

OPTIMIZACION DE LOS TAMAÑOS DE BUFFER EN UN SISTEMA DE COLAS CON CLIENTES DE MULTIPLES CLASES Y DISTINTOS REQUERIMIENTOS DE RETARDO Y PERDIDA

A. Santana, J.M. Quinteiro, C. Herrera y P. Saavedra, Universidad de Las Palmas de Gran Canaria
J. Gómez, Universidad Politécnica de Cataluña

RESUMEN

En este trabajo analizamos el rendimiento de una política de gestión de colas en un conmutador de red, que recibe dos clases de tráfico con distintos requisitos de calidad en lo que se refiere a retardos y pérdidas. Se muestra como pueden obtenerse, de forma recursiva, las probabilidades de estado de este sistema, y a partir de las mismas los retardos medios de ambas clases de tráfico, así como sus tasas de pérdida. Todas estas cantidades intervienen en el cálculo del rendimiento del sistema a través de una función objetivo que representa una ponderación del coste causado por los retardos producidos en el tráfico prioritario (en tiempo real) y las pérdidas originadas al tráfico no prioritario (en tiempo no real).

ABSTRACT

In this paper we analyze the yields of a queue management politic in a conmuter of a network, which receives two classes of traffic with different requisites of quality with respect to retards and losses. We prove how the state probabilities of the system can be obtained recursively and from the mean retards of arrivals of both traffic classes as well as their loss rates. These quantities are linked with the calculus of the yields of the system by means of the objective function that represents a weighting of the associated cost produced by the retards produced by the priority traffic (in real time) and the losses originated by the non priority traffic (in non real time).

1. MOTIVACION Y PRESENTACION DEL MODELO

Las actuales redes de comunicaciones de banda ancha están diseñadas para soportar y acomodar, de manera flexible y eficiente, una amplia variedad de servicios tales como voz, vídeo, datos y sus combinaciones multimedia. Para cumplir con estos objetivos se han diseñado diversos estándares de red, entre los que cabe destacar ATM como uno de los más importantes. En este estándar la información que viaja por la red se organiza en pequeños paquetes de longitud fija llamados *células*. Los diferentes tipos de aplicaciones que hacen uso de la red pueden variar en sus requerimientos de servicio. Así, por ejemplo, las aplicaciones en tiempo real, tales como la videoconferencia, requieren prestaciones extremas en términos de caudal del tráfico (throughput), de retardo, de la variación del retardo (delay jitter, jitter o varianza en el retardo, Figuera-Pasquale, (1995) y de la tasa de pérdidas. La cada vez mayor generalización de estas aplicaciones ha convertido en urgente la necesidad de proporcionar servicios de red con prestaciones garantizadas y desarrollar los algoritmos que soporten estos servicios, Zhang (1995).

Uno de los mecanismos fundamentales para poder proporcionar servicios con prestaciones garantizadas es la elección de la disciplina de servicio de paquetes en el conmutador. En una red de conmutación de paquetes, los paquetes de distintas conexiones interactúan unos con otros en cada conmutador, y sin el control apropiado estas interacciones pueden llegar a afectar a las prestaciones de la red. La disciplina de servicio del nodo de conmutación controla el orden en el que los paquetes son servidos y determina cómo interactúan los paquetes de las distintas conexiones. Una forma de conseguir este objetivo es mediante la implementación, en los buffers del conmutador, de mecanismos de prioridad capaces de controlar el tiempo o espacio dedicado a cada célula. Debido a su simplicidad, rapidez y más bajo coste de implementación, en los conmutadores de alta velocidad se prefieren los mecanismos de prioridad que controlan el espacio (capacidad) disponible en el buffer, en los que deben tener en cuenta el instante de llegada de cada célula y controlar su tiempo de residencia en el conmutador.

En los últimos años, diferentes autores han propuesto una serie de nuevas disciplinas de servicio que proporcionan prestaciones garantizadas por conexión para redes de conmutación de paquetes de alta velocidad, Demers **et al.** (1989), Ferrari-Verna (1990), Golestani (1994), Kalmanek **et al** (1990),

Verma **et al.** (1991), Zhang-Ferrari (1993) y Zhang (1990), así como nuevas técnicas de análisis para el estudio de estas disciplinas, Banerjea-Keshav (1993), Chan (1994), Cruz (1991) y (1991a), Knightly **et al.** (1995), Kurose (1992), Low (1992), Parekh-Gallager (1993), Xie-Lam (1994), Zhang **et al.** (1993), Zhang-Ferrari (1994), Zhang-Knightly (1994), Zhang **et al.** (1994).

La interacción producida entre los distintos paquetes que llegan a un conmutador da lugar a una distorsión en la distribución del tráfico de salida, lo que puede llegar a ocasionar graves perjuicios para aquellos servicios que requieran cierto nivel de sincronía o regularidad en la transmisión de sus células. La única forma de restaurar el patrón de tráfico en el interior de las redes es empleando esquemas de trabajo no conservativos, Zhang (1995). Estos esquemas, si bien consiguen controlar la distorsión del patrón de tráfico en las conexiones, y por tanto son capaces de proporcionar prestaciones con variación del retardo limitado, tienen el inconveniente de que incrementan el retardo medio de los paquetes e infrutilizan la capacidad del canal de salida. Disciplinas de este tipo son Stop-and-Go, Golestani (1990), o HRR (Hierarchical Round Robin) Kalmanek **et al.** (1990) entre otras.

Cuando se utilizan esquemas de trabajo conservativo hay que recurrir a disciplinas basadas en marcas de tiempo ya que éstas tienen la flexibilidad de proporcionar los servicios requeridos. Ejemplos de este tipo ampliamente referenciados en la literatura son WFQ (Weighted Fair Queueing), Demmers **et al.** (1990) y Parekh-Gallager (1993), VirtualClock, Figueira-Pascuale (1995), Zhang (1991), o Dynamic Priority Queueing, Ren-Mark-Wong (1994).

En este trabajo se presenta un novedoso esquema de gestión de buffers basado en rondas, de trabajo conservativo y de tasa garantizada, que tiene en cuenta el patrón de llegada de dos clases de tráfico. En particular, hemos considerado que estas clases de tráfico corresponden, respectivamente, a tráfico en tiempo real (voz, video) y a tráfico en tiempo no real (datos). Los mecanismos basados en rondas consisten básicamente en alternar el servicio entre los distintos canales de entrada al conmutador. Existen múltiples políticas para efectuar las rondas. La más simple es atender una célula de cada canal, desde el primero al último, y vuelta a empezar desde el primer canal. Una variación de esta política consiste en asignar prioridades a los canales en función de la clase de tráfico, atendiendo en cada ronda más células en los canales de mayor prioridad que en los canales de prioridad más baja; por ejemplo, se podrían atender tres células de un canal de video por cada célula de un canal de datos. Otras políticas especifican umbrales: mientras en el canal A haya menos de n células esperando, se atiende al canal B; una vez que en A haya n células, el servicio se dedica exclusivamente a A hasta que en A queden menos de m células.

El mecanismo que presentamos incorpora a los buffers de entrada al conmutador propios de cada canal, un buffer adicional que es compartido por las dos clases de tráfico. Este sistema es gestionado por una política que, en cierta medida, tiene en cuenta los instantes de llegada de las células sin necesidad de utilizar marcas de tiempo que especifiquen dichos instantes. Esta política proporciona una reducción en los retardos de las células de alta prioridad (tiempo real), sin que por ello se vean afectadas considerablemente las prestaciones dadas al tráfico de baja prioridad (tiempo no real).

El diseño del sistema se muestra en la siguiente figura:

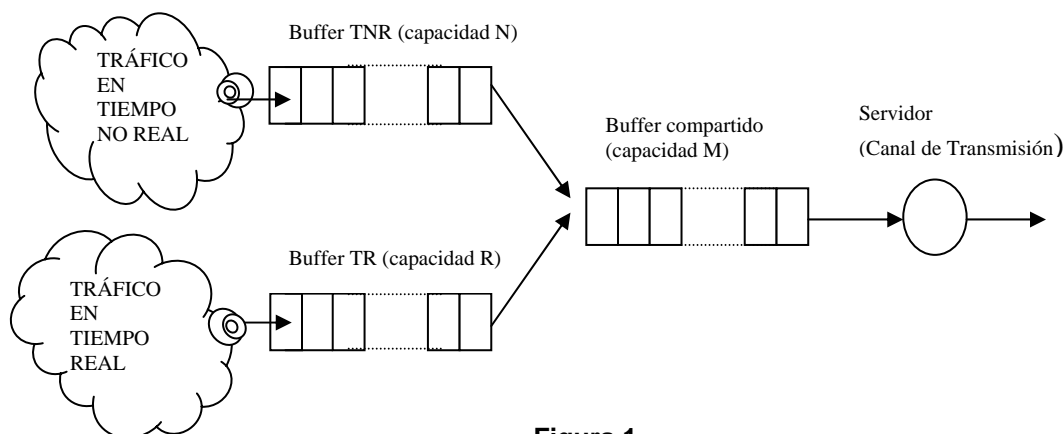


Figura 1.

El tráfico en tiempo real (TR) accede a su propio buffer, que tiene una capacidad limitada para R células. Asimismo, el tráfico en tiempo no real (TNR) accede a un segundo buffer con capacidad N . Las células que salen de ambos buffers se mezclan en un tercer buffer *compartido*, con capacidad para M células. La salida de este tercer buffer es el canal de transmisión, sobre el que se multiplexan ambos tráficos. La gestión del sistema se produce del siguiente modo:

1. En cada instante, en el buffer compartido puede haber sólo una célula TNR, y como máximo $M-1$ células TR.
2. Si no hay ninguna célula TNR en el sistema, la primera célula TNR que llegue al mismo pasa directa e instantáneamente al buffer compartido. Si hay ya una célula TNR en el buffer compartido, cualquier nueva célula TNR que llegue al sistema se incorpora al buffer TNR.
3. Cada vez que llega una célula TR, accede directamente al buffer compartido si en éste hay menos de $M-1$ células TR; en otro caso se incorpora el buffer TR.
4. Cada vez que una célula TR (TNR) es transmitida, la primera célula del buffer TR (TNR), si es que hay alguna, pasa instantáneamente a la cola del buffer compartido.
5. Cuando un buffer de entrada está lleno, las células que lleguen al mismo son rechazadas.

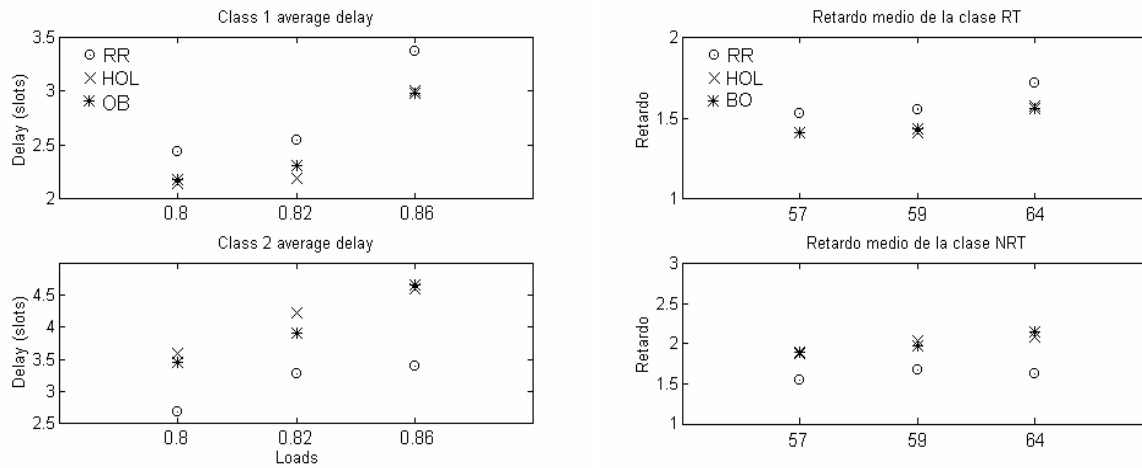
Como puede apreciarse en esta descripción, el mecanismo de gestión de estas colas es muy sencillo y, por tanto, su funcionamiento en la práctica es muy rápido, lo que constituye una de sus principales ventajas desde el punto de vista técnico. Además hay que destacar que las operaciones de selección e introducción de las células en el buffer de salida no generan ningún retardo adicional cuando en dicho buffer existen células o el enlace de salida está ocupado.

Los parámetros de control del sistema son los tamaños de los buffers, N , R y M . Estos valores deben elegirse cuidadosamente, de forma que se cumplan los requisitos exigidos por el tráfico de la red. Estos requisitos se concretan en minimizar el retardo para el tráfico TR y minimizar las pérdidas para el tráfico TNR, sin que ello se consiga a costa de incrementar en demasía las pérdidas para el tráfico TR y el retardo para el tráfico TNR. Si valoramos en α el coste del retardo por unidad de tiempo y por célula para el tráfico TR y en β el coste de la pérdida de una célula en el tráfico TNR, nuestro objetivo es encontrar los valores de N , R y M que minimizan el coste medio por unidad de tiempo a largo plazo para el sistema:

$$\phi(M, N, R) = \lim_{t \rightarrow \infty} \frac{\alpha E[Y_{TR}(t)] + \beta E[L_{TNR}(t)]}{t}$$

donde $E[Y_{TR}(t)]$ es el retardo medio total acumulado por los clientes TR hasta t , y $E[L_{TNR}(t)]$ es el número medio de células TNR perdidas hasta t . Este mínimo debe hallarse con la restricción de que el tiempo medio de espera de los clientes TNR no supere un umbral W_1 y la tasa media de pérdida para los clientes TR no rebase tampoco el umbral g_1 .

Hemos llevado a cabo diversas simulaciones comparando esta política de gestión con otras dos políticas basadas en rondas. Para la comparación, hemos utilizado una política RR (*Round Robin*), consistente en atender k células de clase TR por cada célula de clase TNR, y una política HOL (*Head of the Line*), consistente en vaciar la cola TR cada vez que atiende a un número predeterminado j de células TNR. Estas simulaciones se llevaron a cabo con el simulador YATS (*Yet Another Tinny Simulator*) de redes ATM, debiendo diseñarse las clases de objetos de red que implementaran estas políticas de gestión. El código de estos objetos se encuentra disponible en [YATS]. El objetivo de las simulaciones ha sido comprobar el comportamiento de las tres políticas bajo distintas condiciones de carga y relaciones de servicio para cada clase de cliente. En todos los casos se han dispuesto los parámetros de control de los distintos mecanismos de gestión descritos de forma que las tasas de servicio dedicadas a cada tipo de tráfico sean lo más parecidas posibles en todos los casos, para que tenga sentido la comparación. En particular, los resultados que se muestran a continuación corresponden al caso en que, por término medio, se atienden dos células TR por cada célula TNR. Hemos denotado como OB (*Output Buffer*) a los resultados obtenidos con la nueva política propuesta:



Los gráficos de la izquierda corresponden a condiciones de alta carga, y los de la derecha a condiciones de carga media de la red. En el caso de alta carga, las simulaciones se realizaron con una carga inicial del 80 % (con un 60 % correspondiente al tráfico TR y un 20 % al TNR); para apreciar el comportamiento de las distintas políticas cuando un tipo de tráfico aumenta su caudal, se analizó el efecto del incremento de la tasa de cada una de las clases de tráfico en un 11.1 %; pasando a generar un caudal del 86.6 % cuando el incremento es del tráfico TR y del 82.2 % cuando el incremento es del tráfico TNR.

En los gráficos de la derecha se muestra el comportamiento a media carga; ahora el caudal del tráfico TR es del 42 % y el del TNR el 14 %, generando una carga total del 57.1 %. Al igual que en el caso anterior, se analizó el efecto del incremento de la tasa de cada una de las clases de tráfico en un 16.6 %, pasando a un caudal del 64.2% cuando el incremento es del tráfico TR y del 59.5 % cuando el incremento es del tráfico TNR.

Como puede apreciarse en estas gráficas, la política presentada es, en general, mejor que la HOL (produce casi siempre retardos menores o iguales que ésta para las dos clases de tráfico en todas las condiciones de carga), y reduce los tiempos de espera para el tráfico TR en comparación con la política RR, aunque a costa de incrementar los tiempos de espera del tráfico TNR. Debe destacarse también que en las simulaciones se observa que la nueva política disminuye los retardos máximos en ambos tipos de tráfico, causando por tanto una menor distorsión en el patrón del tráfico a la salida del sistema. Esta comprobación empírica de las buenas propiedades de esta política de gestión invita a realizar un estudio analítico de la misma, que nos permita disponer en cada caso de los valores óptimos de los parámetros de control del sistema.

En lo que sigue supondremos que ambas clases de tráfico llegan al sistema según sendos procesos de Poisson, de parámetros λ_R y λ_N respectivamente. Asimismo, supondremos que el tiempo de servicio sigue una distribución de probabilidad general, idéntica para ambas clases de tráfico, aunque posteriormente particularizaremos los resultados para servicio determinista. Ello es razonable para una red de comunicaciones, donde la velocidad del canal (que es el servidor del sistema) es constante e idéntica cualquiera que sea el tipo de información que se transmita. Más discutible es la hipótesis del proceso de Poisson para las llegadas, que se utiliza básicamente porque permite la suficiente simplicidad en el tratamiento analítico del problema. No obstante, la técnica empleada es generalizable a modelos con llegadas de Erlang o con distribuciones tipo fase, que permiten abarcar tipos de tráfico más generales. En cualquier caso, el mecanismo de gestión y control presentado ha demostrado su eficiencia en simulaciones con tráficos de entrada que no son de Poisson. En particular, se han realizado diversas pruebas cuando las células de ambas clases siguen patrones de tráfico a ráfagas, generados mediante la mezcla de varias fuentes ON-OFF. En cada caso, durante los periodos OFF, de duraciones exponenciales con medias diferentes según la fuente de procedencia, no se generan células; durante los periodos ON, de duraciones también exponenciales se generan células espaciadas regularmente. En las simulaciones puede observarse que, con estos patrones de tráfico, el comportamiento a grandes rasgos de los retardos producidos en el conmutador siguiendo las tres políticas señaladas no difiere del observado para el tráfico de Poisson, aunque los valores son, desde luego, distintos.

2. DESCRIPCION DEL ESTADO DEL SISTEMA

El estado del sistema de colas anterior en el instante t puede describirse mediante la terna $(N_{TR}(t), N_{TNR}(t), S_{TNR}(t))$ donde:

$N_{TR}(t)$: es el número total de clientes TR en el sistema en el instante t .

$N_{TNR}(t)$: es el número total de clientes TNR en el buffer TNR en el instante t .

$S_{TNR}(t)$: es la posición que ocupa el (único) cliente TNR en la cola compartida, en caso de haber alguno de estos clientes en dicha cola; en caso contrario esta variable vale 0.

Dado que los buffers son finitos, este sistema alcanza necesariamente el equilibrio. Llamaremos entonces:

$$(N_{TR}, N_{TNR}, S_{TNR}) = \lim_{t \rightarrow \infty} N_{TR}(t), N_{TNR}(t), S_{TNR}(t)$$

De acuerdo con la política de gestión del sistema, sólo son posibles los estados de la forma:

$$(k, 0, 0), \quad 0 \leq k \leq R+M-1$$

$$(k, n, m), \quad m-1 \leq k \leq R+M-1, \quad 0 \leq n \leq N, \quad 1 \leq m \leq M$$

que constituyen un total de $M + R + M(N+1) \left(R + \frac{M+1}{2} \right)$ estados posibles.

Nuestro primer objetivo será determinar las probabilidades en el equilibrio, π_{ijk} , de los distintos estados del sistema. Para ello, de igual modo que en el tratamiento clásico de las colas M/G/1, podríamos calcular estas probabilidades a partir de la cadena de Markov encajada en los instantes de salida. Para ello basta calcular las probabilidades de transición entre dos instantes de salida del sistema sucesivos. Esta es una tarea sencilla pero trabajosa. Así, llamando a_k a la probabilidad de que durante un tiempo de servicio lleguen k clientes TR, y c_j a la probabilidad de que lleguen j TNR tendríamos, por ejemplo:

$$p_{000, k00} = a_k c_0, \quad 0 \leq k \leq R+M-3$$

El cálculo es más complicado cuando entre dos salidas se producen llegadas de las dos clases de clientes; si en la primera salida no quedó ningún cliente TNR en la cola compartida, hay que determinar el orden de llegada para saber en qué posición de esta cola queda el primer cliente TNR que accede a ella. Si llamamos:

$\varphi_{rs}(j) = \text{Prob}$ (si entre dos salidas sucesivas llegan r clientes TR y s clientes TNR, el primero de los clientes TNR llegue en la j -ésima posición, $1 \leq j \leq r+s$)

puede probarse que:

$$\varphi_{rs}(j) = \frac{\binom{r+s-j}{s-1}}{\binom{r+s}{s}}$$

y tendríamos, por ejemplo,

$$p_{000, knm} = \frac{\lambda_R}{\lambda_R + \lambda_N} a_k c_{n+1} \varphi_{kn+1}(m), \quad 1 \leq k \leq R+M-3, \quad 1 \leq m \leq M-2, \quad 0 \leq n \leq N-1$$

Si denotamos por E_m al conjunto de estados para los que $S_{TNR} = m$ (ordenando dentro de este conjunto los estados en orden creciente de N_{TR} y N_{TNR}), puede demostrarse que la matriz de transiciones es de la forma:

$$P = \begin{matrix} & E_0 & E_1 & E_2 & & E_{M-1} & E_M \\ \begin{matrix} E_0 \\ E_1 \\ E_2 \\ E_3 \\ \vdots \\ E_M \end{matrix} & \begin{pmatrix} A_{00} & A_{01} & A_{02} & \dots & A_{0M-1} & A_{0M} \\ A_{10} & A_{12} & A_{12} & \dots & A_{1M-1} & A_{1M} \\ 0 & A_{21} & 0 & \dots & 0 & 0 \\ 0 & 0 & A_{32} & \dots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \dots & A_{MM-1} & 0 \end{pmatrix} \end{matrix}$$

donde las matrices A_{ij} a su vez pueden descomponerse en subcajas de matrices triangulares o muy huecas. Si bien es posible obtener las probabilidades estacionarias π_{ijk} a partir de esta matriz mediante la resolución del sistema:

$$\begin{aligned} \pi P &= \pi \\ \sum \pi_{ijk} &= 1 \end{aligned}$$

resulta un problema numérico complejo. Por ello procederemos de otro modo para encontrar las probabilidades π_{ijk} .

3. CALCULO DE LAS PROBABILIDADES ESTACIONARIAS DEL SISTEMA

El proceso $\{(N_{TR}(t), N_{TNR}(t), S_{TNR}(t)), t \geq 0\}$ es regenerativo. Este proceso se regenera a sí mismo en aquellos instantes en que se produce una llegada que encuentra el sistema vacío. Si definimos un *ciclo* como el tiempo transcurrido entre dos llegadas que encuentran el sistema vacío, y llamamos:

T = Longitud de un ciclo

T_{ijk} = Tiempo total, durante un ciclo, en que el sistema se encuentra en el estado (i,j,k) .

Entonces, de acuerdo con la teoría de procesos regenerativos:

$$\pi_{ijk} = \frac{E[T_{ijk}]}{E[T]} \quad (1)$$

En particular, debido a que las llegadas son de Poisson, se tiene $E[T_{000}] = 1/(\lambda_R + \lambda_N)$, y por tanto:

$$\pi_{000} = \frac{1}{(\lambda_R + \lambda_N)E[T]} \quad (2)$$

Siguiendo la línea argumental de Tijms (1986) es posible construir ecuaciones recursivas para calcular las probabilidades π_{ijk} utilizando las expresiones anteriores. Para ello definimos:

N_{ijk} = Número de clientes atendidos durante un ciclo que a su salida dejan el sistema en el estado (i,j,k) .

$A_{(i,j,k),(a,b,c)}$ = Tiempo medio, durante un servicio que comenzó con el sistema en el estado (i,j,k) , en que el sistema permanece en el estado (a,b,c) .

Observando ahora que:

- el primer servicio de un ciclo comienza con un cliente presente (que puede ser TR o TNR),
- durante un servicio el sistema puede alcanzar el estado $(a,0,0)$ solamente si comenzó en un estado $(i,0,0)$ con $i \leq a$,

- durante un servicio, el sistema puede alcanzar el estado (a,b,c) con $1 \leq c \leq M$ sólo si el servicio comenzó con el sistema en algún estado (i,j,c) , con $i \leq a$, $j \leq b$ (ya que la posición c del cliente TNR en la cola compartida no se altera durante un tiempo de servicio), o bien si el sistema empezó en algún estado $(i,0,0)$ con $i \leq a$, $i < c$,

podemos escribir las siguientes ecuaciones:

Si $c - 1 \leq a \leq R + M - 1$, $1 < c \leq M$, $0 \leq b \leq N$:

$$E[T_{abc}] = \frac{\lambda_R}{\lambda_R + \lambda_N} A_{(1,0,0)(a,b,c)} + \sum_{i=c-1}^a \sum_{j=0}^b E[N_{ijc}] A_{(i,j,c)(a,b,c)} + \sum_{i=1}^a E[N_{i00}] A_{(i,0,0)(a,b,c)}$$

Si $0 \leq a \leq R + M - 1$, $0 \leq b \leq N$:

$$E[T_{ab1}] = \frac{\lambda_N}{\lambda_R + \lambda_N} A_{(0,0,1)(a,b,1)} + \sum_{i=0}^a \sum_{j=0}^b E[N_{ij1}] A_{(i,j,1)(a,b,1)} \quad (3)$$

Si $0 \leq a \leq R + M - 1$:

$$E[T_{a00}] = \frac{\lambda_R}{\lambda_R + \lambda_N} A_{(1,0,0)(a,0,0)} + \sum_{i=0}^a E[N_{i00}] A_{(i,0,0)(a,0,0)}$$

Dado que las llegadas y salidas en este sistema se producen de una en una, cuando el sistema está en equilibrio el número medio $E[N_{abc}]$ de salidas que dejan el sistema en el estado (a,b,c) es igual al número medio de llegadas que encuentran el sistema en este estado. Ahora bien, puesto que las llegadas se producen según un proceso de Poisson, (siempre que el sistema disponga de espacio disponible para alojar a cada nueva llegada), este último número medio es $\lambda_{abc}^* [T_{abc}]$, donde:

$$\lambda_{abc}^* = \begin{cases} \lambda_R + \lambda_N & \text{si } a < R + M - 1 \text{ y } b + c < N + 1 \\ \lambda_R & \text{si } a < R + M - 1 \text{ y } b + c \geq N + 1 \\ \lambda_N & \text{si } a \geq R + M - 1 \text{ y } b + c < N + 1 \\ 0 & \text{si } a \geq R + M - 1 \text{ y } b + c \geq N + 1 \end{cases}$$

Por tanto:

$$E[N_{abc}] = \lambda_{abc}^* E[T_{abc}] \quad (4)$$

Sustituyendo estos valores en el sistema de ecuaciones (3), llamando $p_R = \lambda_R / (\lambda_R + \lambda_N)$, $p_N = \lambda_N / (\lambda_R + \lambda_N)$, dividiendo todos los términos por la duración media del ciclo, $E[T]$, y teniendo en cuenta (1) y (2), se llega a:

Si $c - 1 \leq a \leq R + M - 1$, $1 < c \leq M$, $0 \leq b \leq N$:

$$\pi_{abc} = p_R \lambda_{abc}^* \pi_{000} A_{(1,0,0)(a,b,c)} + \sum_{i=c-1}^a \sum_{j=0}^b \lambda_{abc}^* \pi_{ijc} A_{(i,j,c)(a,b,c)} + \sum_{i=1}^a \lambda_{abc}^* \pi_{i00} A_{(i,0,0)(a,b,c)}$$

Si $0 \leq a \leq R + M - 1$, $0 \leq b \leq N$:

$$\pi_{ab1} = p_N \lambda_{abc}^* \pi_{000} A_{(0,0,1)(a,b,1)} + \sum_{i=0}^a \sum_{j=0}^b \lambda_{abc}^* \pi_{ij1} A_{(i,j,1)(a,b,1)} \quad (5)$$

Si $0 \leq a \leq R + M - 1$:

$$\pi_{a00} = p_R \lambda_{abc}^* \pi_{000} A_{(1,0,0)(a,0,0)} + \sum_{i=0}^a \lambda_{abc}^* \pi_{i00} A_{(i,0,0)(a,0,0)},$$

Si en este sistema se dividen todas las ecuaciones por π_{000} , llamando $\theta_{abc} = \pi_{abc} / \pi_{000}$, pueden obtenerse con facilidad de modo recursivo los valores θ_{abc} . Las probabilidades π_{abc} se obtienen luego fácilmente sin más que observar que:

$$\pi_{000} + \sum_{(a,b,c) \neq (0,0,0)} \pi_{abc} = 1 \Rightarrow \sum_{(a,b,c) \neq (0,0,0)} \theta_{abc} = \frac{1}{\pi_{000}}$$

de donde:

$$\pi_{000} = \frac{1}{1 + \sum_{(a,b,c) \neq (0,0,0)} \theta_{abc}} \quad (6)$$

$$\pi_{000} = \frac{\theta_{abc}}{1 + \sum_{(a,b,c) \neq (0,0,0)} \theta_{abc}}$$

Obviamente, aún falta determinar los términos $A_{(i,j,k)(a,b,c)}$.

Comencemos por calcular $A_{(i,j,c)(a,b,c)}$ para $1 \leq c \leq M$, $0 \leq i \leq a$, $0 \leq j \leq b$. Para ello definimos la variable aleatoria:

$$\chi_{(i,j,k)(a,b,c)}(t) = \begin{cases} 1 & \text{si en el instante } t \text{ el sistema está en el estado } (a,b,c), \text{ y aún} \\ & \text{no ha terminado un servicio que comenzó en el instante } 0 \\ & \text{con el sistema en estado } (i,j,c) \\ 0 & \text{en otro caso} \end{cases}$$

Es evidente entonces que:

$$A_{(i,j,k)(a,b,c)} = E \left[\int_0^{\infty} \chi_{(i,j,c)(a,b,c)}(t) dt \right] = \int_0^{\infty} E[\chi_{(i,j,c)(a,b,c)}(t)] dt \quad (7)$$

Ahora bien:

$$E[\chi_t] = P(\chi_t = 1) = (1 - B(t)) e^{-\lambda_R t} \frac{(\lambda_R t)^{a-i}}{(a-i)!} e^{-\lambda_N t} \frac{(\lambda_N t)^{b-i}}{(b-i)!} \quad (8)$$

donde $B(t)$ es la función de distribución del tiempo de servicio. En el caso particular de servicio determinista de duración D se tiene:

$$B(t) = \begin{cases} 0 & t < D \\ 1 & t \geq D \end{cases} \quad (9)$$

Luego, en este caso:

$$A_{(i,j,c)(a,b,c)} = \int_0^D e^{-\lambda_R t} \frac{(\lambda_R t)^{a-i}}{(a-i)!} e^{-\lambda_N t} \frac{(\lambda_N t)^{b-i}}{(b-i)!} dt \quad (10)$$

Si llamamos:

$$G(m,n) = \int_0^D e^{-\lambda_R t} \frac{(\lambda_R t)^m}{m!} e^{-\lambda_N t} \frac{(\lambda_N t)^n}{n!} dt \quad (11)$$

$$F(m,n) = e^{-(\lambda_R + \lambda_N)D} \frac{\lambda_R^m \lambda_N^n}{m! n!} \frac{D^{m+n}}{\lambda_R + \lambda_N}$$

puede comprobarse fácilmente la relación recursiva:

$$G(m,n) = \frac{\lambda_R}{\lambda_R + \lambda_n} G(m-1,n) + \frac{\lambda_N}{\lambda_R + \lambda_n} G(m,n-1) - F(m,n), \quad m \geq 1, n \geq 1 \quad (12)$$

con valores iniciales:

$$G(m,0) = \frac{1}{\lambda_R + \lambda_N} \left[\left(\frac{\lambda_R}{\lambda_R + \lambda_N} \right)^m - e^{-(\lambda_R + \lambda_n)D} \sum_{h=0}^n \left(\frac{\lambda_R}{\lambda_R + \lambda_N} \right)^h \frac{(\lambda_R D)^{m-h}}{(m-h)!} \right] m \geq 0 \quad (13)$$

$$G(0,n) = \frac{1}{\lambda_R + \lambda_N} \left[\left(\frac{\lambda_N}{\lambda_R + \lambda_N} \right)^n - e^{-(\lambda_R + \lambda_n)D} \sum_{h=0}^n \left(\frac{\lambda_N}{\lambda_R + \lambda_N} \right)^h \frac{(\lambda_N D)^{n-h}}{(n-h)!} \right] n \geq 0 \quad (14)$$

De esta forma,

$$A_{(i,j,c)(a,b,c)} = G(a-i, b-j), \quad 1 \leq c \leq M, 0 \leq i \leq a, 0 \leq j \leq b \quad (15)$$

y su valor puede calcularse utilizando recursivamente (12) con los valores iniciales (13) y (14).

Para calcular ahora $A_{(i,0,0)(a,b,c)}$, con $1 \leq c < M$ debemos notar que para que durante un tiempo de servicio se pase del estado $(i,0,0)$ al (a,b,c) , deben haber llegado $a-i$ clientes TR y $b+1$ clientes TNR, con la condición de que el primer cliente TNR haya llegado cuando hay exactamente $c-1$ clientes TR en el sistema. Utilizando la propiedad del proceso de Poisson de que si en un periodo $(0,t)$ se producen n ocurrencias del mismo, éstas se distribuyen uniformemente en el intervalo, la probabilidad de que la primera llegada de un cliente TNR sea la que ocupe la posición c en la cola compartida habiendo llegado $a-i$ clientes TR y $b+1$ clientes TNR en $(0,t)$, viene dada por:

$$\frac{(a-i)! (a+b+1-c)!}{(a-c+1)! (a+b+1-i)!} (b+1) \quad (16)$$

Procediendo ahora como para el cálculo de $A_{(i,j,c)(a,b,c)}$, llegamos a:

$$A_{(i,0,0)(a,b,c)} = \frac{(a-i)! (a+b+1-c)!}{(a-c+1)! (a+b+1-i)!} (b+1) G(a-i, b+1) \quad (17)$$

Un razonamiento similar, teniendo en cuenta ahora que para que un cliente TNR ocupe la posición M debe haber llegado cuando en el sistema hay $M-1$ ó más clientes TR nos permite obtener:

$$A_{(i,0,0)(a,b,M)} = \frac{(a-i)! (a+b+2-M)!}{(a-M+1)! (a+b+1-i)!} G(a-i, b+1) \quad (18)$$

Por último, también de modo muy sencillo se deduce que:

$$A_{(i,0,0)(a,0,0)} = G(a-i, 0) \quad (19)$$

De esta forma, las ecuaciones (15), (17), (18) y (19) junto con (12) nos proporcionan el esquema recursivo preciso para obtener finalmente las probabilidades en el equilibrio π_{ijk} a partir de (5) y (6).

4. PROBABILIDADES DE PERDIDA

Un cliente TR es rechazado por el sistema cuando a su llegada no hay espacio en buffer para alojarlo; ésto ocurre cuando en total en el sistema hay $R+M-1$ clientes TR. Por tanto, la probabilidad de pérdida en el estado estacionario para los clientes TR viene dada por:

$$\gamma_{TR} = \pi_{R+M-1,0,0} + \sum_{c=1}^M \sum_{b=0}^N \pi_{R+M-1,b,c} \quad (20)$$

Dado que durante un ciclo tratan de acceder al sistema por término medio $\lambda_R E[T]$ clientes TR, si llamamos $L_{TR}(T)$ al número de clientes TR rechazados por ciclo, se tiene, haciendo uso de (2):

$$E[L_{TR}(T)] = \gamma_{TR} \lambda_R E[T] = \frac{\gamma_{TR} \lambda_R}{(\gamma_R + \lambda_N) \pi_{000}} \quad (21)$$

Asimismo, la probabilidad de pérdida en el estado estacionario para los clientes TNR viene dada por:

$$\gamma_{TNR} = \sum_{a=0}^{R+M-1} \sum_{c=1}^M \pi_{aNc} \quad (22)$$

y el número medio de clientes TNR rechazados por ciclo es:

$$E[L_{TNR}(T)] = \gamma_{TNR} \lambda_N E[T] = \frac{\gamma_{TNR} \lambda_N}{(\lambda_R + \lambda_N) \pi_{000}} \quad (23)$$

6. TIEMPOS DE ESPERA

Situémonos en nuestro sistema en el preciso instante en que se acaba de completar un servicio (se transmite una célula), que ha dejado el sistema en estado (a,b,c) , y sea ω_{ibc} , con $i \leq a$, el tiempo que aún debe esperar en cola el cliente TR que tiene delante $i - 1$ clientes TR. Obviamente, si $i \leq M - 1$:

$$\omega_{ibc} = \begin{cases} \sum_{j=1}^{i-1} X_j & \text{si } i < c \text{ ó } c = b = 0 \\ \sum_{j=1}^i X_j & \text{si } i \geq c \geq 1 \end{cases} \quad (24)$$

siendo X_j la duración de un tiempo de servicio. Ahora bien, si $i \geq M$, el tiempo de espera de este cliente no se verá afectado por los nuevos clientes TR que lleguen a partir de ahora, pero sí podría verse afectado por los clientes TNR que llegasen a partir de este momento. En efecto, existe la posibilidad de que durante el tiempo que tarde nuestro cliente TR en llegar al buffer compartido, en virtud de la política de gestión empleada puedan acceder a este buffer clientes TNR que llegaron después que él, pero que encontraron ninguna o poca cola en su buffer específico y pudieron entrar pronto en el buffer compartido. Podemos establecer entonces las siguientes relaciones de recurrencia para el tiempo de espera de nuestro cliente TR:

- Si $c = 0$:

$$\omega_{i00} = \begin{cases} X + \omega_{i-1,0,0} & \text{con prob. } \beta_0 \\ X + \omega_{i-1,j,M-1} & \text{con prob. } \beta_{j+1}, 0 \leq j < N \\ X + \omega_{i-1,N,M-1} & \text{con prob. } \beta_{N+1}^* \end{cases} \quad (25)$$

donde X es la duración de un tiempo de servicio y β_k es la probabilidad de que durante un tiempo de servicio lleguen al sistema k clientes TNR. Por su parte, β_k^* es la probabilidad de que durante un servicio lleguen k o más clientes TNR.

- Si $c = 1$:

$$\omega_{i01} = \begin{cases} X + \omega_{i,0,0} & \text{con prob. } \beta_0 \\ X + \omega_{i,j,M} & \text{con prob. } \beta_{j+1}, 0 \leq j < N-1 \\ X + \omega_{i,N-1,M} & \text{con prob. } \beta_N^* \end{cases} \quad (26)$$

$$\omega_{ib1} = \begin{cases} X + \omega_{i,b-1+j,M} & \text{con prob. } \beta_j, 0 \leq j < N-b, 0 < b < N \\ X + \omega_{i,N-1,M} & \text{con prob. } \beta_{N-b}^* \end{cases} \quad (27)$$

$$\omega_{k,N,1} = X + \omega_{k,N-1,M} \quad (28)$$

- Si $1 < c \leq M$:

$$\omega_{ibc} = \begin{cases} X + \omega_{i-1,b+j,c-1} & \text{con prob. } \beta_j, 0 \leq j < N-b, 0 \leq b < N \\ X + \omega_{i-1N,c-1} & \text{con prob. } \beta_{N-b}^* \end{cases} \quad (29)$$

$$\omega_{i,N,c} = X + \omega_{i-1,N,c-1} \quad (30)$$

De estas ecuaciones pueden obtenerse las siguientes relaciones de recurrencia para los tiempos medios de espera:

- Si $i \leq M - 1$:

$$E[\omega_{ibc}] = \begin{cases} (i-1)E[X] & \text{si } i < c \text{ ó } c = b = 0 \\ iE[X] & \text{si } i \geq c \geq 1 \end{cases} \quad (31)$$

- Si $i \geq M$:

$$\begin{aligned} E[\omega_{i00}] &= E[X] + \beta_0 E[\omega_{i-1,0,0}] + \sum_{j=0}^{N-1} \beta_{j+1} E[\omega_{i-1,j,M-1}] + \beta_{N+1}^* E[\omega_{i-1,N,M-1}] \\ E[\omega_{i01}] &= E[X] + \beta_0 E[\omega_{i,0,0}] + \sum_{j=0}^{N-2} \beta_{j+1} E[\omega_{i,j,M}] + \beta_N^* E[\omega_{i,N-1,M}] \\ E[\omega_{ib1}] &= E[X] + \sum_{j=0}^{N-b} \beta_j E[\omega_{i,b-1+j,M}] + \beta_{N-b}^* E[\omega_{i,N-1,M}], \quad 0 < b < N \\ E[\omega_{iN1}] &= E[X] + E[\omega_{i,N-1,M}] \\ E[\omega_{ibc}] &= E[X] + \sum_{j=0}^{N-b} \beta_j E[\omega_{i-1,b+j,c-1}] + \beta_{N-b}^* E[\omega_{i-1,N,c-1}], \quad 0 \leq b < N \\ E[\omega_{iNc}] &= E[X] + E[\omega_{i-1,N,c-1}] \end{aligned} \quad (32)$$

Las ecuaciones (32) junto con los valores iniciales dados por (31) permiten obtener recursivamente las esperanzas $E[\omega_{ijk}]$ para todos los estados (i,j,k) posibles. La resolución de estas ecuaciones será más o menos difícil en función de cuál sea la función de distribución de probabilidad del tiempo de servicio. En el caso particular de tiempo de servicio determinista de duración D se tiene:

$$E[X] = D \quad (33)$$

$$\beta_k = e^{-\lambda_N D} \frac{(\lambda_N D)^k}{k!}, \quad \beta_k^* = 1 - \sum_{j=0}^{k-1} e^{-\lambda_N D} \frac{(\lambda_N D)^j}{j!}$$

Ahora bien, las esperanzas obtenidas en (32) corresponden a los tiempos medios de espera en cola medidos a partir del momento en que termina un servicio. El tiempo de espera global de un cliente TR que a su llegada encuentra el sistema en estado (a,b,c) es:

- Si $a < M - 1$:

$$W_{abc} = \begin{cases} X_{RES} + \sum_{i=1}^{a-1} X_i & \text{si } b = c = 0 \\ X_{RES} + \sum_{i=1}^a X_i & \text{si } c > 1 \end{cases} \quad (34)$$

donde X_{RES} es el tiempo de servicio residual que falta para que termine el servicio del cliente que ocupa la cabecera de la cola compartida a la llegada del cliente TR.

- Si $a \geq M - 1$:

Si durante el tiempo de servicio residual del cliente que ocupa la cabecera de la cola llegan nuevos clientes TR, estos no afectan al tiempo de espera del cliente TR que acaba de llegar. Sin embargo, por la misma razón señalada más arriba, sí que afectan los clientes TNR que lleguen durante este tiempo. Si llamamos α_k a la probabilidad de que durante el tiempo de servicio residual posterior a la llegada del cliente TR lleguen k clientes TNR, y α_k^* a la probabilidad de que lleguen k ó más, tenemos:

$$W_{abc} = \begin{cases} X_{RES} + \omega_{a-1,j,M} & \text{con prob. } \alpha_{j+1}, 0 \leq j \leq N & (b = c = 0) \\ X_{RES} + \omega_{a-1,N,M} & \text{con prob. } \alpha_{N+1}^* & (b = c = 0) \\ X_{RES} + \omega_{a,0,0} & \text{con prob. } \alpha_0 & (c = 1, b = 0) \\ X_{RES} + \omega_{a,j,M} & \text{con prob. } \alpha_{j+1}, 0 \leq j \leq N-2 & (c = 1, b = 0) \\ X_{RES} + \omega_{a,N-1,M} & \text{con prob. } \alpha_N^* & (c = 1, b = 0) \\ X_{RES} + \omega_{a,b+j-1,M} & \text{con prob. } \alpha_j, 0 \leq j \leq N-b & (c = 1, b > 0) \\ X_{RES} + \omega_{a,N,M} & \text{con prob. } \alpha_{N-b}^* & (c = 1, b > 0) \\ X_{RES} + \omega_{a-1,b+j,c-1} & \text{con prob. } \alpha_j, 0 \leq j \leq N-b & (c > 1) \\ X_{RES} + \omega_{a-1,N,c-1} & \text{con prob. } \alpha_{N-b}^* & (c > 1) \\ X_{RES} + \omega_{a-1,0,0} & \text{con prob. } \alpha_0 & (b = c = 0) \end{cases} \quad (35)$$

De modo similar a como hicimos anteriormente para los tiempos de espera ω_{ijk} , podemos ahora hallar a partir de (35) el tiempo medio de espera en cola para un cliente TR que a su llegada encuentra el sistema en el estado (a,b,c) :

$$\begin{aligned} E[W_{a00}] &= E[X_{RES}] + \alpha_0 E[\omega_{a-1,0,0}] + \sum_{j=0}^N \alpha_{j+1} E[\omega_{a-1,j,M}] + \alpha_{N+1}^* E[\omega_{a-1,N,M}] \\ E[W_{a01}] &= E[X_{RES}] + \alpha_0 E[\omega_{a,0,0}] + \sum_{j=0}^N \alpha_{j+1} E[\omega_{a,j,M}] + \alpha_{N+1}^* E[\omega_{a,N-1,M}] \\ E[W_{ab1}] &= E[X_{RES}] + \sum_{j=0}^{N-b} \alpha_j E[\omega_{a,b+j-1,M}] + \alpha_{N-b}^* E[\omega_{a,N,M}], \quad 0 < b \leq N \\ E[W_{abc}] &= E[X_{RES}] + \sum_{j=0}^{N-b} \alpha_j E[\omega_{a-1,b+j,c-1}] + \alpha_{N-b}^* E[\omega_{a-1,N,c-1}], \quad 0 \leq b \leq N, c > 1 \end{aligned} \quad (36)$$

Estas esperanzas pueden calcularse recursivamente utilizando como valores iniciales las esperanzas que se obtienen directamente de (34):

$$E[W_{a00}] = E[X_{RES}] + (a-1)E(X), 1 \leq a \leq M-1 \quad (37)$$

$$E[W_{abc}] = E[X_{RES}] + aE(X), \text{ si } c > 1$$

y la condición obvia: $E[W_{000}] = 0$.

Como ya ocurrió con los valores medios $E[\omega_{ijk}]$, la obtención de las esperanzas en (36) depende de la dificultad de cálculo de las probabilidades α_k . En el caso particular de servicio determinista de duración D puede probarse que:

$$a_k = \int_0^D e^{-\lambda_N(D-t)} \frac{[\lambda_N(D-t)]^k}{k!} \frac{1}{D} dt = \frac{1}{\lambda_N} \left[1 - \frac{1}{\lambda_N D} \left(k+1 - e^{-\lambda_N D} \sum_{j=0}^k (j+1) \frac{(\lambda_N D)^{k-j}}{(k-j)!} \right) \right] \quad (38)$$

En este caso, además $E[X_{RES}] = D/2$.

Denotemos ahora por W^{TR} (respectivamente, W^{TNR}) a la variable aleatoria que mide el tiempo de espera de un cliente TR (respectivamente, TNR) arbitrario, una vez que el sistema ha alcanzado el equilibrio. Llamemos, asimismo, N_{TR} (respectivamente, N_{TNR}) al número de clientes TR (respectivamente, TNR) presentes en el sistema cuando éste se encuentra en equilibrio.

Estamos en condiciones de poder calcular el tiempo medio de espera de un cliente TR arbitrario. Si condicionamos por el estado del sistema a la llegada de este cliente tenemos:

$$E[W^{TR}] = \sum_{(a,b,c)} E[W_{abc}] \pi_{abc} \quad (39)$$

El valor de esta esperanza se calcula haciendo uso de (36) y (37), con las probabilidades estacionarias halladas en (5) y (6).

Por último, el tiempo medio de espera para los clientes TNR puede hallarse con la utilización de la fórmula de Little. En este sistema concreto, la fórmula de Little adopta la forma:

$$E[N_{TR}] + E[N_{TNR}] = \lambda_R(1-\gamma_{TR})E[W^{TR}] + \lambda_N(1-\gamma_{TNR})E[W^{TNR}] \quad (40)$$

esto es, el número medio total de clientes en el sistema en el equilibrio debe ser igual a la suma de las tasas efectivas de llegada de cada clase de clientes por sus tiempos medios de espera respectivos. Los números medios $E[N_{TR}]$ y $E[N_{TNR}]$ pueden calcularse fácilmente a partir de las probabilidades π_{ijk} :

$$E[N_{TR}] = \sum_{a=0}^{R+M-1} a \pi_{a00} + \sum_{a=0}^{R+M-1} a \sum_{b=0}^N \sum_{c=1}^M \pi_{abc} \quad (41)$$

$$E[N_{TNR}] = \sum_{a=0}^{R+M-1} \sum_{b=0}^N \sum_{c=1}^M (b+c) \pi_{abc}$$

y de (40) se despeja fácilmente:

$$E[W^{TNR}] = \frac{E[N_R] + E[N_D] - \lambda_R(1-\gamma_{TR})E[W^{TR}]}{\lambda_N(1-\gamma_{TNR})} \quad (42)$$

7. OPTIMIZACION DEL RENDIMIENTO DEL SISTEMA

Como se ha dicho en el epígrafe 1, los valores de M , N y R deben elegirse de modo que se minimice el coste medio por unidad de tiempo a largo plazo que suponen los retardos para los clientes TR y las pérdidas para los TNR:

$$\phi[M, N, R] = \lim_{t \rightarrow \infty} \frac{\alpha E[Y_{RT}(t)] + \beta E[L_{TNR}(t)]}{t}$$

Este mínimo debe hallarse con la restricción de que $E[W^{TNR}] < W_1$ y $\gamma_{TR} < g_1$. Dado que el proceso $(N_{TR}(t), N_{TNR}(t), S_{TNR}(t))$ es regenerativo, la teoría de la renovación nos permite calcular el límite anterior de la forma:

$$\phi[M, N, R] = \frac{\alpha E[Y_{TR}(T)] + \beta E[L_{TNR}(T)]}{E[T]} \quad (43)$$

esto es, el coste medio para el sistema a largo plazo coincide con el coste medio durante un periodo de renovación. La ecuación (2) nos permite calcular $E[T]$, y la ecuación (23) nos proporciona $E[L_{TNR}(T)]$. Nos falta por determinar solamente $E[Y_{TR}(T)]$, el tiempo de espera acumulado por todos los clientes TR atendidos durante un periodo de renovación. Esta cantidad se halla fácilmente como:

$$E[Y_{TR}(T)] = \lambda_R (1 - \gamma_{TR}) E[T] E[W^{RT}] \quad (44)$$

esto es, el número medio de clientes TR que llegan durante un periodo de renovación, multiplicado por el tiempo medio de espera de cada uno, ya obtenido en (39).

Dado que todos los términos que aparecen en (43) se han calculado numéricamente, no podemos hallar una fórmula explícita para obtener los valores de M , N y R que minimizan (43), y éstos han de hallarse mediante un algoritmo de búsqueda que vaya incrementando y/o decrementando sucesivamente los valores de estos parámetros cumpliendo además las restricciones $E[W_{TNR}] < W_1$ y $\gamma_{TR} < g_1$. Debe señalarse que la forma recursiva de computar todos los elementos necesarios para el cálculo de las funciones involucradas en (43) simplifica mucho los cálculos necesarios en el algoritmo de búsqueda, ya que los términos obtenidos en una iteración pueden ser reutilizados en la siguiente sin necesidad de recalcularlos.

7. CONCLUSIONES, ESTADO ACTUAL Y DESARROLLOS FUTUROS

En este trabajo hemos analizado el rendimiento de una política de gestión de colas en un conmutador de red tal como se muestra en la Figura 1, que recibe dos clases de tráfico con distintos requisitos de calidad en lo que se refiere a retardos y pérdidas. Hemos visto como, de forma recursiva, es posible calcular numéricamente las probabilidades de estado de este sistema, y a partir de las mismas hemos mostrado como obtener los retardos medios de ambas clases de tráfico, así como sus tasas de pérdida. Todas estas cantidades intervienen en el cálculo del rendimiento del sistema a través de una función objetivo que representa una ponderación del coste causado por los retardos del tráfico prioritario (en tiempo real) y las pérdidas causadas al tráfico no prioritario (tiempo no real). Los valores de los parámetros óptimos de control del sistema se obtienen finalmente mediante la aplicación de un algoritmo adecuado que minimice la función objetivo. En el momento actual estamos trabajando en el desarrollo de un algoritmo eficiente adecuado a este fin, así como en la búsqueda de métodos aproximados para el cálculo de los términos de la función objetivo, que permitan obtener aproximaciones de la misma sin requerir tantos cálculos como los presentados en este trabajo. Esta tarea es particularmente interesante cuando los valores previsibles de M , N ó R son grandes.

Futuras extensiones de este trabajo incluyen:

- (1) La comparación de la política de gestión aquí estudiada con una mayor variedad de políticas actualmente implementadas en sistemas de colas de características similares, evaluando los pros y contras de cada una de ellas y estableciendo una clasificación de políticas que permita elegir la mejor en cada caso particular.
- (2) Aunque en este trabajo hemos considerado el caso de tráficos en tiempo real y en tiempo no real, la política estudiada resulta eficiente para repartir los recursos del sistema entre dos tráficos en tiempo real o dos tráficos en tiempo no real. Resulta interesante la generalización de la política aquí

mostrada a conmutadores que reciban más de dos clases de tráfico, e incluir la posibilidad de que en el buffer compartido pueda haber simultáneamente dos o mas clientes TNR.

- (3) Hemos considerado llegadas al sistema según un proceso de Poisson. Un problema abierto es el análisis de llegadas que no sean de Poisson. La generalización más inmediata de este modelo es a llegadas con distribuciones tipo fase o MAP's (*Markovian arrival processes*).
- (4) Por último señalemos que también resulta de interés el estudio de esta política de gestión en sistemas en tiempo discreto, tales como los que se emplean para modelizar la transmisión de células en ATM.

REFERENCIAS

- BANERJEA, A. and S. KESHAV (1993): "Queueing delays in rate controlled ATM networks", in Proc. IEEE INFOCOM'93, 547-556, San Francisco, CA.
<http://emily.mathcs.emory.edu/~cheung/papers/places/tenet/BanKes 93.ps>
- CHANG, C. (1994): "Stability, queue length, and delay of deterministic and stochastic queueing networks", **IEEE Trans. Automatic Contr.**, 39, 913-931.
- CRUZ, R. (1991): "A calculus for network delay, Part I: Network elements in isolation", **IEEE Trans. Inform. Theory**, 37, 114-121, 1991.
- _____ (1991): "A calculus for network delay, Part II: Network analysis", **IEEE Trans. Inform. Theory**, 37, 121-141.
- DEMERS, A.; S. KESHAV and S. SHENKAR (1990): "Analysis and simulation of a fair queueing algorithm", in Proceedings of the ACM SIGCOMM'89, 3-12. Also in **J. Internetworking Res. and Experience**, 3-26.
- FIGUEIRA, N. and J. PASQUALE (1995): "Leave-in-Time: A New Service Discipline for Real-Time Communications in a Packet-Switching Data Network", **ACM**, 25(4),
<http://www.acm.org/sigcomm/ccr/archive/1995/conf/figueira.ps>
- _____ (1995): "An upper bound on delay for the virtualclock service discipline", **IEEE/ACM Transactions on Networking**, 3(4).
- FERRARI, D. and D.C. VERMA (1990): "A scheme for real-time channel establishment in wide-area networks", in **J. Select. Areas Commun. IEEE**, 8(3), 368-379,
<http://emily.mathcs.emory.edu/~cheung/papers/places/tenet/FerVer90a.ps>
- GOLESTANI, S.J. (1990): "A stop-and-go queueing framework for congestion management", in **Proceedings of the ACM SIGCOMM'90**, Philadelphia, PA, 8-18.
- _____ (1994): "A self-clocked fair queueing scheme for broadband applications", in **Proc. IEEE INFOCOMM '94**, Toronto, CA, 636-646.
- JING-FEI, REN.; JON W. MARK and JOHNNY W. WONG (1994): "A dynamic priority queueing approach to traffic regulation and scheduling in B-ISDN", **GLOBECOM '94**.
- KALMANEK, C.R.; H. KANAKIA and S. KESHAV (1990): "Rate controlled servers for very high-speed networks", in **IEEE Global Telecommun. Conf.**, San Diego, CA, 300.3.1-300.3.9.
<http://emily.mathcs.emory.edu/~cheung/papers/places/tenet/KaKaKe90.ps>
- KUROSE, J. (1991): "On computing per-session performance bounds in high-speed multi-hop computer networks", in **Proc. ACM Sigmetrics'92**, Conf., 128-139.
- KNIGHTLY, E.; D. WREGE J. LIEBEHERR and H. ZHANG (1995): "Fundamental limits and tradeoffs for providing deterministic bounding interval dependent traffic models", in **Proc. IEEE INFOCOM'95**, Boston, MA.

- LIN, A.Y. (1991): "Priority queueing strategies and buffer allocation protocols for traffic control at an ATM integrated broadband switching system. **IEEE Journal on selected areas in communications**, 9.
- LIN, A.Y. (1991): "Priority queueing strategies for traffic control at a multichannel ATM switching system", **GLOBECOM '91**, 8B.2.1, 234-238
- LOW, S. (1992): "Traffic control in ATM networks", Ph.D. dissertation, Univ. Calif. Berkeley
- PAREKH, A.K. and R.G. GALLAGER (1993): "A generalized processor sharing approach to flow control in integrated services networks: the single-node case", in **Transactions on Networking**, IEEE/ACM.
- TIJMS, H.C. (1986): "Stochastic Modelling and Analysis. A computational approach", Wiley,
- VERMA, D.; H. ZHANG, and D. FERRARI (1991): "Guaranteeing delay jitter bounds in packet switching networks", in **Proc. TRICOMM'91**, Chapel Hill, NC, 35-46.
- XIE, G. and S. LAM (1994): "Delay guarantee of virtual clock server", **Tech. Rep. TR-94-24**, Dept. Computer Sci., Univ. Texas at Austin, Also in 9th IEEE Workshop on Computer Commun.
- YATS (----): Available in <http://www.ifn.et.tu-dresden.de/TK/yats/yats.html>.
- YE, J. and S. LIN (1991): "Analysis of Multimedia Traffic Queues with Finite Buffer and Overload Control - Part I: Algorithm", CH2979-3/91/0000-1464, IEEE.
- ZHANG, H. (1995): "Service Disciplines For Guaranteed Performance Service in Packet-Switching Networks", **Proceedings of the IEEE**, 83(10),
- ZHANG, H and D. FERRARI (1993): "Rate-controlled static-priority queueing", in Proc. IEEE INFOCOMM'93, San Francisco, CA, 227-236. Also in **Journal of High Speed Networks: Special Issue on Quality of Service**, 3(4).
- _____ (1994): "Rate-controlled Service Disciplines", **Journal of High Speed Networks**, 3(4): 3889-412, <ftp://n1.sp.cs.cmu.edu/pub/hzhang/HighSpeedNetworks94.ps.Z>
- ZHANG, L. (1991a): "VirtualClock: a new traffic control algorithm for packet switching networks", in Proc. of ACM SIGCOMM'90, Philadelphia, PA, 19-29. Also in **ACM Trans. Comput. Syst.**, 9, 101-124.
- _____ (1995): "Service Disciplines For Guaranteed Performance Service in Packet-Switching Networks", **Proceedings of the IEEE**, 83(10), Also in [ftp://n1.sp.cs.cmu.edu/pub/hzhang/IEEE Proceedings'95.ps.Z](ftp://n1.sp.cs.cmu.edu/pub/hzhang/IEEE_Proceedings'95.ps.Z)
- ZHANG, H. and E. KNIGHTLY (1994): "Providing end-to-end statistical performance guarantees with interval dependent stochastic models", in **ACM Sigmetrics'94**, Nashville, TN, 211-220.
- ZHANG, Z.; D. TOWSLEY and J. KUROSE (1994): "Statistical analysis of generalized processor sharing scheduling discipline", in **Proc. ACM SIGCOMM'94**, London, UK,
- ZHANG, Z. [1994] : "Improving utilization for deterministic service in multimedia communication", In **1994 Int. Conf. On Multimedia Computing and Syst.**, Boston, MA, 295-305.