

**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 Aspas

CAPÍTULO 3

Aproximaciones para colas M/G/s con distribución hipoexponencial del tiempo de servicio y prioridad

Tal como se ha visto en el capítulo anterior, en los sistemas PAMR que cursan solamente tráfico vocal y que realizan la gestión por mensaje, encontramos que el tiempo de ocupación del canal tiene un coeficiente de variación menor que la unidad, con lo cual puede ser caracterizado mediante funciones de distribución de tipo hipoexponencial. El retardo medio que sufrirán las llamadas en este caso será inferior al calculado mediante la ecuación (1.2) debido a la mayor regularidad en el tiempo de servicio con respecto a la cola M/M/s.

El objetivo de este capítulo es proporcionar aproximaciones para el retardo medio de acceso que sufren ambas categorías de llamadas, prioritarias y no prioritarias, en un sistema PAMR, dados el tráfico ofrecido por las llamadas de cada nivel de prioridad y el número de canales s disponibles. Se ha considerado únicamente el caso de prioridad HOL y dos niveles de prioridad si bien los resultados pueden ser generalizados a sistemas con más niveles. Para otro parámetro indicativo del GoS, la probabilidad de demora PD , la aproximación de la ecuación (1.1) es muy buena cualquiera que sea la distribución del tiempo de servicio mientras el proceso de llegadas siga siendo de Poisson tal como ya se ha comentado [TIJ86]. En cuanto a otros parámetros sugeridos por la ITU-R, tales como los percentiles de llamadas que superan cierto retardo, no estamos en condiciones de ofrecer aproximaciones. Siempre es posible aplicar desigualdades estadísticas como la de Markov [KLE75] para acotar dichos percentiles, aunque al trabajar con cotas superiores corremos el riesgo de sobredimensionar el sistema.

3.1 Aproximaciones del tiempo medio de espera en la cola M/D/s

En esta sección asumimos todas las hipótesis del sistema M/M/s presentado en la figura 1.3, excepto la de que el tiempo de servicio está distribuido exponencialmente, la cual sustituimos por la hipótesis de que todas las llamadas tienen la misma duración fija. De este modo la media de la duración de la llamada es igual a la duración de cada una de ellas, y tanto la varianza como el coeficiente de variación son nulos. Se asume también que la cola es de capacidad ilimitada (hipótesis de cola infinita). Difícilmente encontraremos un sistema de telefonía móvil en el que la duración de la llamada sea fija y al que podamos aplicar por tanto el modelo M/D/s, sin embargo la finalidad del estudio de este modelo persigue los siguientes puntos:

- Es útil como cota inferior del retardo medio de acceso, dado que la distribución determinística del tiempo de servicio es el mejor caso que puede darse [BAR95]. Si bien desde un punto de vista de ingeniería son en general preferibles las cotas superiores (peor caso), este resultado puede ser útil para conocer el mínimo número de canales sin el cual no es posible implementar el sistema.
- Tal como se verá en los apartados que siguen, este estudio va a servir de punto de partida en la obtención de resultados para colas en las que el tiempo de servicio está distribuido de forma más regular que exponencial (hipoexponenciales).

Existe una solución analítica exacta para el cálculo de las probabilidades de cada uno de los estados del sistema y la distribución del tiempo de espera en cola para la disciplina de cola FCFS (recordemos que el número medio de unidades en la cola está relacionado con el tiempo medio de espera a través de la relación de Little [KLE75]). Por tanto es posible calcular de forma exacta la probabilidad de demora PD , el retardo medio de acceso y los percentiles de que el retardo de acceso sea superior a cierta cota. En [TIJ86] se detallan los procedimientos que se deben seguir para calcular la distribución del tiempo de espera en cola, notando el autor que en ciertos casos (de carga o número de canales altos) los tiempos de cómputo para obtener un grado de exactitud satisfactorio pueden ser muy elevados. En [KIM94] se ofrece la siguiente fórmula para el cálculo del tiempo medio de espera en la cola M/D/s (la notación utilizada es la misma que en el capítulo 1):

$$W(M/D/s) = \frac{1}{\mu} \sum_{j=1}^{\infty} \sum_{k=j+1}^{\infty} \left\{ \frac{(jA)^{k-1}}{(k-1)!} - \frac{(jA)^k s}{Ak!} \right\} \exp(-jA) \quad (3.1)$$

En cualquier caso la complejidad de los cálculos a realizar es muy alta y la convergencia del método propuesto es muy lenta cuando crece el número de canales del sistema s o cuando las cargas son altas. Por esta razón existen tablas conteniendo los valores exactos de las probabilidades de demora y el tiempo medio de espera en cola, así como algunos percentiles del tiempo de espera [SEE85a].

La complejidad de las soluciones exactas para la cola M/D/s ha motivado la búsqueda de soluciones aproximadas de cómputo sencillo. De ellas destacamos la aproximación debida a Cosmetatos [COS76], según la cual la relación entre el tiempo medio de espera en una cola M/D/s y el de una cola igual pero con distribución exponencial del tiempo de servicio es:

$$R_D = \frac{W(M/D/s)}{W(M/M/s)} = \frac{1}{2} \left\{ 1 + \frac{(1-\rho)(s-1)(\sqrt{4+5s}-2)}{16\rho s} \right\} \quad (3.2)$$

donde $\rho = A/s$ representa la carga de tráfico por canal. Por conveniencia hemos introducido la relación R_D que se utilizará frecuentemente en esta tesis. De hecho la mayoría de los investigadores han tratado de relacionar sus aproximaciones para los retardos en colas M/G/s con el retardo de una cola M/M/s con el mismo tráfico y número de canales a través de la relación R_G [KIM94]. Cuando la distribución G es conocida (en este caso una D) la representamos explícitamente en la notación. La aproximación de la ecuación (3.2) es excelente para cargas medias y altas, y además es de cómputo extremadamente sencillo. El resultado obtenido es exacto para el caso de un solo servidor obteniéndose $R_D=0,5$ [COH82] ya que concuerda con la fórmula de Pollaczek-Khinchin (P-K) que proporciona el tiempo medio de espera en la cola M/G/s [KLE75]. También se obtiene el resultado asintótico $R_D \rightarrow 0,5$ [KIM94] para carga cercana al 100% ($\rho \rightarrow 1$). Sin embargo el resultado es absurdo cuando la carga es muy baja o el número de canales es muy alto, ya que aparece $R_D > 1$ tal como puede observarse en las figuras 3.1 y 3.2 en las que se representan los valores del tiempo medio de espera relativo para distintas aproximaciones (recordemos que el retardo medio en la cola M/D/s debe ser siempre menor que en la cola M/M/s equivalente [KIM94]).

En casos de baja carga y ante la imposibilidad de utilizar la ecuación (3.2) debido a su inconsistencia, podemos utilizar una aproximación para la cola M/G/s [SEE86] particularizada al caso M/D/s en [BAR96a]. A diferencia de otras

aproximaciones existentes para la cola M/G/s, ésta no necesita el valor de $W(M/D/s)$ para su cálculo, lo cual la hace apta para obtener precisamente dicho retardo medio. En la notación m_i representa el i -ésimo momento ordinario del tiempo de servicio.

$$R_G = \frac{W(M/G/s)}{W(M/M/s)} = \left\{ (1-\rho)\gamma_1 \frac{s}{m_1} + \rho \frac{m_2}{2m_1^2} \right\} \quad (3.3)$$

para el caso de tiempo de servicio determinístico tenemos [HOR84]:

$\gamma_1 = \frac{m_1}{s+1}$ y $m_2 = m_1^2$ de lo cual con algunas manipulaciones algebraicas se obtiene la siguiente fórmula:

$$R_D = \left\{ \frac{(1-\rho)s}{s+1} + \frac{\rho}{2} \right\} \quad (3.4)$$

que como puede fácilmente observarse es exacta para el caso asintótico de carga cercana al 100% ($\rho \rightarrow 1$) y para el caso de un solo servidor al igual que la anterior aproximación de la ecuación (3.2). Aunque esta estimación pueda resultar poco exacta para determinadas situaciones de carga tal como se aprecia en las figuras 3.1 y 3.2, no es en ningún caso inconsistente. A diferencia de otras aproximaciones presentadas, esta aproximación es exacta cuando la carga tiende a cero (carga muy baja), razón por la cual se va a utilizar más adelante en los cálculos relativos a las llamadas prioritarias (que en las condiciones de interés de estudio del sistema deberían suponer una carga baja). Así para el caso de muy baja carga se obtiene el resultado asintótico [KIM94]:

$$\rho \rightarrow 0 \Rightarrow R_D \rightarrow \frac{s}{s+1} \quad (3.5)$$

Además de las dos aproximaciones citadas para el retardo medio de acceso existen otras en la literatura. Algunas aproximaciones útiles se encuentran en [SHA90] y [KIM94]. En esta última referencia se puede hallar un estudio muy exhaustivo de las aproximaciones del tiempo de espera en colas M/G/s y también uno de los pocos estudios de sensibilidad existentes para aproximaciones al tiempo medio de espera en la cola M/D/s.

En [BAR95] se presenta la una aproximación que, aunque es en general menos precisa que la ecuación (3.2), es consistente para todas las situaciones de carga alta y baja. Esta estimación es exacta para el caso de plena carga (tendente al 100%) pero no

lo es para el caso de un solo servidor ni el de carga muy baja: si $\rho \rightarrow 0 \Rightarrow R_D \rightarrow 1$ cuando el valor exacto debería ser $R_D = s/s+1$. La fórmula para el cálculo de esta estimación es:

$$\frac{1}{R_D} \approx 1 + \rho^{\frac{s}{22\rho}} \quad (3.6)$$

En la figura 3.1 se representan las tres aproximaciones introducidas en este apartado para un sistema de 2 canales y carga variable entre el 0% y el 100%. La figura 3.2 representa la misma relación para un conjunto de 30 canales. En las gráficas, y especialmente en la figura 3.2 para la cual las probabilidades de demora son más bajas, puede observarse la inconsistencia de la ecuación (3.2) en determinados casos. Los valores exactos están tomados de [SEE85a] que no proporciona valores cuando la probabilidad de demora es muy baja (debido a baja carga o elevado número de canales). El valor representado como exacto para carga nula es el valor asintótico de la ecuación 3.5.

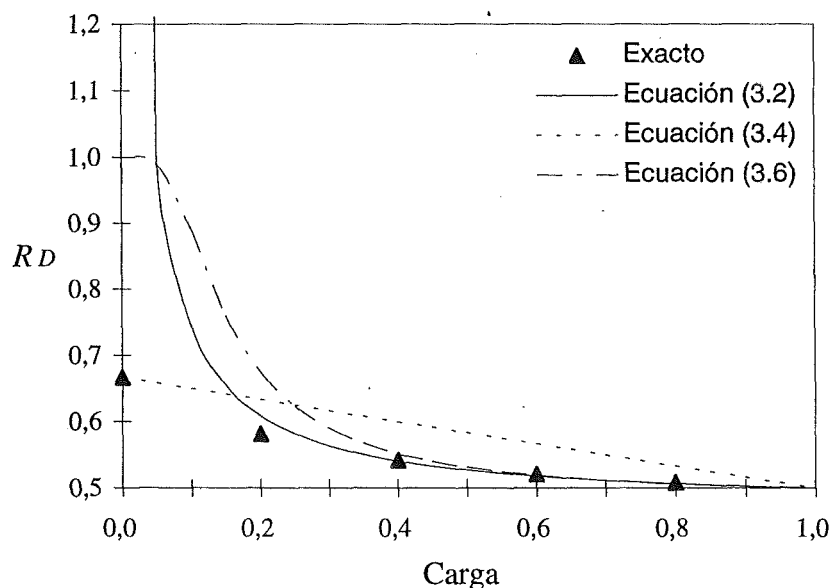


Figura 3.1 Valores aproximados y exactos de la relación entre el retardo de la cola M/D/s y el de la M/M/s para un conjunto de 2 canales y carga variable.

Para la probabilidad de demora en un sistema M/D/s, tal como se ha citado, una excelente aproximación es la probabilidad demora en la cola M/M/s. En este caso la probabilidad de demora en M/M/s es además una cota superior [TIJ86].

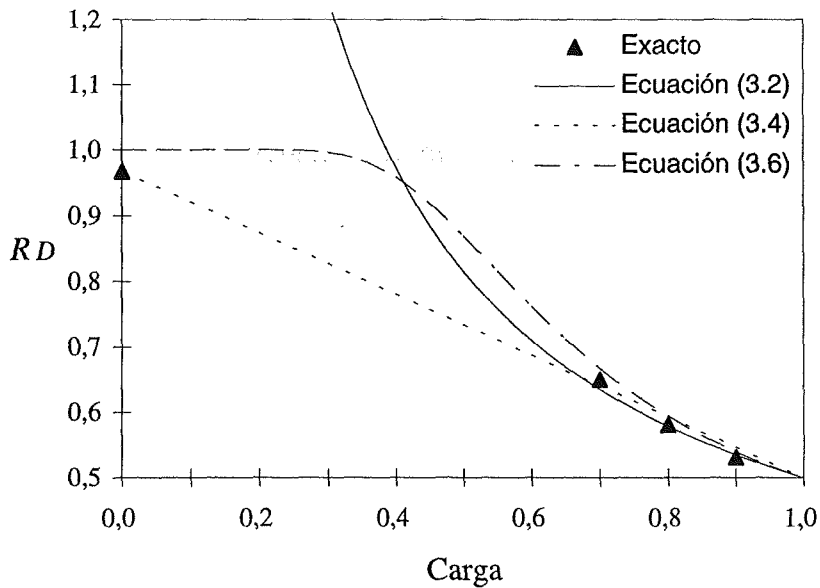


Figura 3.2 Valores aproximados y exactos de la relación entre el retardo de la cola M/D/s y el de la M/M/s para un conjunto de 30 canales y carga variable.

3.2 Tiempo medio de espera en la cola M/M/s con prioridad HOL

En el apartado 1.5 se ha citado la existencia de una solución analítica exacta para la cola M/M/C con prioridad HOL y un número cualquiera de niveles de prioridad. Si el nivel de prioridad que ostenta un tráfico determinado se indica mediante un subíndice i y entendiendo que a menor índice más prioritaria es la llamada, puede calcularse el tiempo medio de espera en cola para una categoría determinada de prioridad como [GRO74]:

$$W_i(M/M/s) = PD \frac{1/\mu}{s(1-\sigma_i)(1-\sigma_{i-1})} \quad (3.7)$$

$$\sigma_i = \sum_{k=1}^i A_k / s \quad A = \sum_{\forall k} A_k \quad (3.8)$$

donde $W_i(M/M/s)$ representa el retardo medio de acceso de las llamadas de categoría i , y σ_i la suma de todas las cargas de tráfico generadas por las llamadas de la categoría en estudio y todas las de más prioridad tal como se reflejas en la ecuación

(3.8). Para la probabilidad de demora PD , que es la misma para todas las llamadas con independencia del rango de prioridad que ostentan, sigue siendo válida la ecuación (1.1) ya que el tráfico ofrecido es la suma de los tráficos generados individualmente para cada nivel de prioridad, y la existencia de un sistema de prioridades HOL altera el orden de las llamadas en la cola pero no la probabilidad de que todos los canales estén ocupados.

Aplicando la ecuación (3.7) para el caso concreto de un sistema con dos niveles de prioridad se obtiene que los retardos para las llamadas de cada rango son:

$$W_1(M/M/s) = PD \frac{1/\mu}{s(1-p\rho)} \quad (3.9)$$

$$W_2(M/M/s) = PD \frac{1/\mu}{s(1-p\rho)(1-\rho)} \quad (3.10)$$

donde p representa el tanto por uno de llamadas prioritarias. Obviamente debe cumplirse que el tiempo de espera medio para todas las llamadas sea el obtenido mediante la ecuación (1.2), que debe coincidir con:

$$W(M/M/s) = pW_1(M/M/s) + (1-p)W_2(M/M/s) \quad (3.11)$$

Estos resultados exactos son útiles como primera aproximación a un sistema en general y como cota superior del retardo en sistemas en los que el coeficiente de variación del tiempo de servicio sea menor que la unidad (tiempo de servicio distribuido de forma más regular que el exponencial).

Un parámetro útil de interés en los sistemas con prioridades es la ganancia de prioridad, definida como la relación entre el retardo que sufrirían las llamadas en el caso de que no hubiera mecanismo de prioridades y el retardo actual de las llamadas prioritarias. En el caso de colas M/M/s se puede obtener dicho parámetro relacionando las ecuaciones (1.2) y (3.9), quedando así en función de la carga como:

$$GP(M, \rho) = \frac{W(M/M/s)}{W_1(M/M/s)} = \frac{1-p\rho}{1-\rho} \quad (3.12)$$

ecuación que se representa en la figura 3.3, en la cual se observa que la ganancia es mejor (y por otra parte más necesaria) con cargas mayores y que es inútil si el porcentaje de prioridad es alto al tender la ganancia de prioridad a uno. Obsérvese que en este caso la ganancia de prioridad no depende del número de canales disponibles, a

diferencia de lo que va a ocurrir con otras distribuciones de la duración de la llamada. La máxima ganancia se obtiene cuando el porcentaje de llamadas prioritarias tiende a cero y es:

$$GP(M, \rho \rightarrow 0) = \frac{1}{1-\rho}$$

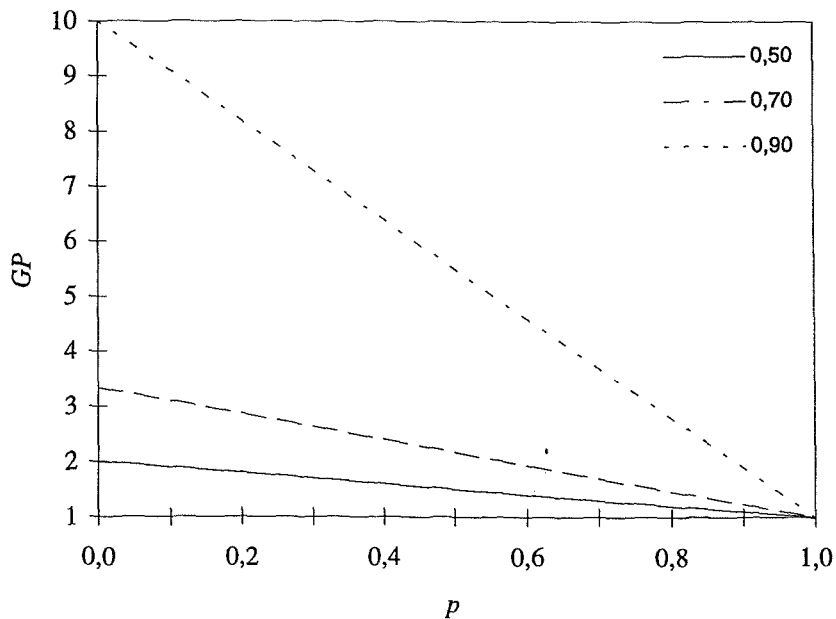


Figura 3.3 Ganancias de prioridad GP para diversas cargas y prioridades p con distribución exponencial de la duración de la llamada. En este caso la ganancia no depende del número de canales.

En base a lo expuesto, las condiciones de evaluación de interés de un sistema de telefonía móvil PAMR con prioridades son las siguientes:

- Alta carga: la utilización de prioridades es necesaria si la carga es alta y por tanto el retardo medio de acceso para todas las llamadas puede ser también alto. Se entiende que en situaciones de carga media si el sistema está bien dimensionado el retardo de acceso es pequeño y aceptable para todas las llamadas.
- Bajo porcentaje de llamadas prioritarias: la prioridad rebaja sensiblemente el retardo de acceso en el caso de que sean pocas las llamadas que la poseen. Por otro lado y tal como se verá más adelante, en esta situación no se incrementa

de forma sensible el retardo que sufren las llamadas no prioritarias respecto al que sufrirían en un sistema sin prioridad.

3.3 La cola M/D/s con dos niveles de prioridad

En esta sección se propone un método para calcular de forma aproximada el tiempo medio de acceso a una cola en la cual la duración media del servicio está distribuida de forma determinística, manteniendo las hipótesis de llegadas de Poisson (población infinita) y cola infinita, a la vez que se introducen dos niveles de prioridad HOL. Para este modelo no existe una solución analítica exacta que proporcione el retardo medio. Si bien este modelo se encontrará solo ocasionalmente en sistemas de telefonía móvil reales (principalmente debido a la dificultad de encontrar sistemas reales con distribución determinística de la duración de la llamada), su interés es el citado en el apartado 3.1: para establecer una cota inferior del GoS y como punto de partida para el análisis de sistemas en los que el tiempo de servicio está distribuido de forma hipoexponencial. Hay que tener en cuenta también que en sistemas que no son de telefonía, así como en tráfico vocal o no vocal de redes fijas si es frecuente encontrar una distribución determinística del tiempo de servicio [KUE96].

El método propuesto se basa en una combinación correcta de las ecuaciones introducidas en los dos apartados anteriores, y en la comprobación numérica de los resultados del método propuesto comparándolos con los valores exactos. Los pasos para el cálculo son los siguientes [BAR96a]:

Paso 1: El retardo medio de acceso para las llamadas prioritarias debe calcularse combinando las ecuaciones (3.4) (dado que las llamadas prioritarias deberían ofrecer una carga baja y la citada ecuación aparece como la más apta) y (3.9) para obtener:

$$W_1(M/D/s) = W_1(M/M/s) \left\{ \frac{(1-p\rho)s}{s+1} + \frac{p\rho}{2} \right\} \quad (3.13)$$

Paso 2: El retardo medio para todas las llamadas puede ser obtenido de la ecuación (3.2) ya que la carga ofrecida por todas las llamadas conjuntamente no es baja (precisamente la situación de interés es con cargas altas tal como se ha visto) y la ecuación propuesta es muy precisa.

Paso 3: El retardo medio para las llamadas no prioritarias puede obtenerse de la ecuación homóloga de la (3.11) dado que el retardo medio para todas las llamadas en una cola M/D/s es:

$$W(M/D/s) = pW_1(M/D/s) + (1-p)W_2(M/D/s) \quad (3.14)$$

donde la única incógnita es precisamente el retardo medio de las llamadas no prioritarias.

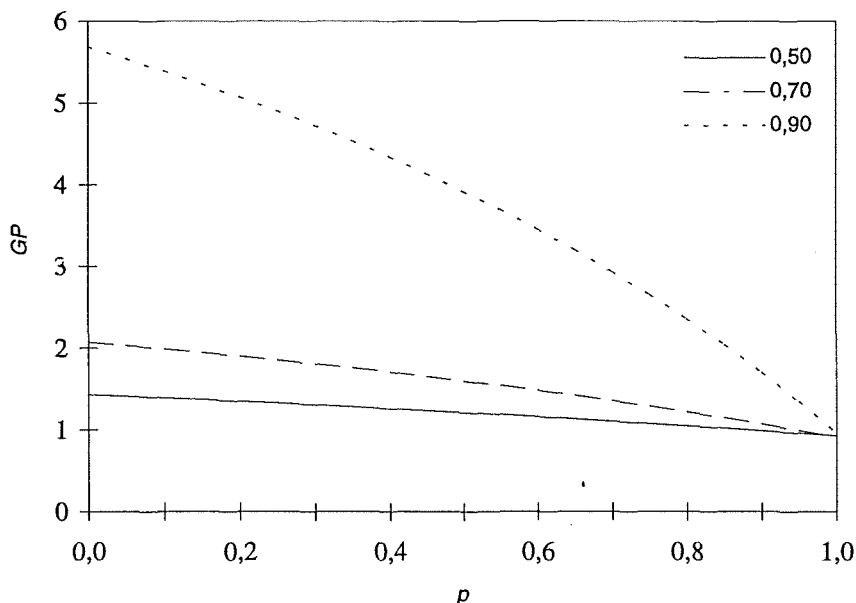


Figura 3.4 Ganancia de prioridad para un sistema de 10 canales con diversas cargas y duración determinística de la llamada.

De acuerdo a las aproximaciones realizadas podemos calcular la ganancia de prioridad en la cola $M/D/s$ como:

$$GP(D, \rho) = \frac{W(M/D/s)}{W_1(M/D/s)} = \frac{1-p\rho}{1-\rho} \times \frac{R_D}{R_{D1}} \quad (3.15)$$

donde el primer factor es justamente $GP(M, \rho)$ y el segundo será siempre menor que la unidad, ya que R_D es decreciente conforme la carga aumenta, tal como puede deducirse de cualquiera de las aproximaciones referidas. De acuerdo con la metodología propuesta R_D se calculará conforme a la ecuación (3.2) y R_{D1} por tratarse de llamadas prioritarias y por tanto en baja carga de acuerdo a la ecuación (3.4) o (3.13). De ello se deduce que la utilización de prioridades da un resultado más espectacular con distribución exponencial que con distribución determinística de la duración de la llamada al ser $GP(D, \rho) < GP(M, \rho)$. Podemos fácilmente conjeturar

generalizando este resultado que para coeficientes de variación del tiempo de servicio menores que uno obtendremos ganancias de prioridad peores que para si la distribución fuera exponencial y mejores que si la distribución fuera determinística. Esta circunstancia previsible es demostrada en la sección 3.5. En la figura 3.4 se representa la ganancia de prioridad para un sistema de 5 canales y diversas situaciones de carga y porcentaje de llamadas prioritarias.

3.4 Resultados numéricos para la cola M/D/s con prioridad

En las figuras 3.5 y 3.6 se muestran los resultados de los retardos medios de acceso para llamadas prioritarias y no prioritarias respectivamente en un sistema de 5 canales con una carga del 90% (se ofrecen en total 4,5 Erlang) en el que las llegadas son de Poisson y la duración de la llamada es determinística y está normalizada a 1 unidad de tiempo. De este modo el eje de ordenadas representa el retardo de acceso normalizado con respecto al tiempo medio de servicio. En la gráfica se comparan los tiempos calculados de acuerdo al procedimiento descrito con los obtenidos mediante simulación para porcentajes de llamadas prioritarias de hasta el 50%. También se representan los tiempos calculados con las ecuaciones (3.8) y (3.9) para la cola M/M/s como referencia de lo que constituiría una primera aproximación y cota superior en ausencia de la metodología de cálculo propuesta en la sección anterior. Puede observarse la eficacia del sistema de prioridades, especialmente cuando el porcentaje de dichas llamadas es bajo, en cuyo caso se obtiene un importante beneficio en el retardo de acceso con un coste bajo para el resto de llamadas.

De cara a corroborar la aptitud de la metodología descrita para el cálculo de los retardos, ésta ha sido aplicada a una gran variedad de situaciones variando la carga, el número de canales y el porcentaje de llamadas prioritarias, proporcionando en todos los casos resultados muy satisfactorios. Otras situaciones de carga y número de canales aparecen en [BAR96a]. La figura 3.7 representa una situación de carga y número de canales diferente, aunque las conclusiones a las que se puede llegar son las ya expuestas con el ejemplo anterior.

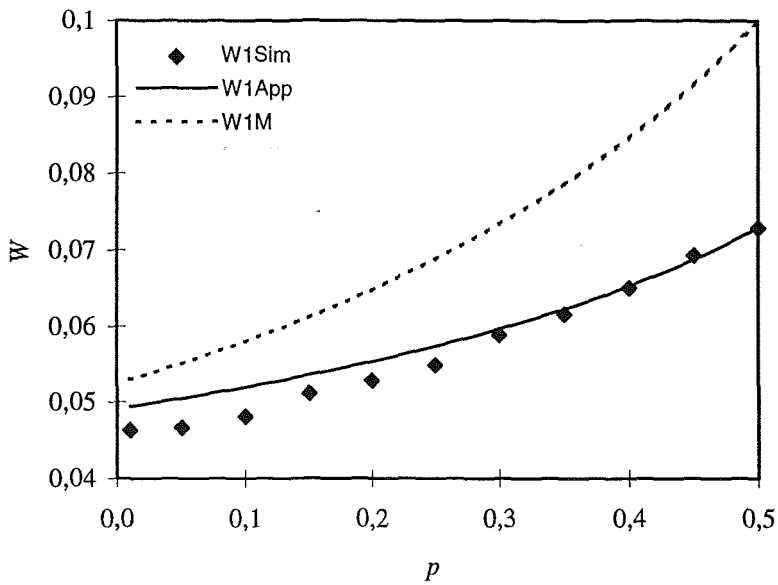


Figura 3.5 Retardo medio de acceso a la cola M/D/s para las llamadas prioritarias con carga del 90% y 5 canales.

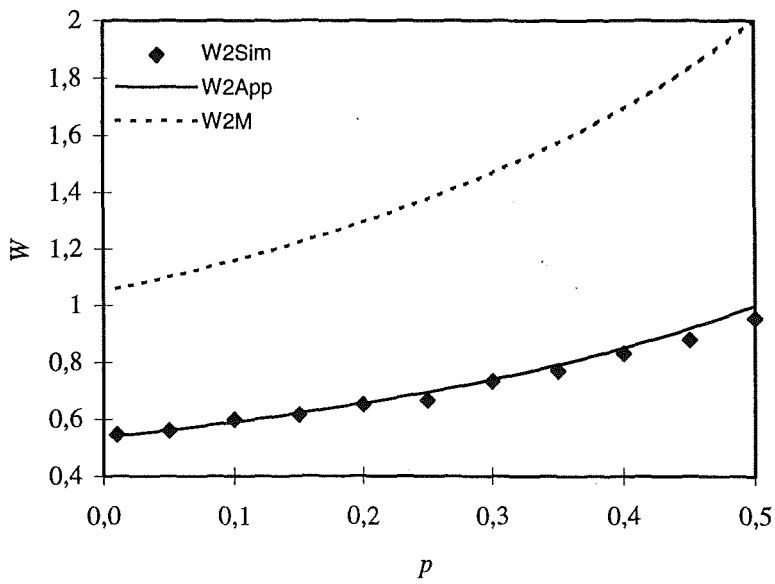


Figura 3.6 Retardo medio de acceso a la cola M/D/s para las llamadas no prioritarias, con carga del 90% y 5 canales.

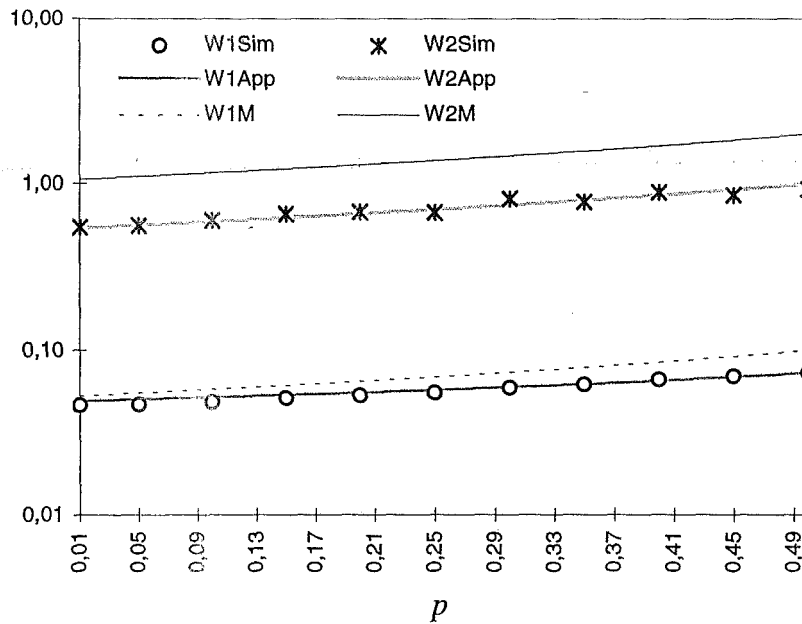


Figura 3.7 Retardo medio de acceso en cola M/D/s para las llamadas prioritarias y no prioritarias a un sistema con 15 canales y carga del 95%.

3.5 La cola M/G/s con prioridad

La aproximación presentada en la sección anterior es extremadamente precisa, pero solamente es válida si se asume que la duración de la llamada está distribuida de forma determinística, situación que tal como se ha comentado raramente se producirá en sistemas de telefonía móvil. El resultado del método presentado constituye en realidad una cota inferior del retardo de acceso. Tal como se ha comprobado en el capítulo 2, la distribución de la duración de la llamada en sistemas PAMR en los que se proporciona servicio de voz únicamente, es de forma hipoexponencial con coeficientes de variación menores que uno y por tanto con menos dispersión que si el tiempo de servicio estuviese distribuido exponencialmente. De este modo el retardo medio calculado para la cola M/M/s conforme a la ecuación (1.2) puede interpretarse como una cota superior.

Hasta este punto podemos calcular ambas cotas superior e inferior del retardo medio en un sistema PAMR con servicio de voz y prioridades. De cualquier modo la cota inferior tiene poco interés en evaluación ya que no garantiza el buen comportamiento del sistema, mientras que la cota superior tiende a sobredimensionar el número de canales necesarios para conseguir un determinado GoS. Así sería

interesante poder ofrecer una buena estimación de dicho retardo en lugar de las cotas mencionadas. En esta sección se introduce el procedimiento para calcular de forma aproximada el retardo medio de acceso a una cola $M/G/s$ en la que el tiempo de servicio sigue una distribución hipoexponencial, de forma muy parecida al método propuesto en la sección 3.3.

De las distintas aproximaciones existentes para tiempo medio de espera en la cola $M/G/s$ (ver sección 5.2) la introducida por Kimura [KIM94] presenta un excelente compromiso ofreciendo una buena precisión y una baja complejidad de cómputo. Dicha aproximación es como sigue:

$$W(M/G/s) = \frac{1+c^2}{\frac{2c^2}{W(M/M/s)} + \frac{1-c^2}{W(M/D/s)}} \quad (3.16)$$

donde c representa el coeficiente de variación del tiempo de servicio. Esta fórmula requiere el conocimiento de R_D a diferencia de la ecuación (3.3) en la que dicho parámetro no es necesario para el cálculo. Recordemos que la ecuación (3.3) ha sido utilizada precisamente para obtener R_D en situaciones de baja carga por lo que era necesario acudir a una aproximación que no incluyera dicha relación, aun a costa de una menor precisión en la estimación. Sin embargo en este punto y de cara a la aproximación que se está buscando, interesa una aproximación precisa. La relación R_D puede obtenerse de las cualquiera de las ecuaciones presentadas en la sección 3.2 de forma aproximada.

De modo similar a como se procedió en la sección 3.3, los pasos propuestos para el cálculo del retardo medio son los siguientes [BAR97a]:

Paso 1: El retardo medio para las llamadas prioritarias $W_1(M/G/s)$ debe calcularse utilizando la aproximación de Kimura conforme a la ecuación (3.16) como:

$$W_1(M/G/s) = \frac{1+c^2}{\frac{2c^2}{W_1(M/M/s)} + \frac{1-c^2}{W_1(M/D/s)}} \quad (3.17)$$

donde $W_1(M/D/s)$ se obtiene de la ecuación (3.13), dada la baja carga que suponen las llamadas prioritarias en el sistema.

Paso 2: El retardo medio para todas las llamadas se obtendrá de la ecuación (3.16) en la que $W(M/D/s)$ puede obtenerse de la ecuación (3.2) ya que aquí no estamos situación de baja carga.

Paso 3: El retardo medio para las llamadas no prioritarias $W_2(M/G/s)$ se obtiene de la siguiente ecuación del retardo medio de acceso para todas las llamadas:

$$W(M/G/s) = pW_1(M/G/s) + (1-p)W_2(M/G/s) \quad (3.18)$$

en la que la única incógnita es precisamente el retardo buscado.

La ganancia de prioridad obtenida para la cola M/G/s siguiendo la metodología propuesta en este apartado es:

$$GP(G, \rho, c^2) = \frac{W(M/G/s)}{W_1(M/G/s)} = \frac{\frac{2c^2}{W_1(M/M/s)} + \frac{1-c^2}{W_1(M/D/s)}}{\frac{2c^2}{W(M/M/s)} + \frac{1-c^2}{W(M/D/s)}} \quad (3.19)$$

ecuación de la que después de algunas manipulaciones algebraicas sencillas e introduciendo los valores ya obtenidos anteriormente de $GP(M, \rho)$ y $GP(D, \rho)$ obtenemos:

$$GP(G, \rho, c^2) = GP(D, \rho) \frac{2c^2 R_{1D} + 1 - c^2}{2c^2 R_D + 1 - c^2} \quad (3.20)$$

Esta ganancia coincide con los resultados ya presentados en secciones anteriores. Así para $c^2=1$ (distribución exponencial introducida en la sección 3.2) y $c^2=0$ (distribución determinística introducida en la sección 3.3) la ganancia de prioridad coincide respectivamente con las ecuaciones (3.12) y (3.15).

Debido a que la carga ofrecida por las llamadas prioritarias es una parte del total de llamadas y que R_D es decreciente cuando la carga aumenta [KIM84], siempre se cumple:

$$0,5 < R_D < R_{1D} < 1 \quad (3.21)$$

con lo que sustituyendo valores en la ecuación (3.20) fácilmente obtenemos que la ganancia de prioridad con distribución hipoexponencial ($0 < c^2 < 1$) es mayor que para el caso determinístico y menor que para el exponencial, quedando demostrado lo que habíamos conjeturado en la sección 3.3, es decir:

$$GP(D, \rho) < GP(G, \rho, c < 1) < GP(M, \rho)$$

Este hecho se refleja en la figura 3.8 en la que se representa la ganancia de prioridad para un sistema con 10 canales y carga del 90% en tres situaciones diferentes del coeficiente de variación de la duración de la llamada: determinística, general con $c^2=0,3$ y exponencial.

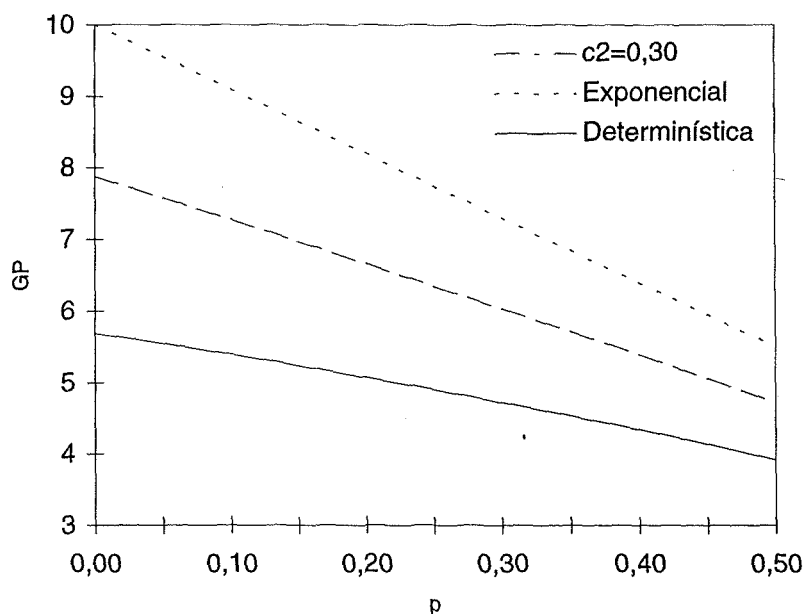


Figura 3.8 Ganancia de prioridad para un sistema con 5 canales y carga del 90% para tres coeficientes de variación de la duración de la llamada.

3.6 Resultados numéricos en la cola M/G/s

En las figuras 3.9 y 3.10 se representan los retardos de acceso para las llamadas prioritarias y no prioritarias respectivamente a un sistema PAMR con las siguientes características:

Número de canales: 5

Carga: 90% (total de tráfico ofrecido de 4 Erlang)

Duración media de la llamada: 1 unidad de tiempo

Coefficiente de variación de la duración de la llamada: 0,33

Dicho coeficiente de variación es consistente con los resultados de las medidas del capítulo 2. En las simulaciones se ha utilizado una función de distribución Erlang-3 (E_3), cuyo coeficiente de variación es de 1/3, para la duración de la llamada. De hecho en la bibliografía consultada es habitual el uso de funciones de distribución de Erlang- k para comparar aproximaciones de la distribución general (G) hipoexponencial [COS76, BOX79, HOO84, TIJ86, KIM94]. En las gráficas se observa que aún sin ser tan precisa como la aproximación de la cola M/D/s, es mucho mejor como estimación la aproximación que la cota superior representada por la cola M/M/s.

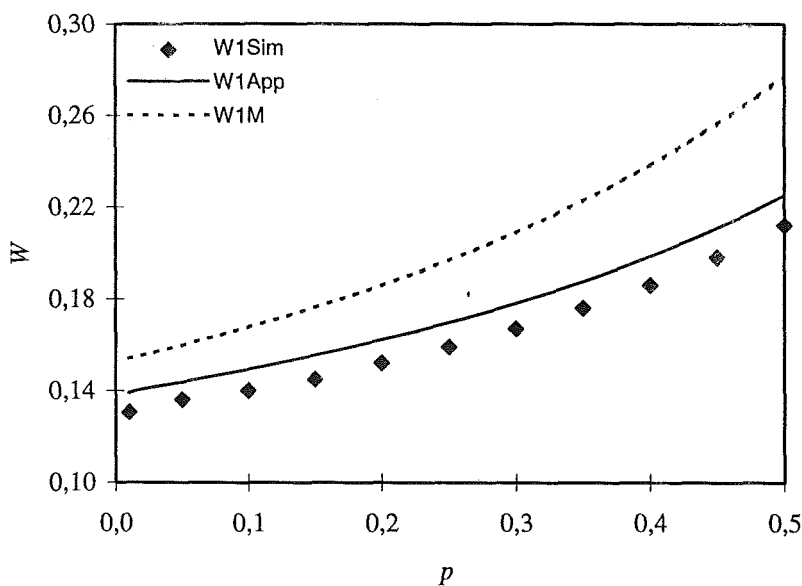


Figura 3.9 Retardo soportado por las llamadas prioritarias en un sistema PAMR con 5 canales y carga del 90% para distintos valores del porcentaje de prioridad.

En la tabla 3.1 figuran los valores del retardo medio de acceso de las llamadas prioritarias para diferentes coeficientes de variación del tiempo de servicio. Las funciones de densidad de probabilidad de la duración del servicio utilizadas en las simulaciones son de nuevo funciones de Erlang: Erlang-2 para $c^2=0.5$ y Erlang-4 para $c^2=0.25$. El retardo medio está expresado en milisegundos para una duración media de la llamada de 1 segundo. Se observa que en todos los casos el tiempo calculado de forma aproximada siguiendo la metodología propuesta es muy exacto y superior al obtenido mediante simulación. Este hecho contribuye a la posible aplicación de este método en entornos de ingeniería en los que, en caso de error, resulta más conveniente situarse en un peor caso.

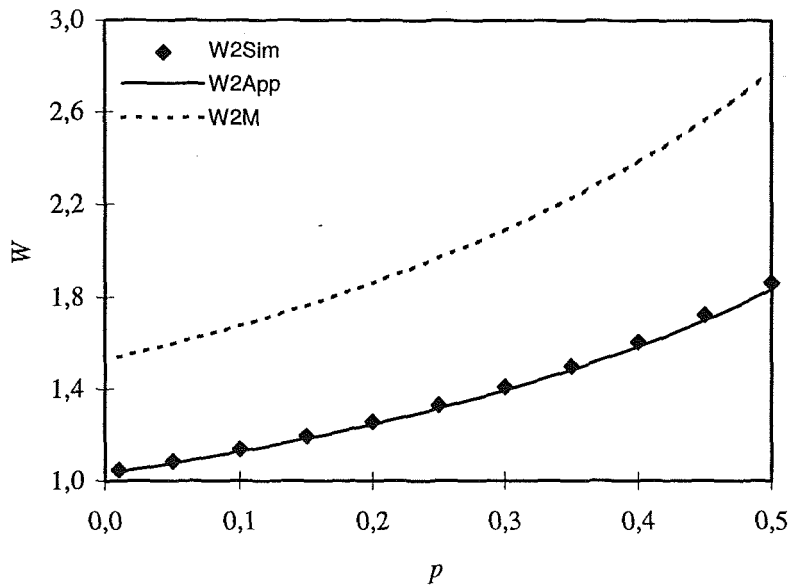


Figura 3.10 Retardo soportado por las llamadas no prioritarias en un sistema PAMR con 5 canales y carga del 90% para distintos valores del porcentaje de prioridad.

c^2	Carga	p	$s=5$		$s=2$		$s=0$	
			Sim.	Appr.	Sim.	Appr.	Sim.	Appr.
0,25	0,90	0,1	137	146	51,3	54,3	27,5	28,6
	0,95	0,2	158	169	66,4	69,0	39,1	39,4
	0,90	0,1	148	158	56,2	58,7	30,0	30,8
	0,95	0,2	173	184	72,9	74,8	42,1	42,7
0,5	0,90	0,1	145	155	53,7	56,2	28,2	29,3
	0,95	0,2	169	179	68,5	71,3	39,1	40,4
	0,90	0,1	159	169	58,8	61,3	31,0	32,0
	0,95	0,2	186	197	75,6	78,3	43,3	44,3

Tabla 3.1 Retardo medio de las llamadas prioritarias bajo diferentes condiciones de carga, prioridad, número de canales y coeficiente de variación de la duración de la llamada.

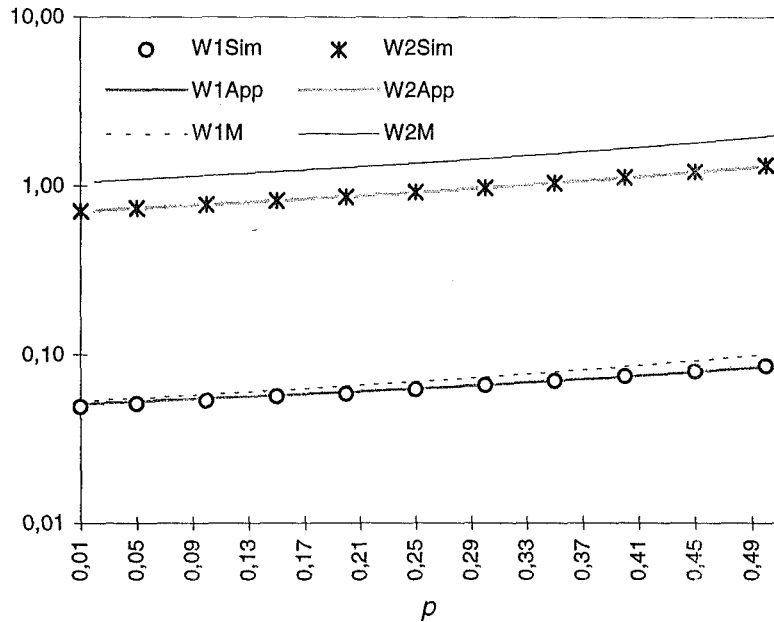


Figura 3.11 Tiempos de espera para las llamadas prioritarias y no prioritarias en un sistema con 15 canales y carga del 95%.

Al igual que en el caso de la cola M/D/s, también se ha probado esta metodología de cálculo bajo distintas condiciones de carga, número de canales y coeficiente de variación de los tiempos de servicio, obteniéndose siempre estimaciones muy cercanas a las obtenidas mediante simulación. La figura 3.11 representa los tiempos de acceso en otro de los ejemplos realizados.

En el caso real de un sistema PAMR la f.d.p. del tiempo de servicio no será una Erlang- k tal como se ha venido utilizando en la simulación de las comparaciones realizadas. En el capítulo 2 se ha comprobado que otras f.d.p. como las lognormales o las Erlang- j,k se ajustan mejor a la realidad de un sistema PAMR. De cualquier modo, un cambio en otros momentos de la distribución de la duración de la llamada, más allá del segundo, afectará muy poco al retardo medio de acceso al sistema.

3.7 Sumario

En este capítulo se ha realizado una revisión de algunos resultados analíticos exactos para la cola M/M/s con prioridad, y otros aproximados analíticos o heurísticos para las colas M/D/s y M/G/s sin prioridad. Una correcta combinación de los resultados citados conduce a la obtención aproximada del retardo medio de acceso en las colas M/D/s y M/G/s con dos niveles de prioridad. Las aproximaciones

Capítulo 3

presentadas en este capítulo poseen dos características importantes en toda aproximación: exactitud del resultado y facilidad de cómputo. El método que se presenta ha sido comprobado numéricamente en un entorno similar al que se produce en sistemas PAMR, y más concretamente a aquel en que interesa obtener evaluaciones, es decir:

- Carga alta.
- Más de un canal.
- Dos niveles de prioridad
- Baja proporción de llamadas prioritarias.
- Distribución hipoexponencial del tiempo de servicio.

La exactitud de la estimación del GoS presentada (en este caso retardo medio de acceso) frente a la alternativa de utilizar el modelo $M/M/s$ como cota superior, permite evitar el gasto excesivo de radiocanales en sistemas PAMR y contribuye por tanto al ahorro espectral.

La aproximación presentada puede encontrar otras aplicaciones en campos fuera de la ingeniería de telecomunicación en los que se producen tiempos de servicio hipoexponenciales. Se puede conjeturar que algunas de las condiciones bajo las que se ha realizado el estudio podrían ser relajadas manteniendo la exactitud del resultado. De este modo posteriores investigaciones deberían tratar de generalizar los resultados de este capítulo a sistemas con más de dos niveles de prioridad y a tiempos de servicio hiperexponenciales.