificada con parámetros ( $Q=6, M=2, N=6$ ) y dispone de 175056 cuadros (29166 GoPs). En este caso, la síntesis realizada se ha llevado a cabo con las consideraciones de un retardo de hasta 800 intervalos de GoP en la función de

autocovarianza y no se ha fijado el valor máximo de tasa binaria de la serie temporal al estado de actividad de todas las minifuentes.

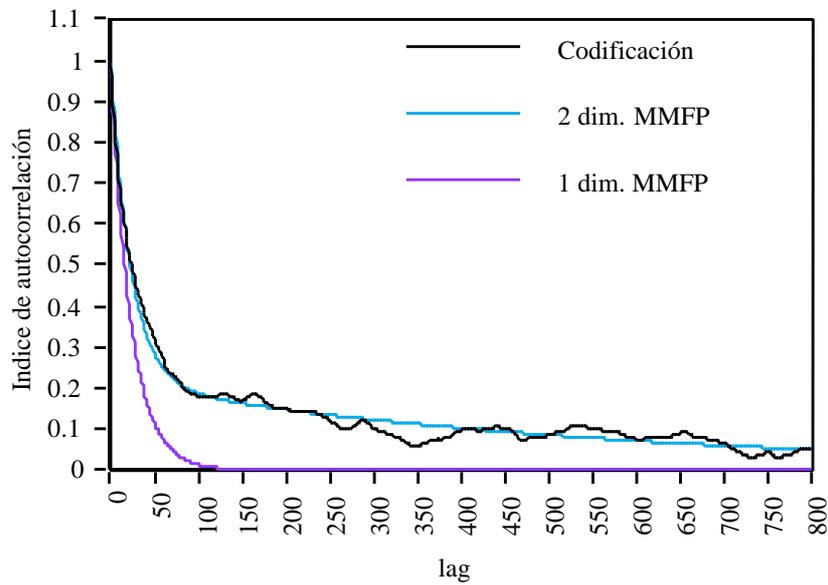


Figura 3.4.13. Autocovarianza de la serie temporal de los datos y de los modelos  $M/M/\infty/S$  y  $M/M/\infty/S_1+S_2$

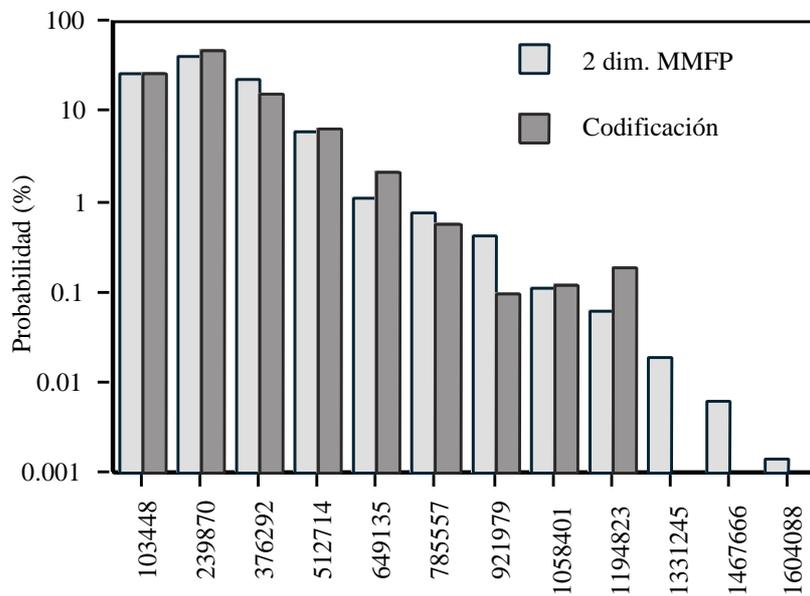


Figura 3.4.14. Histograma en escala logarítmica de las realizaciones del tráfico real y del modelo  $M/M/\infty/S_1+S_2$

### 3.5. Suavizado y Predicción 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 2 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.

Si se observan las serie temporales del tráfico MPEG-I, se aprecia una variación lenta del valor medio de la tasa dependiendo de la actividad y complejidad de la escena. Este efecto a largo término dificulta la elaboración de predictores, dada la aparente no estacionariedad de la media de la serie temporal. Para sintetizar un buen predictor es necesario capturar el efecto a largo plazo.

En este trabajo se ha desarrollado un nuevo predictor basado en la síntesis de un modelo autoregresivo integrativo de media móvil (ARIMA) del tráfico VBR MPEG. Se han utilizado tres secuencias para elaborar y evaluar el predictor, "Live in Central Park", "Jurassic Park" y "Geografía de Catalunya", de 33000, 175000 y 51000 imágenes respectivamente. Estas secuencias han sido codificadas con parámetros  $(Q=9, M=2, N=6)$ ,  $(Q=6, M=2, N=6)$  y  $(Q=9, M=2, N=6)$ . 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 será la salida de un filtro FIR ( $s_n$ ), 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. 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 refleja en la función de autocovarianza de la serie temporal, como se ilustra en la figura 3.5.1. 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 las 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:

$$AR(z) = (1 - 0.6950 z^{-6} - 0.3 z^{-12})^{-1} \quad (3.5.1)$$

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:

$$\begin{aligned} MA(z) = & 1 - 0.7618 z^{-1} + 0.1136 z^{-2} - 0.1676 z^{-3} + 0.0195 z^{-4} - 0.0451 z^{-5} \\ & - 0.1691 z^{-6} + 0.0386 z^{-7} + 0.0397 z^{-8} - 0.0268 z^{-9} + 0.0523 z^{-10} \\ & - 0.0371 z^{-11} - 0.1910 z^{-12} + 0.1351 z^{-13} \end{aligned} \quad (3.5.2)$$

Finalmente, el modelo ARIMA tendrá como función de transferencia:

$$H(z) = MA(z) AR(z) (1-z^{-1})^{-1} \quad (3.5.3)$$

Para evaluar el comportamiento del predictor sintetizado se ha realizado un análisis de los errores de predicción para todas las secuencias. En la figura 3.5.2 se presenta la autocorrelación de los residuos y los intervalos de confianza para el 99%.

Se puede observar que el modelo ARIMA es adecuado para la predicción. Este modelo no se puede emplear para sintetizar tráfico MPEG VBR dado que la generación de serie temporal tiene un valor medio variante no acotado. Para explotar el efecto de estacionariedad, una vez extraído el comportamiento a largo término, se propone como trabajo futuro emplear una parte integrativa del tipo  $(1-z^{-1})^H$ , dado que, las secuencias integrativas de este tipo presentan un comportamiento de tráfico autosemejante. El desarrollo de este método integrativo puede realizarse como se muestra en [Garr93].

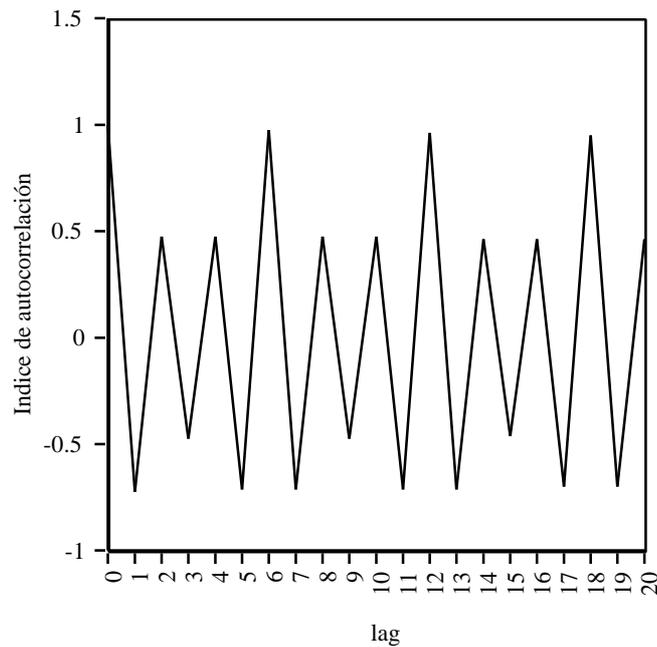


Figura 3.5.1. Índice de autocorrelación de la serie temporal "Live in Central Park" una vez extraída la parte integrativa codificada con  $(Q=9, M=2, N=6)$

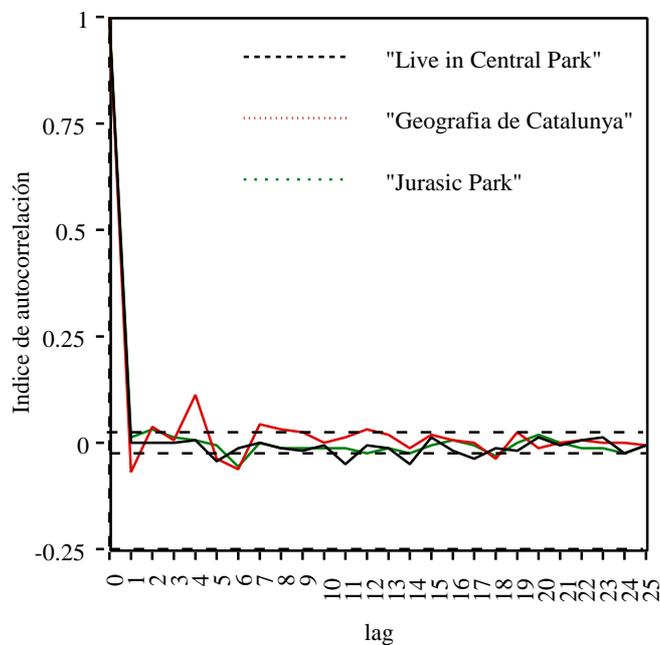


Figura 3.5.2. Índice de autocorrelación de las series residuales de la predicción con el filtro ARIMA de las series temporales de la tasa binaria por cuadro de las secuencias codificadas MPEG VBR

### 3.6. Conclusiones

El trabajo presentado en este capítulo se centra en la caracterización del tráfico generado por un codificador de vídeo MPEG VBR. Se han presentado las distintas técnicas de modelado en ATM y se han considerado aquellas que se ajustan mejor al comportamiento del tráfico de vídeo MPEG VBR. Se han sintetizado cuatro modelos diferentes para el tráfico generado por el codificador durante un intervalo de cuadro. Estos modelos se basan en un nuevo planteamiento de la síntesis de tráfico para cada modo de codificación (I, P o B) del algoritmo MPEG. La generación artificial del tráfico se ha realizado visitando los modelos de cada modo de codificación de igual manera a la fijada con los parámetros N y M en el codificador. Las técnicas de modelado empleadas se han basado en los procesos ARMA, DAR y MMFP. Sobre el modelo MMFP se ha introducido una variante, también analizada, en la cual se han tenido en cuenta las relaciones de las tasas binarias generadas en los modelos de codificación P y B con la tasa del modelo I. Esta nueva mejora se ha denominado modelo de fluidos condicional. La comparación de los modelos hallados ha revelado que los modelos de fluidos capturan mejor el comportamiento del tráfico real de vídeo, puesto que, aproximan muy bien no sólo las funciones de autocovarianza y de distribución, sino también los parámetros descriptores del tráfico MPEG VBR.

A partir de la consideración de que el tráfico generado por el codificador será suavizado antes de su entrega a la red, ya que las características periódicas de la codificación MPEG VBR reducen las ventajas de la multiplexación estadística, se ha desarrollado un nuevo modelo de tráfico para la tasa binaria generada en intervalos de duración de un GoP. Para la caracterización de este tráfico se ha empleado un modelo de fluidos. Se ha elegido este tipo de modelo por su tratabilidad matemática en el análisis de las prestaciones de la multiplexación estadística y por los resultados obtenidos en la comparación de modelos realizada anteriormente. Este modelo de fluidos se caracteriza por estar compuesto de dos tipos de minifuentes diferentes, lo cual da lugar a una estructura bidimensional de estados de la fuente. La estructura bidimensional de este modelo ha permitido ajustar perfectamente los histogramas obtenidos del tráfico real. La síntesis del modelo se ha desarrollado en base al comportamiento presentado por la función de autocovarianza. Se ha demostrado que esta función se puede descomponer en un término de decaimiento a corto plazo y otro a largo plazo. El decaimiento a largo plazo muestra el efecto de semejanza de los valores generados durante un intervalo de tiempo prolongado.

Finalmente, se ha propuesto un nuevo predictor de tráfico de vídeo MPEG VBR. Este predictor se basa en considerar la generación de tráfico como un proceso ARIMA. En este caso, la parte integrativa se ha mostrado especialmente adecuada para la síntesis de la dependencia a largo término, de forma que la predicción realizada a corto plazo es

muy precisa. La relación a corto plazo ha sido capturada por las partes autorregresivas y de media móvil. Cabe destacar que la relación temporal a corto plazo se mantiene invariante para todas las codificaciones generadas con los mismos parámetros de codificación. Asimismo, los valores de los coeficientes del predictor ARIMA se muestran insensibles al paso de cuantificación, por lo que se puede mantener el mismo predictor a lo largo de una codificación que varíe el paso de cuantificación.