

**ESCOLA TÈCNICA SUPERIOR D'ENGINYERIA DE  
TELECOMUNICACIÓ DE BARCELONA**

**TRÁFICO DE TELEFONÍA MÓVIL:  
CARACTERIZACIÓN E  
IMPLICACIONES DEL TIEMPO DE  
OCUPACIÓN DEL CANAL**

Autor: Francisco Barceló Arroyo  
Director: Josep Paradells Aspás

## CAPÍTULO 5

### *Aproximación del tiempo medio de espera en colas $M/H_2/s$*

En este capítulo la atención se centra en el cálculo del tiempo medio de espera en colas cuya distribución de la duración del tiempo de servicio es hiperexponencial-2. Es uso común en la literatura utilizar la distribución hiperexponencial balanceada  $H_2^b$  (en la distribución  $H_2^b$  el conjunto de servidores dedica el 50% de su tiempo a cada uno de las dos clases de llamada que componen la distribución  $H_2$ ) para comparar resultados de aproximaciones para la cola  $M/G/s$ , dados los dos primeros momentos de la distribución  $G$  con coeficiente de variación del tiempo de servicio mayor que la unidad [KIM94, SEE85a, TIJ86, HOR84]. Cuando se dispone de los tres primeros momentos de la distribución del tiempo de servicio y siempre asumiendo un coeficiente de variación del tiempo de servicio mayor que uno, puede utilizarse la distribución  $H_2$  no balanceada como comparación [MYS90]. Desde esta perspectiva podía haberse mencionado en el título la cola  $M/G/s$  en lugar de la  $M/H_2/s$ , sin embargo hemos creído más riguroso mencionar la cola en base a la cual se realizan las comparaciones dejando al lector la posible generalización a colas  $M/G/s$ .

El estudio realizado en este capítulo no contempla prioridades. La aproximación se presenta en dos etapas: primero se considera una distribución del tiempo de servicio hiperexponencial balanceada completamente definida por dos momentos, y luego se amplía la aproximación a una distribución hiperexponencial-2 cualquiera definida por los tres primeros momentos.

## 5.1 Interés de la cola $M/H_2/s$ en comunicaciones móviles

El estudio de la cola  $M/H_2/s$  tiene dos aplicaciones bien diferenciadas en el terreno de las comunicaciones móviles:

- Tal como se desprende de las medidas analizadas en el capítulo 4 y de datos de la bibliografía [BAR96b] son diversos los sistemas en los que el coeficiente de variación del tiempo de servicio es mayor que la unidad. La estadística de los mismos no tiene por que ajustarse a una distribución hiperexponencial-2 estricta, y de hecho puede que ajuste mejor a otras distribuciones tal como queda comprobado en el capítulo 4. Sin embargo y tal como se ha mencionado en la introducción a este capítulo, los resultados obtenidos calculando el GoS como si dicha distribución fuese  $H_2$  serán en general muy buenos, especialmente si los comparamos a los que se obtendrían de la cola  $M/M/s$ .
- Cada vez es más habitual la integración de distintos servicios sobre un mismo conjunto de servidores tal como se ha visto en la sección 4.6. En telefonía móvil donde la velocidad de transmisión del medio está lejos de alcanzar las velocidades de los medios físicos como la fibra óptica, cabe contemplar la integración de voz y datos, y especialmente la de voz con mensajes cortos. En estos casos, si suponemos que la duración de cada uno de los dos servicios por separado está distribuida exponencialmente con diferente duración media, tenemos que la duración global está distribuida conforme a una hiperexponencial. Las probabilidades o pesos de cada clase de llamada serán las que resulten de sus tasas de llegadas. Esta circunstancia aplica también a otras combinaciones de servicios como llamadas de despacho e interconexión en sistemas troncales, o voz y fax en telefonía móvil pública.

De este modo las soluciones a la cola  $M/H_2/s$  pueden ser utilizadas como aproximación a la  $M/G/s$  tal como se propone en la introducción, y así podrán aplicarse a aquellos sistemas en los que disponemos de medidas de la media y coeficiente de variación del tiempo de servicio, sin considerar la forma exacta la distribución que subyace a las medidas tomadas. En otros casos la cola en estudio será útil para dimensionar sistemas en los que en base a hipótesis razonables el tiempo de servicio resulte distribuido conforme a una hiperexponencial-2. Los casos en los que dos servicios se integran en una misma red en base a un solo conjunto de servidores resultan efectivamente en una distribución del tiempo de servicio  $H_2$  si cada uno está distribuido exponencialmente con distinta media. Si tratamos la integración de tres o

más servicios estando la duración de cada uno de ellos distribuida exponencialmente con distinta media, ya no sirve el modelo  $M/H_2/s$ . En tal caso resulta una hiperexponencial- $k$ , de la que se pueden calcular los dos o tres primeros momentos y calcular los resultados para la distribución hiperexponencial-2 con los mismos momentos tal como se propone más adelante. De este modo estamos realizando la aproximación de la cola  $M/G/s$  por la  $M/H_2/s$ , tal como se propone en el primer párrafo de este apartado.

Tráfico	$c^2$	Fuente	Observaciones
Transmisión PAMR	1,4		Ver sección 4.2
Transmisión PAMR	2,3	[COH84]	Máxima de diversas muestras sobre monocanal
Mensajes PAMR	2,4	[COH84]	Máxima de diversas muestras sobre monocanal
Celular	1,7		Ver sección 4.5
Celular	2,9-4,1	[JED96]	De la distribución ajustada; para diversas series
Ráfagas (ITU-T)	2,5	[LEE86]	Ver sección 4.7
Ráfagas (PAMR)	2,7-3,1	[HAC83]	Ver sección 4.7
Telefonía fija	2,8	[BOL94]	Una línea telefónica individual
Telefonía fija	4,0	[BOL94]	Grupo de enlaces homogéneo (zona residencial)
Telefonía fija	4,6	[BOL94]	Grupo de enlaces no homogéneo

**Tabla 5.1** Algunos coeficientes de variación al cuadrado para distintos tipos de tráfico de telecomunicación.

Además del campo de las telecomunicaciones móviles, la cola  $M/H_2/s$  y en general la  $M/G/s$  con coeficiente de variación mayor que uno encuentran aplicaciones en otros aspectos de la ingeniería de teletráfico y de la investigación operativa en general. Es de destacar la tendencia actual a modelar la duración de la ocupación de los enlaces de telefonía fija en base a distribuciones hiperexponenciales y especialmente a distribuciones de “cola pesada” (“heavy-tailed”) [BOL94a, BOL94b, DUF94]. Estas últimas distribuciones decaen conforme a una potencia negativa de la variable temporal (más lentamente que las exponenciales) e incluyen a las logonormales y combinaciones de las mismas utilizadas en otros capítulos de este mismo trabajo para modelar tráficos. La tabla 5.1 refleja algunos de los coeficientes de variación del tiempo de ocupación del canal que se pueden encontrar tanto en telefonía móvil como fija.

## 5.2 Revisión de algunas aproximaciones existentes

En la literatura son abundantes y diversas las formas que se proponen para calcular el tiempo medio de espera en colas  $M/H_2/s$  y  $M/G/s$ . Ya que no es el propósito de esta tesis el proporcionar una lista exhaustiva de las mismas, describiremos brevemente las más destacadas, con especial énfasis en las que tienen aplicación en el entorno de ingeniería que consideramos. En dicho entorno son importantes tanto la exactitud de la aproximación como la sencillez de cálculo. Tal como se ha comentado en otros puntos de esta tesis, la aproximación deja de ser interesante cuando la complejidad de los cálculos requeridos exige más tiempo que las simulaciones necesarias para obtener buenos resultados. El orden que seguimos en la exposición es en base a complejidad de cálculo decreciente: la complejidad de cálculo es más fácil de discernir que la exactitud (que sería tal vez más adecuada para establecer un orden) debido a que la exactitud depende en gran medida de los valores de carga, número de servidores considerados y coeficiente de variación del tiempo de servicio. No conocemos ningún estudio de sensibilidad objetivo y lo suficientemente exhaustivo que compare las diversas aproximaciones existentes de una forma exhaustiva. Sin embargo debería existir una obvia correlación entre la complejidad de cálculo y exactitud obtenida, ya que si una aproximación es menos exacta y más compleja que otras, los ámbitos en que sea aconsejable su aplicación serán muy escasos.

Existe una solución analítica exacta para la distribución del tiempo de espera en la cola  $G/H_m/s$  [SMI83a, SMI83b]. Esta solución permite no solo calcular el retardo medio de acceso, sino también los percentiles de llamadas demoradas por encima de cierta cota tal como recomienda la ITU-R [CCI90]. Sin embargo el algoritmo de cálculo de dicha solución es extremadamente complejo y la capacidad de cómputo requerida muy elevada. Así es habitual en la literatura utilizar para la obtención de resultados exactos con los que comparar las aproximaciones propuestas los valores tabulados en [SEE85a] para el caso de hiperexponencial balanceada [KIM94], o resultados de simulación cuando la distribución no es balanceada [MYS90] dada la ausencia de resultados tabulados para este último caso. Este es el procedimiento que se utiliza a lo largo de este trabajo.

La aproximación debida a Boxma, Cohen y Huffles [BOX80] es una excelente estimación del tiempo medio de espera en la cola  $M/G/s$  que tiene en cuenta la distribución exacta de la ocupación del servidor. En el artículo citado solamente se realizan comparaciones para distribuciones hipoexponenciales ( $c^2 < 1$ ) y se conjetura que la aproximación mantiene su validez para coeficientes de variación de la duración

del servicio mayores que la unidad. En [TIJ86] pueden encontrarse detalles prácticos sobre la utilización de esta aproximación para una distribución general, a la vez que su aplicación a distribuciones con coeficiente de variación mayor que uno. La notación que se utiliza en el resto de capítulo es la siguiente:

$$b(t) = q\mu_1 e^{-\mu_1 t} + (1-q)\mu_2 e^{-\mu_2 t} = \text{f.d.p. de la duración del servicio,}$$

$$m_1 = 1/\mu = q/\mu_1 + (1-q)/\mu_2 = \text{duración media del servicio,}$$

$$k = \frac{m_2}{2m_1^2} = \frac{c^2 + 1}{2} = \text{segundo momento relativo,}$$

$$W(M/G/s) = \text{tiempo medio de espera en la cola M/G/s,}$$

$$R_G = \frac{W(M/G/s)}{W(M/M/s)} = \text{tiempo medio de espera relativo,}$$

$$r = \mu \frac{q}{\mu_1} = \text{proporción de tiempo de las llamadas cortas.}$$

Así el parámetro  $r$  implica el tercer momento de la distribución del tiempo de servicio y será utilizado en su lugar en esta tesis por razones prácticas. La utilización y el cálculo de la aproximación de Boxma con una distribución del tiempo de servicio hiperexponencial-2 tal como se detalla en [HOO84] es más simple que con una distribución general, ya que el cálculo del parámetro  $\gamma_1$  de la ecuación (5.2) queda reducido a una suma finita. El tiempo medio de espera en la cola M/H<sub>2</sub>/s puede ser aproximado según [BOX80] y [HOO84] por:

$$W(M/H_2/s) = \frac{2kW(M/D/s)W(M/M/s)}{2aW(M/D/s) + (1-a)W(M/M/s)} \quad (5.1)$$

$$a = \frac{1}{s-1} \left( \frac{m_2}{\gamma_1 m_1} - s - 1 \right)$$

$$\gamma_1 = \int_0^\infty \{1 - B_e(t)\}^s dt = \sum_{i=0}^s \frac{\binom{s}{i} r^i (1-r)^{s-i}}{i\mu_1 + (s-i)\mu_2} \quad (5.2)$$

donde  $B_e(t)$  representa el tiempo residual de servicio:

$$B_e(t) = \frac{1}{m_1} \int_0^t \{1 - B(x)\} dx, \quad t \geq 0$$

La ecuación (5.1) puede escribirse de forma más compacta utilizando el tiempo medio de espera relativo  $R_{H2}(r)$ . Esta relación depende de  $r$  entre otros parámetros como la carga  $\rho$ , el número de servidores  $s$  y el coeficiente de variación del tiempo de servicio  $c$ . La dependencia con  $r$  es puesta de manifiesto explícitamente cuando dicha dependencia existe, ya que en otras aproximaciones que se detallan a continuación el parámetro  $r$  no influye en los cálculos (estas últimas solo consideran los dos primeros momentos de la distribución del tiempo de servicio).

$$R_{H2}(r) = \frac{2kR_D}{2aR_D + (1-a)} \quad (5.3)$$

La aproximación lineal presentada en [SEE86] se ha utilizado en el capítulo 3 particularizada para una situación de baja carga. La ecuación (3.3) se puede utilizar en general y puede escribirse del siguiente modo:

$$R_{H2}(r) = (1-\rho)\gamma_1 s \mu + \rho k \quad (5.4)$$

Esta aproximación implica la misma complejidad de cálculo que la anterior al ser necesario el cálculo de  $\gamma_1$  y sin embargo es menos exacta.

La aproximación debida a Kimura [KIM94] fue introducida también en el capítulo 3, y la ecuación (3.16) puede escribirse como:

$$R_G = \frac{2kR_D}{2c^2 R_D + (1-c^2)} \quad (5.5)$$

Esta aproximación no implica el tercer momento: observar que la notación  $R_G$  no indica la dependencia con  $r$ , y que la G representa que en este caso la ecuación no está particularizada para el caso de distribución hiperexponencial-2 como si lo estaba la aproximación de Boxma. La aproximación de Kimura es más sencilla de calcular que las dos anteriores, pero al no implicar el tercer momento proporcionará el mismo resultado cualquiera que sea el valor de este último, no siendo por tanto aplicable a casos extremos del parámetro  $r$  ( $r \rightarrow 0$  ó  $r \rightarrow 1$ ). Las aproximaciones de Boxma y Kimura coinciden para  $a = c^2$ , tal como se deduce de las ecuaciones (5.1) y (5.3).

Por último presentamos la aproximación debida a Cosmetatos [COS76] que es la de cálculo más simple a la vez que la menos exacta:

$$R_G = (1 - c^2)R_D + c^2 \quad (5.6)$$

Esta interpolación lineal con respecto al cuadrado del coeficiente de variación es exacta para los valores  $c=0$  y  $c=1$  que representan las distribuciones determinística y exponencial respectivamente, pero resulta muy burda para  $c^2 > 2$ . En [SEE85] se aplicó la ecuación (5.6) al cálculo de los percentiles de servicios demorados más que una cierta cota con buenos resultados, pero siempre para casos en los que  $c^2 < 2$ .

En tres de las cuatro aproximaciones referidas aparece como necesario el cálculo del parámetro  $R_D$ . El cálculo de dicho parámetro de forma exacta, con la complejidad que implica, no es adecuado para introducir el valor obtenido en una fórmula aproximada. Por ello es recomendable utilizar la excelente aproximación de la ecuación (3.2). Los sucesivos cálculos realizados en esta tesis de las aproximaciones referidas, así como las comparaciones entre los resultados que éstas proporcionan incluyen esta última aproximación. Debe entenderse con cuidado el hecho de que la precisión que se obtiene en este trabajo para las distintas aproximaciones pueda verse mermada por este hecho, con respecto a la que puedan haber calculado los propios autores. De cualquier modo el cálculo de las aproximaciones referidas no debiera pasar por la exigencia del cálculo de un resultado exacto de cómputo casi tan sofisticado como el resultado que se intenta aproximar. De cualquier modo hay que dejar constancia de que las comparaciones que se presentan más adelante están basadas en un uso eminentemente práctico de las aproximaciones referidas.

Hay que citar también la aproximación propuesta por Ma y Mark [MA95] que resulta apta para valores muy altos del coeficiente de variación y del número de servidores (coeficientes de variación al cuadrado de hasta 20 y hasta 200 canales). Sin embargo el cálculo de esta aproximación resulta mucho más complejo que el de cualquiera de las que hemos analizado anteriormente. Por otra parte esta aproximación necesita ser alimentada con el resultado obtenido para un número mucho menor de canales: los autores proponen obtener el resultado exacto para 10 canales. Por lo expuesto, el uso de esta aproximación no es en absoluto práctico en el entorno de comunicaciones móviles.

### **5.3 Aproximación propuesta para la cola M/H<sub>2</sub><sup>b</sup>/s**

La aproximación propuesta en esta tesis es heurística y basada en tests numéricos, de modo que no existe una demostración matemática de la misma. De cualquier modo está basada en dos conceptos intuitivos: una cota superior para el



tiempo medio de espera en la cola  $M/H_2/s$ , y la idea de simetría entre las distribuciones hiperexponencial-2 balanceada y la determinística con respecto a la distribución exponencial.

Si los dos primeros momentos de una distribución hiperexponencial-2 están fijados, una situación límite se produce cuando la duración de las llamadas cortas tiende a cero ( $\mu_1 \rightarrow \infty$ ). En este caso  $r \rightarrow 0$  y para mantener los dos primeros momentos de la distribución debe ocurrir:

$$\frac{1}{\mu_2} = km_1 \qquad 1 - q = \frac{1}{k} \qquad (5.7)$$

Así el sistema no ve las llamadas con subíndice 1 de tipo corto y solamente ve las llamadas largas, que deben tener una duración igual a  $k$  veces la media de la distribución  $H_2$  considerada. La probabilidad de que dichas llamadas ocurran es de  $1/k$ . Este proceso es de hecho una cola  $M/M/s$  con el mismo tráfico ofrecido y únicamente llamadas de duración media  $km_1$  y distribución exponencial negativa. La cota superior del tiempo medio de espera en una cola  $M/H_2/s$  es:

$$W(M/H_2/s) = Erlang-C(A, s) \frac{km_1}{s - A} = kW(M/M/s)$$

$$R_{H_2}(0) = k \qquad (5.8)$$

El concepto de simetría entre las distribuciones determinística e hiperexponencial-2 balanceada conduce a la siguiente aproximación [BAR97c]:

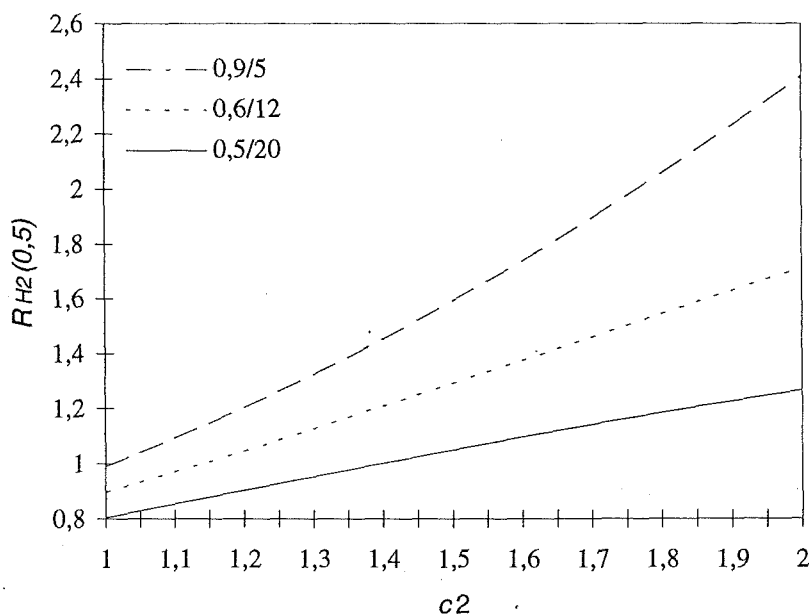
$$R_{H_2}(0,5) = \frac{k}{1 + \frac{(k - 0,5)(1 - \rho)(s - 1)(\sqrt{4 + 5s} - 2)}{16\rho s}} \qquad (5.9)$$

A pesar de la naturaleza heurística de la ecuación (5.9) deben realizarse las siguientes consideraciones:

- El concepto de simetría conduce a utilizar la aproximación de Cosmetatos de la ecuación (3.2) en "sentido inverso". Es notable la semejanza entre la ecuación (3.2) para la aproximación de  $R_D$  y la ecuación (5.9).
- Como sea que existe una sola distribución determinística e infinitas hiperexponenciales-2 balanceadas (una para cada coeficiente de variación) es

necesaria la normalización respecto al parámetro  $k$ , que de hecho representa el segundo momento en la ecuación (5.9).

- El factor  $(k-0,5)$  está basado únicamente en los resultados de tests numéricos.
- A diferencia del resto de aproximaciones presentadas, la nueva aproximación no es exacta cuando el coeficiente de variación tiende a uno, y de hecho proporciona un resultado inconsistente: el retardo medio estimado resulta inferior para la distribución  $H_2$  que para la distribución exponencial. Nótese que la ecuación (3.2) en la cual está basada la (5.9) presenta también resultados inconsistentes para cargas bajas. En la figura 5.1 se representan valores de la ecuación (5.9) notándose que para probabilidades de demora muy bajas el resultado es inferior a 1 si el coeficiente de variación es también bajo.
- Debido a la consideración anterior, la aproximación propuesta no debe ser utilizada para coeficientes de variación cercanos a uno o probabilidades de demora bajas. Estas últimas son siempre debidas a una combinación de baja carga y número de servidores elevado.



**Figura 5.1** Estimación de la ecuación (5.9) para distintas situaciones de carga y número de canales. Nótese que 20 canales a carga de 0,5 dan un tiempo medio de espera igual a 0,0004 veces el tiempo medio de ocupación del canal.

### 5.4 Resultados comparativos para la cola $M/H_2^b/s$

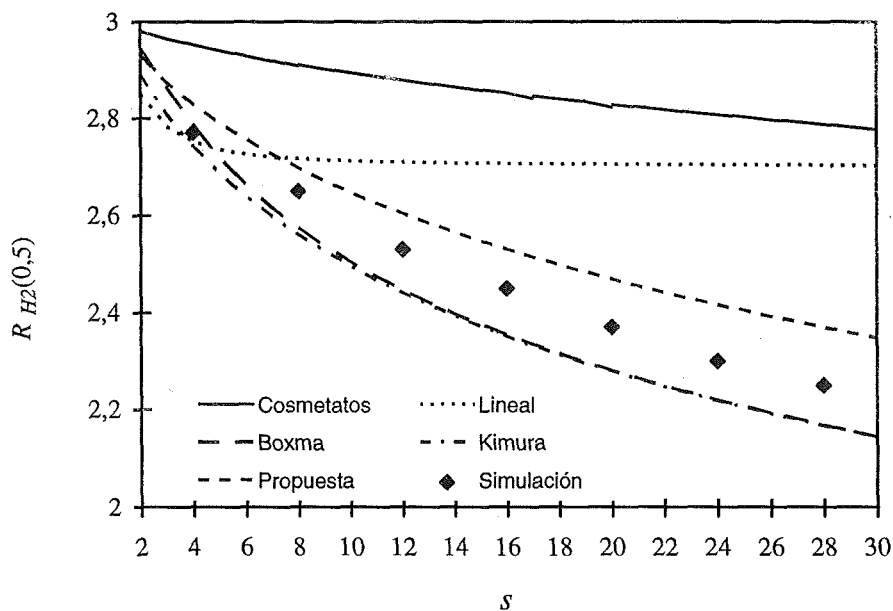
En la sección 5.2 se han presentado cuatro aproximaciones para el cálculo del retardo medio. Para la distribución  $H_2^b$ , las aproximaciones de Boxma y Kimura presentan resultados casi idénticos, si bien la de Boxma requiere una mayor sofisticación computacional. De hecho la distribución  $H_2^b$  queda completamente descrita mediante 2 parámetros, los mismos que necesita la aproximación de Kimura. Por otro lado se ha mencionado en la sección 5.2 y se comprobará más adelante que las otras dos aproximaciones proporcionan resultados menos exactos. Tanto la aproximación propuesta como la de Kimura pueden ser computadas de forma práctica sin necesidad de programar ningún cálculo, de modo que la de Kimura aparece como una aproximación adecuada con la que contrastar la fórmula propuesta en la ecuación (5.9).

	s= Carga	5			12			20		
		Exacto	Kim.	Prop.	Exacto	Kim.	Prop.	Exacto	Kim.	Prop.
$c^2=1,5$	0,6	1,17	1,19	1,15						
	0,7	1,20	1,22	1,20	1,16	1,17	1,13	1,13	1,15	1,08
	0,8	1,22	1,23	1,22	1,19	1,20	1,17	1,17	1,19	1,15
	0,9	1,23	1,24	1,23	1,22	1,23	1,22	1,21	1,22	1,20
$c^2=2,5$	0,6	1,51	1,52	1,53						
	0,7	1,58	1,59	1,60	1,44	1,47	1,48	1,35	1,39	1,39
	0,8	1,64	1,65	1,66	1,55	1,56	1,57	1,48	1,50	1,51
	0,9	1,70	1,70	1,71	1,65	1,66	1,67	1,62	1,63	1,64
$c^2=4$	0,6	1,97	1,91	2,04						
	0,7	2,12	2,09	2,19	1,84	1,81	1,92	1,66	1,62	1,71
	0,8	2,26	2,23	2,30	2,06	2,02	2,13	1,92	1,89	2,01
	0,9	2,39	2,37	2,40	2,29	2,25	2,32	2,21	2,17	2,26

**Tabla 5.2** Resultados comparativos de  $R_{H_2}(0,5)$  para diferentes situaciones de carga, número de canales y coeficiente de variación.

La tabla 5.2 compara valores exactos del tiempo medio de espera en la cola  $M/H_2^b/s$  utilizando como fuente [SEE85a] con los obtenidos mediante la aproximación de Kimura y la propuesta en esta tesis. Para  $c^2=4$  el error cometido por ambas estimaciones es similar, pero mientras que la estimación de Kimura subestima

el retardo, la nuestra lo sobreestima. Esta es una ventaja importante en entornos de ingeniería en los que en la mayoría de los cálculos es preferible una visión pesimista. Para  $c^2=2,5$  ambas estimaciones sobreestiman el retardo mientras que la aproximación de Kimura es ligeramente más exacta. Los casos de baja carga no han sido considerados ya que las probabilidades de demora y retardos que se obtienen son extremadamente bajos y las evaluaciones de teletráfico en el tipo de sistemas que nos concierne deben realizarse siempre con cargas altas. Para  $c^2=1,5$  la aproximación propuesta es todavía correcta para cargas altas, pero si la carga es baja o el número de servidores elevado subestima los retardos.



**Figura 5.2** Tiempo medio de espera relativo para  $c^2=5$  y carga del 85% con número de canales variable.

En la figura 5.2 se presentan los valores de  $R_{H2}(0,5)$  para cada una de las cuatro aproximaciones presentadas en la sección 5.2, así como para la nueva estimación propuesta. Las condiciones en las que se ha realizado la gráfica son una carga del 85% y coeficiente de variación  $c^2=5$ , variando el número de servidores  $s$ . Como referencia se muestran resultados de simulación, ya que para este coeficiente de variación no hay resultados tabulados disponibles. Debe notarse que aunque en [SEE85a] solamente aparecen valores tabulados para coeficientes de variación al cuadrado de hasta  $c^2=4$ , la realidad de los sistemas reflejada en la tabla 5.1 muestra valores de hasta 4,6 para este parámetro. En la figura 5.2 puede observarse como para un número de servidores elevado las aproximaciones de Boxma y Kimura coinciden

subestimando ligeramente el retardo, mientras que la lineal y la de Cosmetatos lo sobrestiman excesivamente. La precisión de la aproximación presentada es similar a las de Boxma y Kimura con la ventaja de sobrestimar en lugar de subestimar el retardo medio de acceso.

La figura 5.3 muestra el crecimiento del tiempo medio de espera relativo en la cola  $M/H_2^b/s$  en función del cuadrado del coeficiente de variación de la duración del servicio, para un sistema con 12 servidores y carga del 85%. Puede observarse como las estimaciones de Cosmetatos y lineal crecen de forma lineal con respecto al cuadrado del coeficiente de variación. Para coeficientes de variación muy grandes el error cometido por las cuatro aproximaciones de la sección 5.2 es del mismo orden de magnitud, mientras que la estimación propuesta presenta un error sensiblemente inferior. En la figura 5.3 se aprecia como la mejor exactitud de las estimaciones de Boxma y Kimura se consigue a costa de un rango de validez inferior.

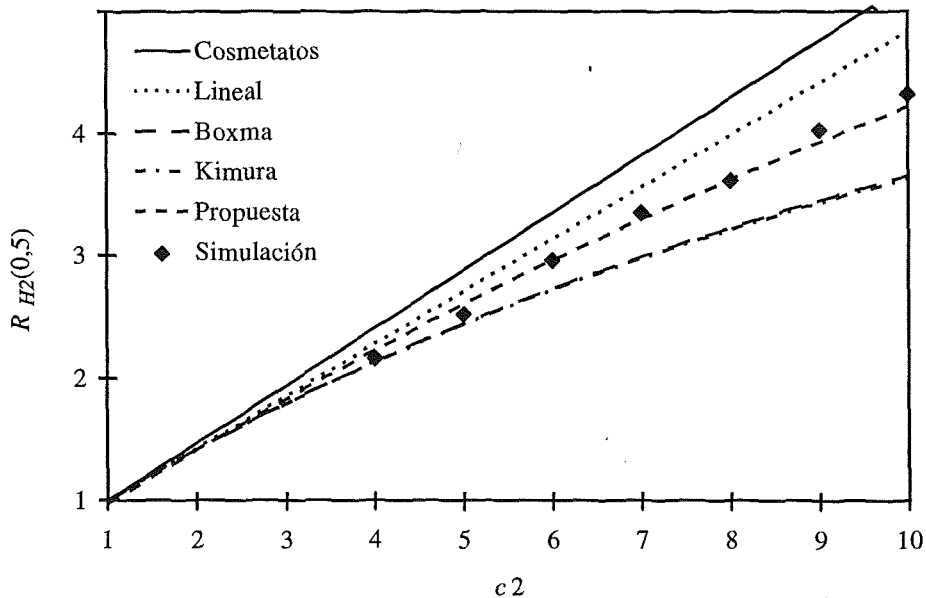


Figura 5.3 Tiempo de espera medio relativo para un sistema con 5 canales y carga del 85%, variando el coeficiente de variación.

## 5.5 Influencia del tercer momento

En las secciones previas se han analizado colas en las que el tiempo de servicio estaba distribuido de acuerdo a una función hiperexponencial-2 balanceada,

la cual queda completamente definida mediante los dos primeros momentos, o lo que es lo mismo, por su duración media y coeficiente de variación. De cara a generalizar la aproximación propuesta a cualquier distribución hiperexponencial-2 de la duración del servicio, nótese que en la sección 5.3 se presentan:

- El valor exacto de  $R_{H2}(0)=k$  para el caso extremo considerado de  $r \rightarrow 0$ , en el cual la duración media de las llamadas de tipo corto es extremadamente baja.
- El valor aproximado de  $R_{H2}(0,5)$  en la ecuación (5.9).

Después de probar numéricamente diferentes opciones (interpolaciones lineales y exponenciales sobre diferentes parámetros que representan el tercer momento de la distribución hiperexponencial-2), se ha comprobado que la opción que ofrece un mejor resultado es la interpolación lineal con respecto a la proporción  $r$  para los dos puntos citados:  $r=0$  y  $r=0,5$ . Esta aproximación puede escribirse después de algunas manipulaciones algebraicas como [BAR97c, d]:

$$R_{H2}(r) = k \left( 1 - \frac{2(k-0,5)(1-\rho)(s-1)(\sqrt{4+5s}-2)}{16\rho s + (k-0,5)(1-\rho)(s-1)(\sqrt{4+5s}-2)} r \right) \quad (5.10)$$

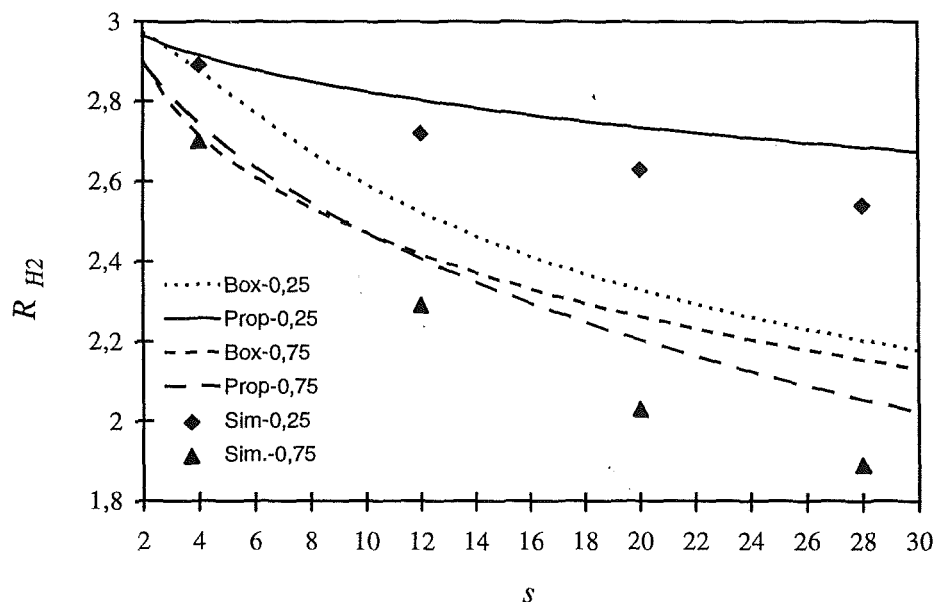
Cuando la proporción de llamadas cortas  $r$  tiende a uno estamos ante otro caso extremo, para el cual la duración media de las llamadas largas tiende a infinito ( $\mu_2 \rightarrow 0$ ) y estas se producen con una probabilidad que tiende a cero. Esta situación es la contraria de la expuesta en la sección 5.3 y para mantener los dos primeros momentos de la distribución todas las llamadas deben ser del tipo corto ( $\rho \rightarrow 1$ ). Este proceso es también una cola M/M/s y el tiempo medio de espera relativo es por tanto  $R_{H2}(1)=1$ . Es fácil comprobar que la ecuación (5.10) propuesta proporciona un valor exacto para  $r \rightarrow 0$  y coincide con la ecuación (5.9) para  $r=0,5$ , sin embargo el resultado será incorrecto para  $r \rightarrow 1$ .

Esta última situación y en general valores altos de  $r$ , aunque son matemáticamente posibles resultan muy difíciles de encontrar en problemas reales de evaluación, ya que implican que el conjunto de servidores dedica la mayor parte del tiempo a servir las llamadas de tipo corto. Para ello la proporción de dichas llamadas debe ser extremadamente alta. Este efecto es amplificado cuando crece el coeficiente de variación dada la mayor dispersión entre las duraciones medias de ambas clases de servicios. Las aplicaciones de colas con distribución hiperexponencial-2 de la duración del servicio y proporción alta de llamadas cortas  $r$  se encuentran más bien en estudios de supervivencia ante fallos: el servicio de tipo corto representa la llamada y

el de tipo largo representa el fallo con duración muy grande y probabilidad de ocurrencia extremadamente baja frente a la de la llamada.

## 5.6 Resultados comparativos

Para contrastar la bondad de la estimación propuesta en la ecuación (5.10) utilizaremos la aproximación de Boxma de la ecuación (5.3) como la que proporciona una mejor aproximación. La aproximación de Kimura de la ecuación (5.5) no incluye el tercer momento, razón por la cual no cabe esperar que proporcione buenos resultados cuando éste es considerado: obviamente la aproximación de Kimura empeorará para casos extremos ( $r \rightarrow 0$  ó  $r \rightarrow 1$ ). Por otro lado la aproximación lineal de la ecuación (5.4) da una peor aproximación. De este modo la aproximación de Boxma, aunque de cómputo más complejo que la propuesta, es la más apropiada para contrastar la fórmula presentada en esta tesis.



**Figura 5.4** Resultados comparativos para carga del 85% y  $c^2=5$ . Las proporciones de tiempo de servidor dedicadas a "llamadas cortas" son del 25% y 75%.

Después de numerosos tests numéricos en diferentes situaciones de carga, coeficiente de variación y proporción, la aproximación presentada aquí ha proporcionado excelentes resultados. La figura 5.4 muestra resultados comparativos (resultados de simulación frente a las aproximaciones de Boxma y propuesta) para una carga del 85% y un coeficiente de variación al cuadrado de  $c^2=5$  variando el

número de servidores o canales para dos valores diferentes de  $r=0,25$  y  $r=0,75$ . Para  $r=0$  la aproximación debe ser exacta dado su carácter de interpolación, mientras que la precisión para  $r=0,5$  ha sido analizada ya en la sección 5.4 (como proceso de cola  $M/H_2^b/s$ ). Las conclusiones a las que lleva el análisis de los tests numéricos son similares a las obtenidas para el caso de distribuciones balanceadas: la aproximación es muy precisa y tiende a sobreestimar el tiempo medio de espera en cola.

En la figura 5.5 se reflejan los tiempos medios de espera relativos en dos situaciones diferentes de carga y número de canales para un mismo coeficiente de variación ( $c^2=3$ ). Puede observarse también la precisión de la aproximación y como ésta sobreestima el valor del tiempo medio de espera. Los valores utilizados son:

caso a) 5 canales y carga del 60%

caso b) 20 canales y carga del 90%

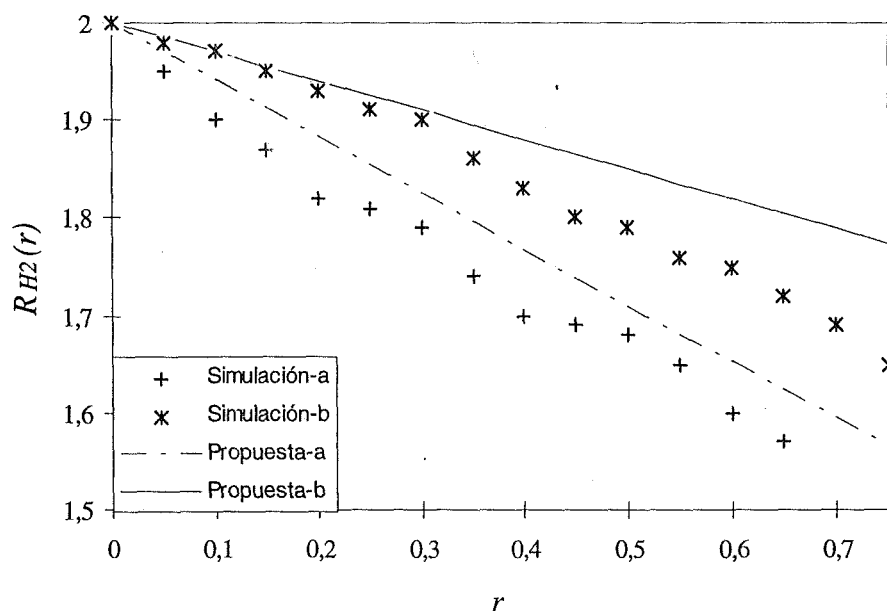
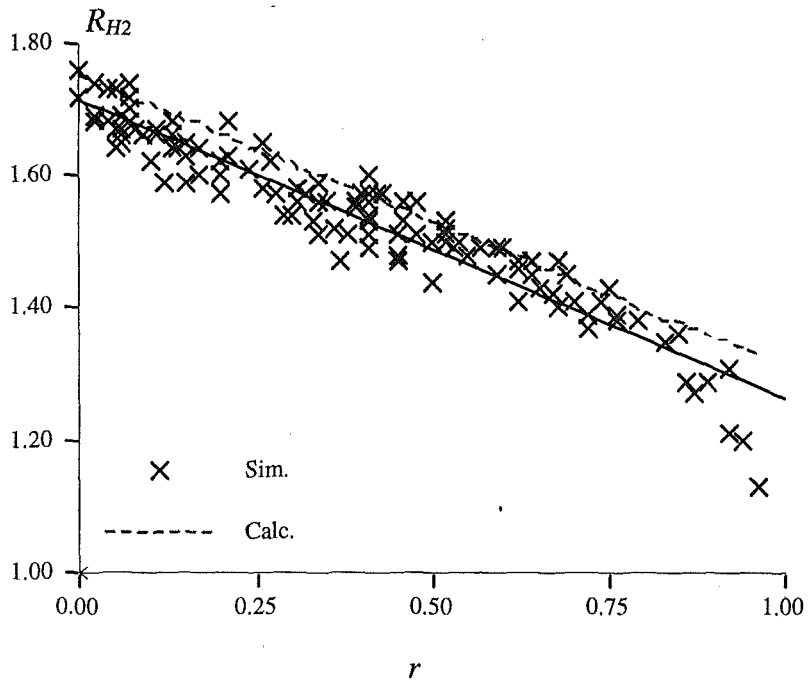


Figura 5.5 Tiempo medio de espera relativo en la cola  $M/H_2/s$  para  $c^2=3$

La figura 5.6 es análoga a la figura 5.5 excepto en el hecho de que se representan 100 resultados de simulación obtenidos a partir de una distribución uniforme de la probabilidad de llamada corta  $q$  (distribución uniforme o rectangular entre el mínimo que puede tomar  $q = 1 - \frac{1}{k}$  de la ecuación (5.7) y  $q = 1$ ). Puede apreciarse como la densidad de puntos de simulación para  $r$  alto ( $r \rightarrow 1$ ) es baja tal



como se describe al final de la sección anterior. Este efecto aumenta con el aumento del coeficiente de variación. Por tanto la fórmula presentada falla sobreestimando el retardo de acceso, pero en aquellos casos más difíciles de encontrar en la realidad. La línea continua muestra la regresión lineal resultante de todos los valores de simulación como referencia de la bondad de la aproximación.



**Figura 5.6** Resultados de  $R_{H2}$  para 5 canales con carga del 60% y coeficiente de variación al cuadrado de 2,5. Nótese que son raros los casos de proporción  $r$  cercana a 1.

## 5.7 Sumario

En muchos casos de ingeniería de teletráfico, y más concretamente en comunicaciones móviles, pueden encontrarse colas  $M/H_2^b/s$ ,  $M/H_2/s$  o  $M/G/s$  con coeficiente de variación del tiempo de servicio mayor que la unidad pudiendo llegar dicho coeficiente a valores altos tal como se refleja en la tabla 5.1. En general no es necesaria una solución exacta para obtener parámetros del GoS ya que las entradas al sistema (especialmente el tráfico ofrecido o la carga) son solamente conocidos de forma aproximada. En estas situaciones son altamente útiles aproximaciones que ofrezcan a la vez precisión en los resultados y simplicidad de cálculos. Respecto a las aproximaciones analizadas en este capítulo podemos concluir que:

- Las aproximaciones lineal y de Boxma consideran los tres primeros momentos de la distribución del tiempo de servicio con una complejidad de cálculo similar, aunque la de Boxma proporciona resultados más exactos.
- Las aproximaciones de Cosmetatos y Kimura consideran solamente los dos primeros momentos con una complejidad de cálculo similar, mientras que la de Kimura proporciona resultados más precisos.
- Todas las aproximaciones citadas son también válidas para coeficientes de variación menores que la unidad.
- La exactitud de todas las aproximaciones referidas se reduce al utilizar en ellas un valor aproximado de  $R_D$ . La utilización de un valor exacto que mejoraría los resultados no es en absoluto práctica.

Teniendo en cuenta estas consideraciones y para coeficientes de variación al cuadrado inferiores a 2, puede recomendarse la aproximación de Kimura si solamente se conocen dos momentos de la distribución del tiempo de servicio. Cuando son conocidos los tres primeros momentos de dicha distribución puede recomendarse la aproximación de Boxma, que aporta una mejor aproximación a costa de añadir complejidad de cálculo (Kimura presenta otra excelente aproximación para este caso en [KIM-94]).

Cuando el cuadrado del coeficiente de variación del tiempo de servicio es mayor que 2 todas las aproximaciones referidas empiezan a dar resultados poco exactos. En este caso las ecuaciones (5.9) y (5.10) se recomiendan según sean conocidos los dos o tres primeros momentos de la distribución del tiempo de servicio respectivamente. Incluso en el caso de tres momentos de la distribución conocidos, la complejidad de cálculo no es mayor que en las aproximaciones de Kimura y Cosmetatos.