

4 Capítulo 4. OPTIMIZACIÓN DEL PROTOCOLO DE ACCESO

4.1 Introducción

En este capítulo vamos a profundizar en el estudio del rendimiento del protocolo de acceso DQRAP/CDMA para condiciones de trabajo genéricas, de cara a optimizar sus prestaciones incluso en el caso de que el tráfico de entrada presente características desfavorables. En particular, fijaremos la atención en la influencia que tienen las características del tráfico ofrecido sobre la eficiencia del sistema.

En el capítulo 3, y de cara a presentar el funcionamiento del protocolo de acceso DQRAP/CDMA, se ha analizado el funcionamiento de dicho protocolo en unas ciertas condiciones concretas de trabajo, para las que éste ha mostrado su buen comportamiento tanto en estabilidad como en retardo y caudal servido efectivo. Sin embargo, es necesario profundizar en el estudio del rendimiento del protocolo para condiciones de tráfico genéricas que puedan poner al sistema en situaciones menos favorables.

En primer lugar, se detallará el análisis realizado de los aspectos del funcionamiento del protocolo para identificar aquellos que puedan limitar su rendimiento. En segundo lugar, se explicarán con detalle las propuestas realizadas de mejora al objeto de incrementar la eficiencia de transmisión del protocolo. Finalmente, se estudiarán de forma exhaustiva estas propuestas y se analizará su eficacia. De este modo, el objetivo final consiste en la elección de los valores adecuados de los parámetros del protocolo que optimizan su rendimiento global, en cualesquiera que sean las condiciones de trabajo.

En una primera fase, y de cara a completar el estudio de todos los aspectos del protocolo susceptibles de ser optimizados, vamos a centrar el análisis en el sistema de detección de secuencias de acceso mostrado en el punto 3.10. Un análisis detallado del mismo permitirá encontrar los valores más adecuados para sus parámetros.

4.2 Análisis del sistema de detección del estado de los minislots

Partiendo del esquema de detección presentado en el punto 3., vamos a analizar en detalle sus propiedades y comportamiento en función de los parámetros que lo conforman. El objetivo final consiste en escoger los valores más adecuados para estos parámetros, de manera que optimicemos el rendimiento del esquema a la hora de detectar las secuencias de acceso.

Si observamos las expresiones del punto 3.10.2, desde la (3.39) hasta la (3.50), donde se muestran los valores de las probabilidades de error en la detección del estado de los minislots de control, comprobamos que estas probabilidades dependen básicamente de tres parámetros:

- El número de filtros del banco de filtros adaptados: F
- La probabilidad de detección de las secuencias de acceso: P_d
- La probabilidad de falsa alarma de los filtros: P_f

El número de filtros (F) acota el número máximo de usuarios que pueden ser servidos en el sistema. Este número está limitado al valor $F(F-1)/2$. Determinamos inicialmente el valor

$F=16$, de manera que puede darse servicio a un máximo de 120 usuarios, que consideraremos suficiente en células de tamaño estándar (macrocélulas).

Por su parte, analizando ahora las expresiones para P_d y P_f (ver las expresiones 3.37 y 3.38), comprobamos que estos parámetros dependen a su vez de:

- *El número de secuencias de acceso transmitidas simultáneamente en un minislot (r):* Este número está relacionado directamente con el número de usuarios que realizan una transmisión de petición de acceso en cada uno de los minislots. Dado el esquema descrito, si k es el número de usuarios que acceden en un cierto minislot, se cumplirá siempre que $r=2 \cdot k$ (ver punto 3.10.1).
- *La probabilidad de falsa alarma de diseño (P_{fD}):* Este valor se utiliza, junto con el número de accesos, para ajustar el umbral de decisión de los bloques correspondientes incluidos en el esquema de la Figura 37 y situados a la salida de los filtros adaptados.
- *El número de accesos simultáneos de diseño (r_D):* Este valor también se utiliza para ajustar el valor del umbral de decisión. Representa el número de secuencias de acceso simultáneas que deben transmitirse en un minislot para que los filtros del sistema tengan la probabilidad de falsa alarma de diseño escogida, es decir el valor P_{fD} .
- *La longitud de la secuencia de acceso (M):* Medida en chips, se ha decidido inicialmente tomar un valor fijo de este parámetro. Este valor es $M=256$.

En primer lugar, y de cara a entender el comportamiento del esquema de detección, se ha estudiado el efecto de las variaciones de P_{fD} y r_D en los valores de P_f y P_d de cada uno de los filtros adaptados. Las siguientes figuras (de la Figura 53 hasta la Figura 56) muestran estos valores en función del número de usuarios que transmiten sus peticiones de forma simultánea (k) para diferentes valores de P_{fD} y r_D .

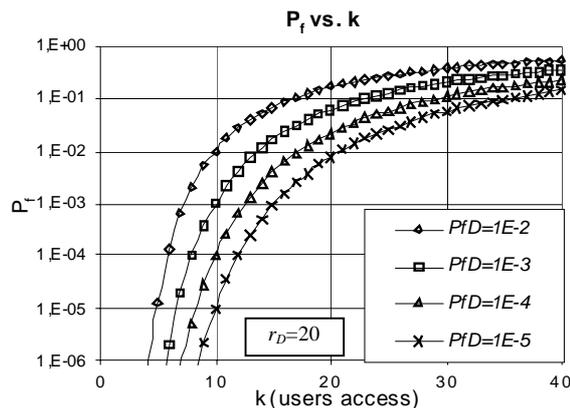


Figura 53. Probabilidad de falsa alarma para distintos valores de P_{fD} con $r_D=20$

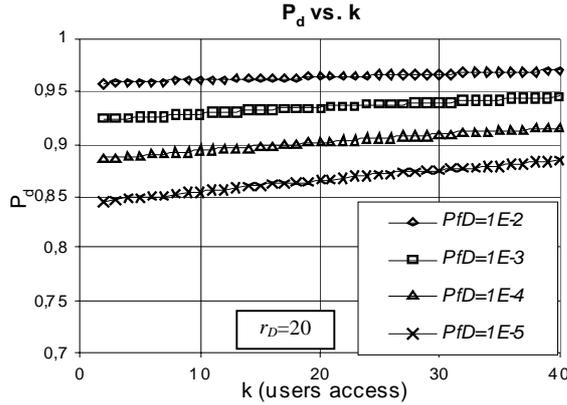


Figura 54. Probabilidad de detección para distintos valores de P_{fD} con $r_D=20$

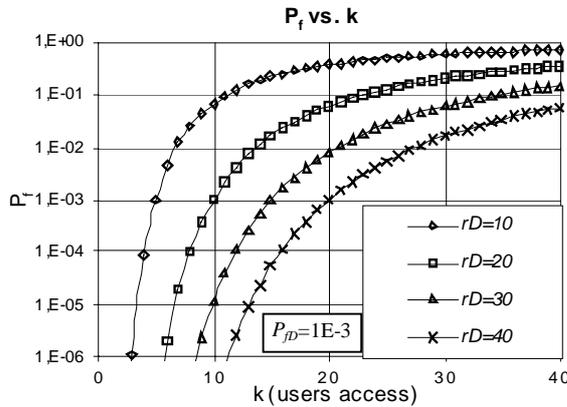


Figura 55. Probabilidad de falsa alarma para distintos valores de r_D con $P_{fD}=10^{-3}$

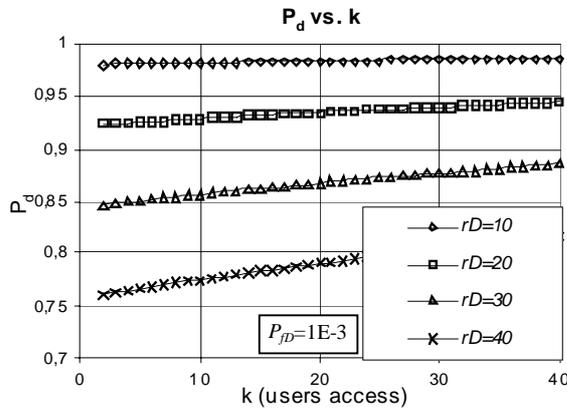


Figura 56. Probabilidad de detección para distintos valores de r_D con $P_{fD}=10^{-3}$

Podemos observar que los valores de P_f y P_d varían en rangos totalmente distintos, pero responden de una manera similar a las variaciones de P_{fD} y r_D . Valores crecientes de estos parámetros también hacen crecer las probabilidades de detección y falsa alarma. Este comportamiento se explica a partir del estudio del esquema de detección propuesto. En efecto, P_{fD} y r_D definen el valor del umbral de decisión de cada filtro adaptado. Cuando este umbral es bajo, cada filtro resulta más sensible a la detección, aumentando la probabilidad de detectar con éxito una secuencia de acceso, pero a su vez también resulta más fácil que la interferencia pueda generar una falsa alarma. Por el contrario, cuando requerimos una detección más precisa de las secuencias de acceso, descendemos el valor de la probabilidad de detección, a

cambio de proteger más el sistema frente a las falsas alarmas. Se muestra de forma evidente, por tanto, el compromiso existente entre ambos valores.

Es interesante reseñar que, mientras que P_d varía suavemente con respecto al número de secuencias simultáneas transmitidas ($r=2 \cdot k$), P_f tiene una variación suave para $r > r_D$ pero decrece abruptamente cuando $r < r_D$. También podemos comprobar que las variaciones de r_D afectan en mucha mayor medida el valor de P_f y P_d que las variaciones de P_{fD} . A su vez, podemos observar que las curvas obtenidas para diferentes valores de r_D (desde 10 a 40) están mucho más distanciadas que las obtenidas variando P_{fD} en tres órdenes de magnitud (desde 10^{-2} hasta 10^{-5}).

De este razonamiento podemos concluir que:

- El valor de r_D deberá usarse para el ajuste grueso de las probabilidades, mientras que P_{fD} servirá para el ajuste fino de los valores. En cualquier caso, deberemos procurar que el sistema esté trabajando, en régimen estable, en la zona en la que $r < r_D$, donde P_f toma sus valores mínimos.

Seguidamente estudiaremos el efecto de la variación tanto de k como de r_D y P_{fD} en las probabilidades de error en la detección de los estados de los minislots, definidas en el punto 3.10.2. Desde la Figura 57 hasta la Figura 64 se muestran estas probabilidades de error para $P_{fD}=10^{-3}$ y valores de r_D entre 10 y 40.

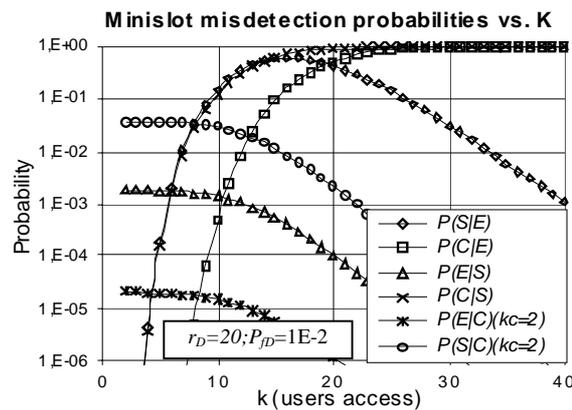


Figura 57. Probabilidades de error en la detección del estado para $r_D=20$ y $P_{fD}=10^{-2}$

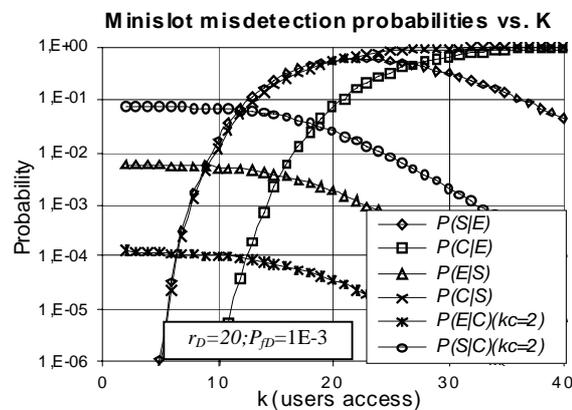


Figura 58. Probabilidades de error en la detección del estado para $r_D=20$ y $P_{fD}=10^{-3}$

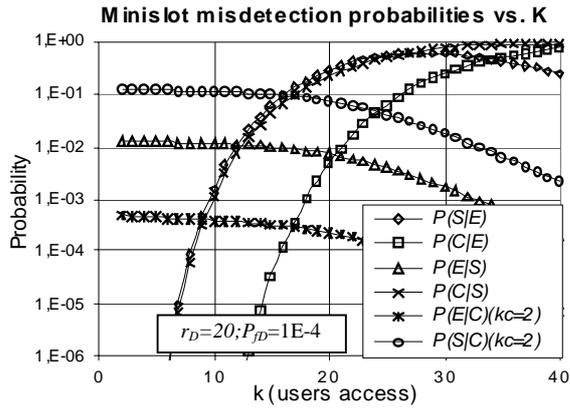


Figura 59. Probabilidades de error en la detección del estado para $r_D=20$ y $P_{fD}=10^{-4}$

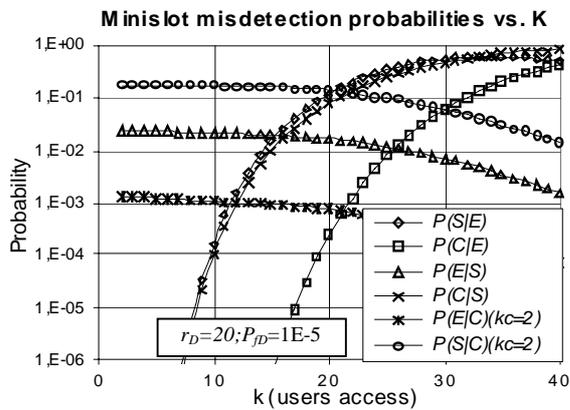


Figura 60. Probabilidades de error en la detección del estado para $r_D=20$ y $P_{fD}=10^{-5}$

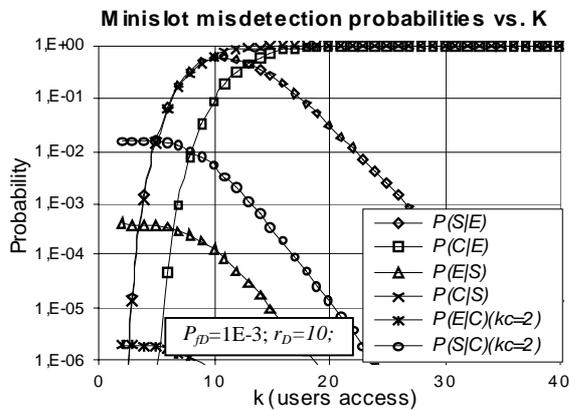


Figura 61. Probabilidades de error en la detección del estado para $r_D=10$ y $P_{fD}=10^{-3}$

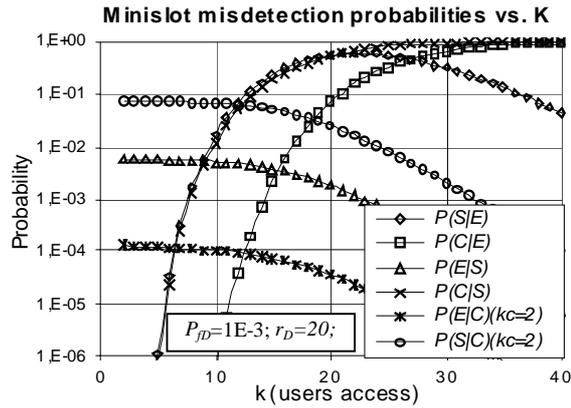


Figura 62. Probabilidades de error en la detección del estado para $r_D=20$ y $P_{fD}=10^{-3}$

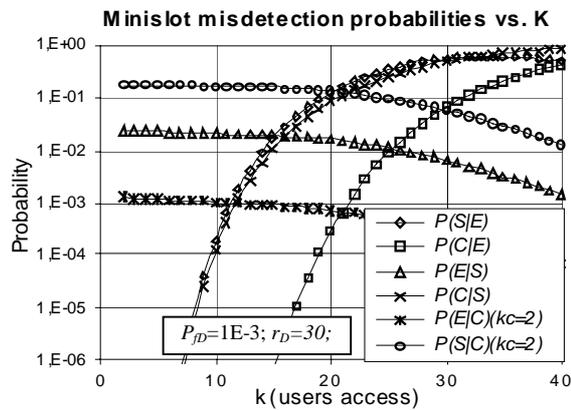


Figura 63. Probabilidades de error en la detección del estado para $r_D=30$ y $P_{fD}=10^{-3}$

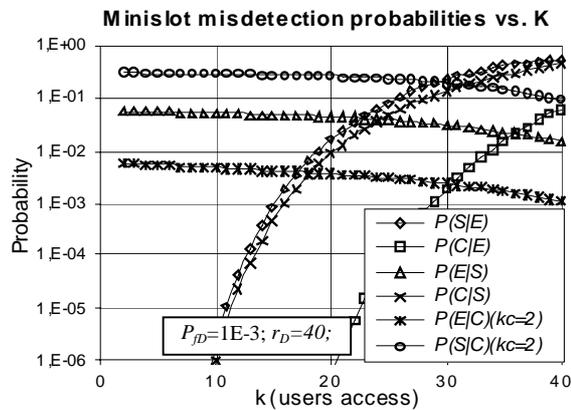


Figura 64. Probabilidades de error en la detección del estado para $r_D=40$ y $P_{fD}=10^{-3}$

Observando estas curvas podemos dividir los errores en la detección en dos grupos claramente diferenciados:

1. *Errores en la detección generados por un número de picos detectados mayor que el número real de secuencias de acceso transmitidas.* En este grupo se encuentran los errores $(S/k=0)$, $(C/k=0)$ y $(C/k=1)$. Este tipo de errores es más frecuente cuando el nivel de interferencia es alto, es decir, cuando hay una carga de tráfico mayor en el acceso y por tanto la probabilidad de falsa alarma es grande. En efecto, se observa como las correspondientes probabilidades del error crecen con k , lo que corrobora este razonamiento. Debe observarse, sin embargo, que la probabilidad del error $(S/k=0)$

alcanza un cierto valor máximo para una cierta carga, y a partir de entonces decrece cuando la carga sigue creciendo. Esto es debido al exceso de interferencia del sistema, que genera un número grande de falsas alarmas que hace muy probable confundir el estado con una colisión, en lugar de con un éxito. En cualquier caso, para la carga ofrecida a la que sucede este fenómeno, los valores de todos los errores son inaceptablemente altos a causa de la interferencia presente en el sistema, es decir, estamos en una zona fuera del rango de utilización del esquema de detección.

2. *Errores producidos por la detección de un número menor de secuencias de acceso de las realmente enviadas al canal.* En este grupo se engloban los errores ($E/k=1$), ($E/k>1$) y ($S/k>1$). Podemos esperar que el comportamiento de las probabilidades de suceso de estos errores se comporten justamente de manera opuesta a los anteriores. En efecto, estos valores decrecen cuando la carga de tráfico aumenta. Los resultados obtenidos de la evaluación matemática del modelo confirman las expectativas.

Para valores pequeños de k , el segundo grupo de errores presenta unos valores de las probabilidades mayores que las del primero. De hecho, los valores de las probabilidades de los errores del primer grupo son valores ciertamente despreciables. Tal y como era de esperar, cuando k aumenta las probabilidades del primer grupo de errores presentan un incremento muy acusado. Estos valores llegan a sobrepasar rápidamente al segundo grupo de probabilidades, cuyos valores decrecen suavemente con k . Un punto interesante en las gráficas es el punto de corte entre los errores ($S/k>1$) y ($S/k=0$), que toman siempre los valores más grandes de cada grupo. Llamaremos k' al número de accesos simultáneos que corresponde a este punto de corte. Observando las gráficas obtenidas para valores iguales de r_D , observamos que el valor de k' crece cuando P_{fD} decrece, así como el valor de la probabilidad de error en la que se produce la intersección. Este hecho indica que la reducción de P_{fD} implica simultáneamente un incremento de la probabilidad de ocurrencia del segundo grupo de errores y una reducción de la probabilidad de ocurrencia para los errores del otro grupo. Las variaciones de r_D producen un efecto análogo. Las curvas obtenidas para los mismos valores de P_{fD} muestran como el punto correspondiente a k' se encuentra en valores de probabilidad cada vez mayores cuando r_D crece.

Tal y como se ha reseñado con anterioridad, las variaciones de r_D generan una variación mayor de las probabilidades que los cambios de P_{fD} . A título de ejemplo, comprobamos que la probabilidad del error ($S/k>1$) aumenta en más de un orden de magnitud cuando r_D pasa de 10 a 40.

En este punto ya estamos en condiciones de justificar la elección de los valores más adecuados para los parámetros P_{fD} y r_D de cara a mejorar el rendimiento general del protocolo. Recordemos que las probabilidades de error son siempre más favorables cuando el número real de secuencias de acceso transmitidas r está por debajo de las de diseño r_D . Por otro lado, y tomando como referencia el valor correspondiente a la intersección definida por k' , es conveniente que los valores de k cumplan que $k < k'$. En este rango de valores, el error ($S/k>1$) es el dominante en el sistema, y por tanto será el primero a considerar para escoger los valores de P_{fD} y r_D . La elección de un valor excesivamente elevado para r_D resultaría en valores inaceptables de la probabilidad del error ($S/k>1$) para cualquier valor de P_{fD} . Por el contrario, valores pequeños de r_D resultan en valores pequeños para k' , lo que tampoco es deseable en el sistema.

Por todo ello, para poder establecer los valores adecuados de los parámetros debemos escoger un umbral razonable para las probabilidades de los errores. Estas probabilidades no deberán

superar el umbral definido. Consideraremos el criterio de que ninguna probabilidad supere el valor 10^{-1} siempre que $k < k'$, y a su vez trataremos de que k' tome el máximo valor posible. De los resultados obtenidos podemos derivar que esta condición se cumple para $r_D \leq 20$. Por otro lado, los valores de P_{fD} que cumplen nuestro requisito deben estar por debajo de 10^{-3} .

Por tanto, los siguientes valores para los parámetros del sistema de detección de secuencias de acceso son los escogidos, que serán utilizados de aquí en adelante en todas las simulaciones:

- | |
|--|
| <ul style="list-style-type: none"> • $r_D=20$ • $P_{fD}=10^{-3}$ |
|--|

Para estos valores, el parámetro k' toma el valor 13, lo que asegura que el sistema funcionará dentro de los límites establecidos para $r < 26$ secuencias de acceso simultáneas.

4.3 Máximo Rendimiento de DQRAP/CDMA

El estudio realizado en el capítulo 3 está basado en un sistema con usuarios de datos que generan mensajes cuyas llegadas siguen una estadística de Poisson, y cuya longitud es una variable exponencial. La media de esta longitud hace que la mayoría de los mensajes no puedan ser enviados en un solo paquete de información. En particular, las gráficas del punto 3.9 se realizaron para mensajes de longitud media 6000 bits, cuando el tamaño de los paquetes es de 600 bits. Esto indica que, en promedio, por cada 10 paquetes de datos se genera una sola petición de acceso.

En estas condiciones, que llamaremos condiciones estándar, el protocolo es capaz de alcanzar un rendimiento muy cercano al óptimo teórico haciendo uso de tan sólo $m=2$ minislots de control por cada slot de datos. El mejor rendimiento respecto al caudal efectivo se obtiene para $m=3$, sin obtener una mejora significativa para valores más grandes de m .

Si recordamos la descripción del sistema realizada en el capítulo 3, el modelo del sistema se divide en dos partes fundamentales: el subsistema de resolución de colisiones y el subsistema de transmisión de datos. Estos dos subsistemas tienen tareas diferentes que funcionan en paralelo: mientras que el primero se encarga de colocar a los usuarios en la cola de transmisión, el segundo se encarga de gestionar esta cola y dar “salida” (servicio) a los usuarios de dicha cola. El mejor rendimiento posible se consigue cuando el subsistema de resolución de colisiones es más rápido que el de transmisión, de modo que el único factor limitante es la capacidad de transmisión del sistema. Esta situación se alcanza minimizando el tiempo que tardan en resolverse las colisiones. Debemos tener en cuenta dos observaciones:

Los usuarios que ya están en la cola de transmisión no generan nuevas peticiones de acceso, y por tanto no pueden generar colisiones.

Los usuarios que están en la cola de transmisión permanecen en el subsistema correspondiente al menos un número de slots de tiempo igual al número de paquetes de datos que contiene el mensaje que están transmitiendo. Esto es, cuando mayor es la longitud de los mensajes, mayor es el tiempo que permanecen en el subsistema de transmisión.

Para una misma carga de tráfico total, podemos concluir que cuanto mayor sea la longitud de los mensajes que generan los usuarios, mayor será el tiempo que permanecerán en el subsistema de transmisión y menor el número de peticiones de acceso que generarán. En conclusión, el escenario que pone más carga en el subsistema de resolución de colisiones (el peor caso respecto a la limitación de rendimiento del sistema global) es aquel en el que, para

una cierta carga de tráfico dada, el número de peticiones de acceso generadas es máximo, es decir, cuando cada uno de los paquetes de datos genera su propia petición. Esto ocurre cuando la longitud de todos los mensajes que generan los usuarios es determinista de valor 1 paquete.

Para apoyar este argumento de carácter general, en la Figura 65 se muestra el retardo de los paquetes normalizado a la longitud de los mensajes (medida en paquetes), para diferentes valores de la misma. Obsérvese que la longitud de los mensajes también indica el tiempo mínimo de transmisión de los mismos, es decir, si todos los paquetes tuvieran un tiempo de acceso nulo, el retardo normalizado de todos ellos sería igual a la unidad. Todos los parámetros del sistema toman los valores estándar (los utilizados en el capítulo 3 para las simulaciones). Es decir, acceso CDMA con $K=17$ códigos disponibles, transmisión de paquetes de longitud $L=600$ bits, ganancia de procesamiento $G_p=64$, probabilidad de falsa alarma de diseño $P_{fD}=10^{-3}$ y número de secuencias de acceso de diseño $r_D=20$. El parámetro L_m indica el número medio de paquetes que tiene cada mensaje generado en el sistema, cuya estadística sigue una variable aleatoria exponencial. La curva que se indica como $L_d=1$ se refiere al caso de tener mensajes de longitud determinista igual a la unidad.

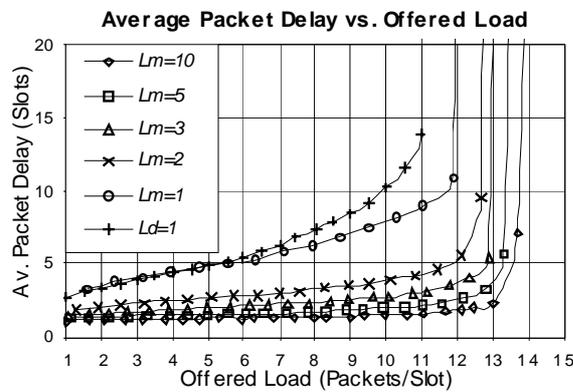


Figura 65. Retardo normalizado de los paquetes

En esta gráfica podemos observar explícitamente como se va degradando la eficiencia del protocolo a medida que el tamaño medio de los mensajes decrece. Esto es debido a la creciente carga de tráfico que sufre el subsistema de resolución de colisiones. Vemos que el caso $L_d=1$ presenta una curva de retardo notablemente más desfavorable que el correspondiente a $L_m=10$, que es lo que hemos llamado caso estándar.

También podemos observar el comportamiento del *throughput* para estas mismas condiciones. La Figura 66 muestra los resultados obtenidos. Por su parte, la Figura 67 muestra una ampliación de la zona de importancia de la Figura 66, donde se puede comprobar que para el caso $L_d=1$ el máximo caudal obtenido es de alrededor de 11'4 paquetes/slot, lejos del óptimo teórico de 14'5 paquetes /slot.

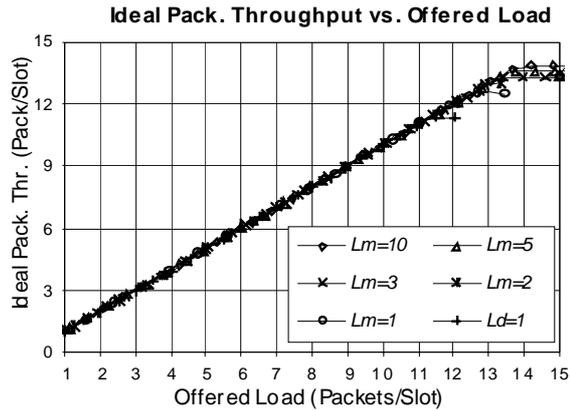


Figura 66. Throughput para diferentes longitudes de paquete

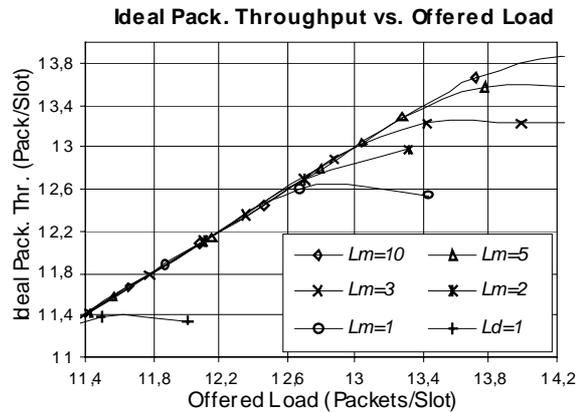


Figura 67. Detalle del throughput para diferentes longitudes de paquete

En la práctica, el caso de tráfico de Poisson con mensajes de longitud fija unitaria es un caso muy especial que se da con poca frecuencia. El uso de técnicas de almacenado o *buffering* pueden permitir agrupar varios paquetes en un solo mensaje de cara a reducir el tráfico de peticiones de acceso. Pocas aplicaciones reales generarán siempre cantidades grandes de mensajes todos ellos de longitud mínima. Sin embargo, este caso de estudio es muy interesante como caso límite para estudiar las capacidades del protocolo DQRAP/CDMA. Por ello, y de cara a optimizar los parámetros del protocolo, nos centraremos en este caso como entorno de trabajo. A partir de este punto, en este capítulo, siempre que no se explicita lo contrario, consideraremos que el tráfico ofrecido por los usuarios es de este tipo (mensajes de longitud determinista igual a 1).

4.3.1 Análisis de parámetros significativos del sistema

Analizamos ahora el comportamiento de algunos parámetros del sistema de cara a poner en evidencia cómo se ven afectados por el incremento de carga del subsistema de resolución de colisiones. La Figura 68 muestra la media de ocupación de la cola de transmisión (DTQ), mientras que la Figura 69 muestra la media de ocupación de la cola de peticiones (CRQ).

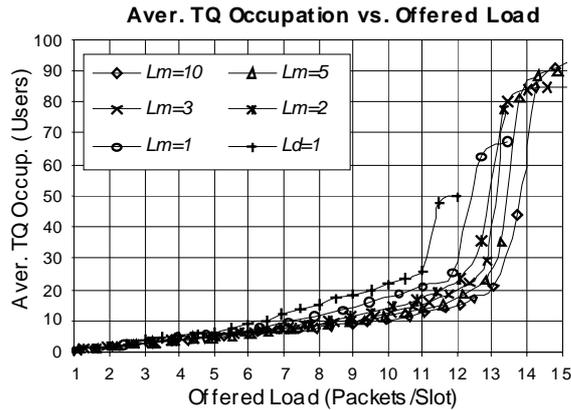


Figura 68. Ocupación media de la cola de transmisión

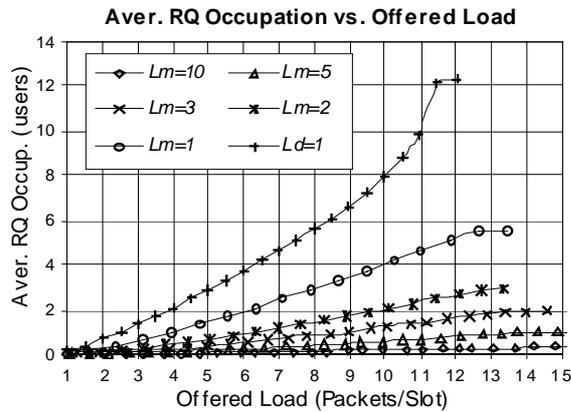


Figura 69. Ocupación media de la cola de peticiones

Estas curvas muestran el crecimiento de la ocupación de las colas del sistema, para cargas totales iguales de tráfico, a medida que se reduce la longitud media de los mensajes generados por los usuarios. Analizando con más detalle el comportamiento de la ocupación de la cola de transmisión, podemos observar que las curvas sufren dos fenómenos:

1. El valor máximo de ocupación se alcanza para cargas de tráfico cada vez mayores cuanto mayor es el tamaño de los mensajes.
2. El valor máximo de ocupación es tanto más pequeño cuanto menor sea la longitud de los mensajes. Obsérvese que el número total de usuarios en las simulaciones es 100, y que para el peor caso ($L_d=1$), el valor de la ocupación media de la cola de transmisión es aproximadamente 50, es decir, la mitad de los usuarios presentes en el sistema.

Es posible comprobar que el primer fenómeno es achacable a uno de los errores de detección de los estados de los minislots detallados en el punto 3.10.2: el error ($S/k=0$). En efecto, este error es el factor limitante que causa este comportamiento. Cada vez que se produce un error de este tipo se genera una posición ‘vacía’ en la cola de transmisión. Cuando la carga del sistema de resolución de colisiones se hace suficientemente grande, la probabilidad de que se produzca el error es cada vez mayor, puesto que aumenta la interferencia que sufren todos los minislots de control. De este modo, un gran número de minislots de control vacíos se transforman en posiciones vacías de la cola de transmisión. Estas posiciones no utilizarán recursos de transmisión, pero sí constituyen una cierta ocupación de la cola.

Por otro lado, el segundo fenómeno se produce también a consecuencia del número creciente de errores en la detección de los estados de los minislots. En este caso, sin embargo, son los errores ($E/k=1$) y ($C/k=1$) los factores limitantes. Estos errores transforman peticiones de transmisión que deberían tener éxito en peticiones perdidas o colisionadas. En ambos casos, el error de detección produce que los usuarios afectados, en contra de lo que deberían, no entren en la cola de transmisión. Si la carga de peticiones es grande, ocurre que estos errores se hacen frecuentes y evitan que la ocupación de la cola de transmisión supere un cierto umbral, con lo que su valor promedio decrece.

La Figura 69 evidencia el incremento de carga de tráfico de peticiones que se genera cuando la longitud media de los mensajes decrece. Este hecho es ilustrado por la mayor ocupación de la cola de resolución de colisiones (mayor número de colisiones producido). En el peor caso ($L_d=1$) la ocupación media de la cola alcanza las 10 unidades. Esto indica que, en promedio, al menos se están generando 10 colisiones por slot, lo cual indica un tráfico medio de al menos 20 accesos por slot.

4.3.2 Análisis del comportamiento de los errores de detección de minislots

En las siguientes figuras (de la Figura 70 a la Figura 75) se muestran la frecuencias relativas (errores por slot) con las que se producen los seis errores posibles de detección en los minislots de control (véase su definición en el punto 3.10.2).

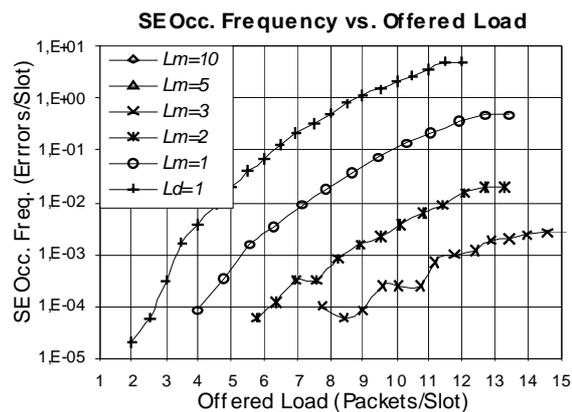


Figura 70. Frecuencia relativa del error ($S/k=0$)

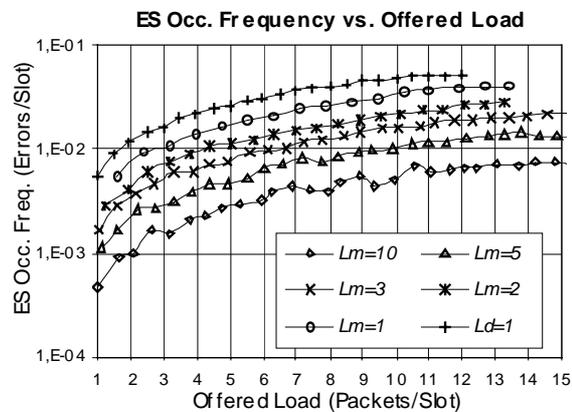


Figura 71. Frecuencia relativa del error ($E/k=1$)

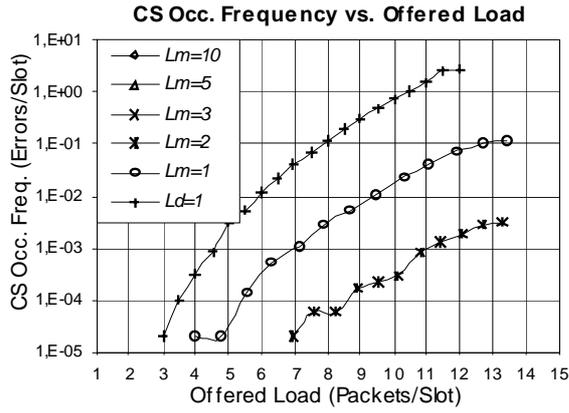


Figura 72. Frecuencia relativa del error ($C/k=1$)

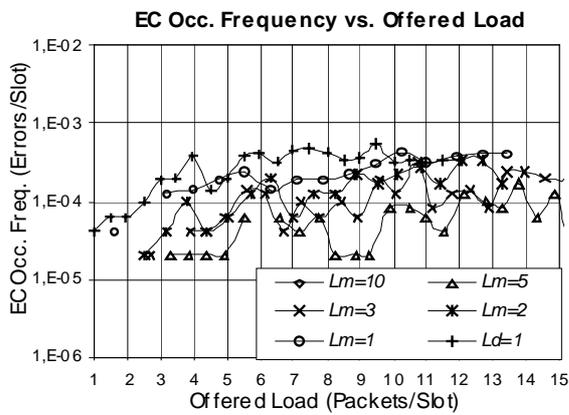


Figura 73. Frecuencia relativa del error ($E/k>1$)

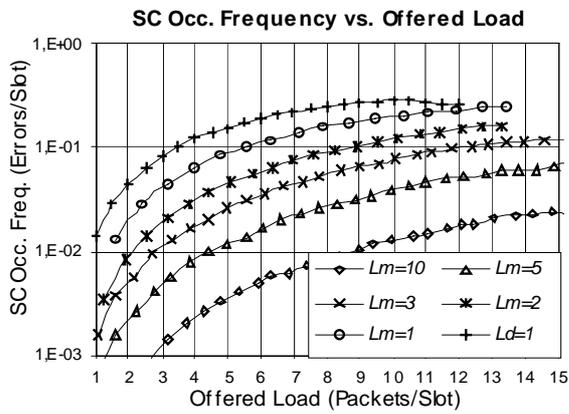


Figura 74. Frecuencia relativa del error ($S/k>1$)

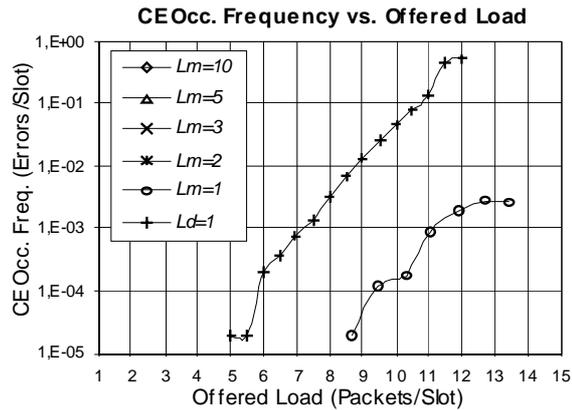


Figura 75. Frecuencia relativa del error ($C/k=0$)

Debemos reseñar que estas frecuencias de error (errores por slot) dependen de dos valores. Por un lado de las probabilidades condicionadas del error (véase el punto 3.10.2) y por otro de la probabilidad absoluta de que se produzca el suceso condicionante. Es decir, a medida que la carga de tráfico aumenta, el número de minislots que no reciben ninguna petición de acceso ($k=0$) va disminuyendo, lo que a su vez afecta a los errores que tienen esta condición previa. Para ilustrar este fenómeno, es posible mostrar el número medio de minislots de control en cada estado por cada slot del sistema. Las Figura 76, Figura 77 y Figura 78 muestran estos valores para las diferentes estadísticas del tráfico ofrecido y cargas totales de tráfico.

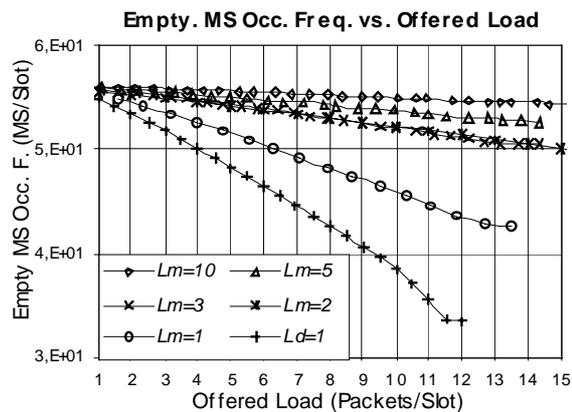


Figura 76. Minislots vacíos por slot en función de la carga

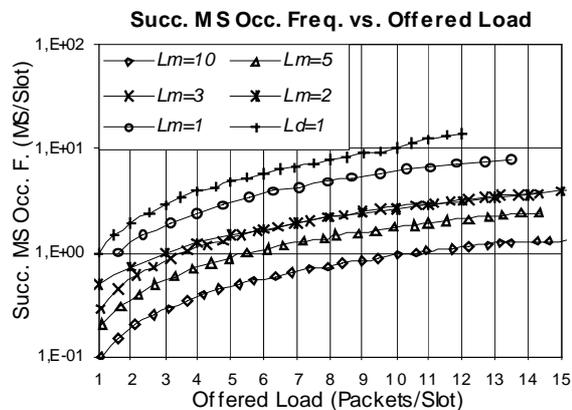


Figura 77. Minislots con un acceso por slot en función de la carga

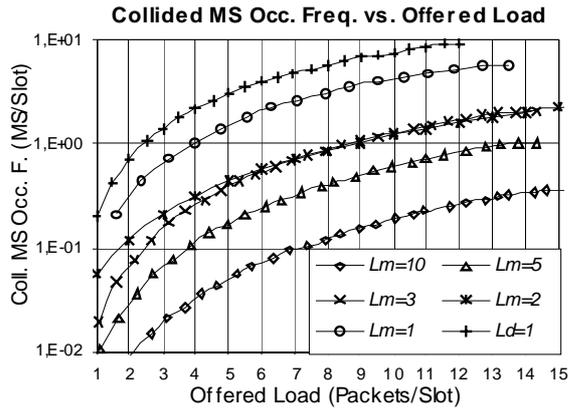


Figura 78. Minislots con más de un acceso por slot en función de la carga

En los siguientes puntos se comentan con detalle las causas y efectos de cada uno de los posibles errores que se dan en el sistema. Recordemos la notación empleada para los errores. La primera letra indica el estado detectado por la estación base, donde:

- E denota el estado *vacío*.
- S denota el estado *éxito*.
- C denota el estado *colisión*.

Por otro lado, el valor de k indica el número de usuarios que realizaron la transmisión de sus peticiones de acceso de forma simultánea en el minislot de control considerado.

- *Error* ($S/k=0$): Tan sólo toma valores significativos para $L_m \leq 3$ paquetes/mensaje. En el caso de tener mensajes de longitud determinista igual a la unidad (el peor caso) alcanza valores incluso por encima de la unidad, indicando que se produce al menos un error de este tipo en cada uno de los slots de tiempo del sistema. Esto confirma lo indicado anteriormente sobre la cota superior de la ocupación de la cola de transmisión. Por tanto, se comprueba que este es el evento limitante cuando la carga del subsistema de resolución de colisiones se hace demasiado alta. Obsérvese que a pesar de que el suceso condicionante ($k=0$) ocurre con menos frecuencia cuanto mayor es esta carga, la frecuencia del error en sí mantiene su crecimiento. Este hecho marca de un modo evidente el abrupto crecimiento del valor de la probabilidad condicionada del error. En el punto 3.10.3 se comentó que el más crítico de los errores posibles en la detección de los estados de los minislots, desde el punto de vista del funcionamiento del algoritmo del protocolo, es el error ($S/k>1$), ya que es el que produce el evento más difícil de recuperar con las reglas del protocolo. Sin embargo, en este punto deberíamos puntualizar que para el caso de tráfico con mensajes pequeños y muy frecuentes, el factor limitante es el error ($S/k=0$).
- *Error* ($C/k=0$): Este error sólo toma valores significativos para $L_m=1$ y para $L_d=1$, donde alcanza valores próximos a la unidad. Contribuye a las pérdidas de rendimiento reduciendo la capacidad del subsistema de resolución de colisiones, ya que se están generando posición no útiles en la cola de peticiones.
- *Error* ($E/k=1$): Este error fuerza a los usuarios afectados a realizar un intento de acceso adicional, y podemos comprobar que para el peor caso, con la carga máxima,

aproximadamente uno de cada diez usuarios que acceden al sistema deben realizar al menos un intento de acceso extra. Como efecto colateral, se aumenta ligeramente la carga de peticiones de acceso.

- *Error (C/k=1)*: Este error hace que los usuarios afectados vayan a la cola de resolución de colisiones en lugar de entrar en la cola de transmisión. Evidentemente, este hecho introduce un retardo adicional en los mensajes afectados y además genera una carga de tráfico añadida en el subsistema de resolución de colisiones. Si observamos la Figura 72, podemos ver que el incremento de frecuencia de este error es muy acusado cuando la longitud media de los mensajes decrece, llegando a sobrepasar el valor unitario para el caso $L_d=1$.
- *Error (E/k>1)*: Constituye el error con menor impacto en el rendimiento del protocolo, sobretodo para cargas altas de tráfico. En efecto, la probabilidad de que se dé este error es cada vez menor cuando aumenta la carga. Cuando se produce, genera un ligero incremento de la carga de tráfico ofrecido al subsistema de resolución de colisiones, en tanto que obliga a los usuarios a realizar una petición que tiene mayor probabilidad de colisionar que la que deberían enviar para resolver la colisión no detectada. Pero el efecto sobre el rendimiento del protocolo va siendo menor a medida que aumenta la carga de tráfico, con lo que resulta no ser un factor limitante de la capacidad.
- *Error (S/k>1)*: Tal y como se apuntó en el punto 3.10.3, este es el más crítico de los errores posibles desde el punto de vista del algoritmo del protocolo. Podemos ver que para el peor caso ($L_d=1$), casi una de cada 10 colisiones son detectadas erróneamente como éxitos. Sin embargo, este error produce un efecto aparentemente paradójico de reducción de la carga del subsistema de resolución de colisiones. En efecto, algunos usuarios entran antes de lo que deberían en la cola de transmisiones. Algunos de ellos, entonces, tienen que compartir una posición en esta cola, y por tanto tardan mucho más de lo debido en salir de la misma. Es decir, se reduce la carga de peticiones a cambio de aumentar el tiempo de servicio del subsistema de transmisión.

Una vez analizados todos los posibles errores y su repercusión como factor limitante de la capacidad de transmisión de un sistema basado en el protocolo DQRAP/CDMA, se van a exponer las propuestas realizadas para la mejora del rendimiento del protocolo, surgidas a partir de este análisis, así como el estudio del grado en el que cumplen su cometido.

4.4 Asignación dinámica de códigos

En este punto, se van a exponer un conjunto de propuestas de modificación de los mecanismos del protocolo DQRAP/CDMA, de cara a mejorar las prestaciones del mismo para el caso en el que el tráfico de entrada consista en mensajes de longitud determinista igual a la unidad ($L_d=1$).

Si recordamos la descripción del protocolo hecha en el capítulo 3, podemos observar que en él se realiza una asignación fija entre cada una de las posiciones de las colas (tanto de peticiones como de transmisión) y los códigos (recursos de transmisión). Si admitimos un máximo de K transmisiones simultáneas, que se realizan con los K códigos de transmisión, a cada una de las posiciones de la cola le corresponde una única secuencia código. Es decir, se establece una relación biunívoca entre los códigos y las posiciones de las colas.

Para la cola de transmisión, esta es la única posibilidad realizable en la práctica. Esto ocurre debido a que tan sólo es posible que cada secuencia código sea utilizada al mismo tiempo por una única comunicación de usuario.

Sin embargo, no ocurre lo mismo con la cola de peticiones (CRQ), donde la asignación fija es tan sólo un caso particular de todos los posibles. En efecto, es posible implementar otro tipo de asignación entre códigos y colisiones a resolver, de modo que podamos mejorar el rendimiento del subsistema de resolución de colisiones.

La propuesta de DQRAP/CDMA descrita en el capítulo 3 tiene una realización práctica simple. Además, tiene la ventaja adicional de que la estación base sólo debe ‘escuchar’ las transmisiones realizadas con los primeros $RQ+1$ códigos. Estos códigos son los RQ códigos asignados a los elementos ocupados de la CRQ, más el primer código libre usado para las nuevas llegadas de peticiones al sistema. Lo cual implica que las posibles falsas alarmas de los filtros no utilizados no afectarán al rendimiento del sistema de detección de las secuencias de acceso.

Por otro lado, si la carga de tráfico ofrecida al sistema es baja, probablemente los $RQ+1$ primeros códigos constituirán tan sólo un subconjunto de los K códigos disponibles, con lo que se estarán manteniendo ciertos recursos sin utilizar. Sin embargo, los resultados mostrados en el punto 3.9 evidencian que, para condiciones estándar ($L_m=10$), la propuesta inicial de DQRAP/CDMA es suficiente para alcanzar un rendimiento muy próximo al máximo teórico.

Por el contrario, cuando la carga de tráfico se vuelve más desfavorable (los mensajes son más cortos) y se mantienen los valores de los parámetros hasta ahora descritos, el rendimiento del protocolo sufre un deterioro significativo. Las causas de este fenómeno son:

Los usuarios que están fuera del sistema (que aún no han entrado en ninguno de los dos subsistemas) tan sólo disponen de los minislots de control de un solo código para enviar sus peticiones. Para situaciones de carga elevada de tráfico, éstos pueden resultar insuficientes. Ello conlleva que la probabilidad de colisión se haga grande y el número de colisiones en el primer intento de acceso pueda crecer de manera importante.

Los usuarios que ya están en la cola de resolución de colisiones (CRQ) disponen de los minislots de un solo código para resolver sus colisiones. Este código es el que les corresponde según su posición en la cola. Esto hace que, por ejemplo, si $m=3$ y se produjo una colisión en la que intervinieron sólo 2 usuarios, existe un 33% de probabilidades de que vuelvan a colisionar en su siguiente contienda.

Este análisis nos da pie a proponer algunas modificaciones al protocolo de cara a mejorar su rendimiento.

4.4.1 Aumento de la tasa de resolución de colisiones

La primera de las propuestas trata de reducir el tiempo que necesitan los usuarios para resolver las colisiones en las que se vieron envueltos, es decir, mejorar la tasa de resolución de colisiones. La manera de hacer esto es reducir la probabilidad de que se generen nuevas colisiones una vez los usuarios ya han entrado en la cola CRQ.

La idea es: aumentar el número de minislots asignados a cada posición de la cola CRQ reduce la probabilidad de tener una nueva colisión en las contiendas que deben resolverse.

Una técnica posible para conseguir este cometido es aumentar el número de minislots por slot de tiempo, es decir, aumentar m . Esta técnica no resulta conveniente, ya que aumenta el tamaño del campo de control de los slots y por tanto reduce la eficiencia en la transmisión de datos.

La solución propuesta consiste en dejar inalterado el valor de m y asignar un cierto número de códigos, que llamaremos C_a , a cada una de las posiciones de la cola de resolución de colisiones. Para que esta asignación sea efectiva, se deberá cumplir que $C_a > 1$. De este modo, cada uno de los usuarios en esta cola dispondrá de $m \cdot C_a$ minislots de control para resolver sus contiendas. Puesto que se cumple que $(m \cdot C_a > m)$, la probabilidad de que se vuelva a producir una colisión será menor que en el diseño inicial. La propuesta descrita de asignación de códigos se muestra gráficamente en la Figura 79.

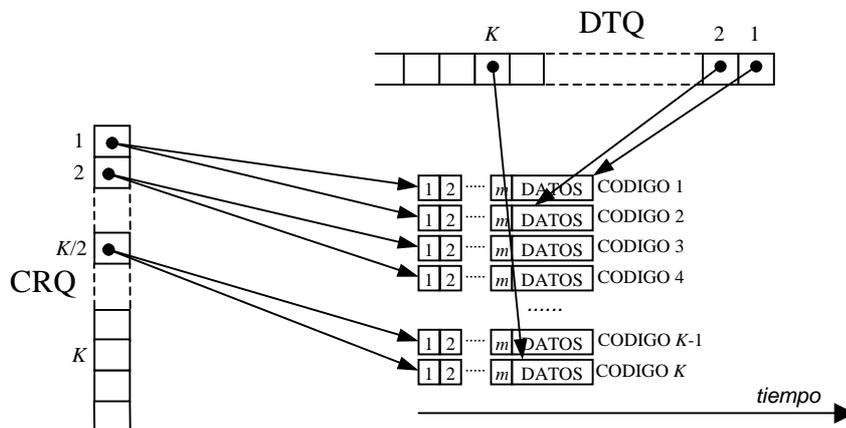


Figura 79. Asignación múltiple de códigos a posiciones de CRQ

En el ejemplo de la figura, se asignan dos códigos a cada posición de la cola de resolución de colisiones ($C_a=2$). Es evidente que el número total de colisiones que podrán tratar de ser resueltas será únicamente la mitad que para el caso de no usar esta asignación múltiple. Sin embargo, gracias al incremento de la tasa de resolución de colisiones (menor probabilidad de generar más colisiones en las contiendas sucesivas), es posible esperar que, en promedio, el número de posiciones ocupadas de CRQ sea menor. Nótese que esta modificación tan sólo requiere de unos ligeros cambios en las reglas del protocolo descritas en 3.4, cambios que serán detallados seguidamente. Ninguno de los elementos esenciales del funcionamiento del algoritmo necesitan ser alterados. El esquema básico y el modelo del protocolo permanecen también sin cambios significativos. Tan sólo se modifican los tiempos de servicio del subsistema de resolución de colisiones.

Es, por tanto, una modificación de realización práctica abordable y que como veremos más adelante mejora de manera sustancial la respuesta del protocolo a situaciones de carga elevada y racheada.

La única regla que debe modificarse es la RTR-2, que queda como sigue:

- **RTR-2:** Si un usuario cumple que $0 < p_{RQ} < \lfloor C / C_a \rfloor$, selecciona al azar una secuencia de acceso en el rango $K_{(p_{RQ}-1) \cdot C_a + 1}$ y $K_{p_{RQ} \cdot C_a}$ de modo que la probabilidad de escoger

cualquiera de estos códigos sea la misma. A continuación elige al azar uno de los m minislots, y transmite una petición de acceso en él haciendo uso del código escogido.

Desde la Figura 80 a la Figura 82 se muestra el retardo de los paquetes y el *throughput* para diferentes valores de C_a .

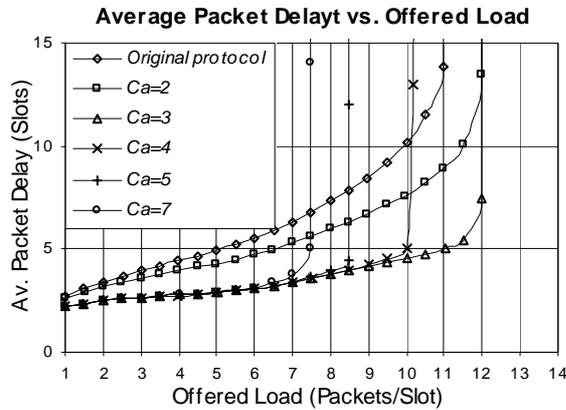


Figura 80. Retardo de paquetes medio para diferentes C_a

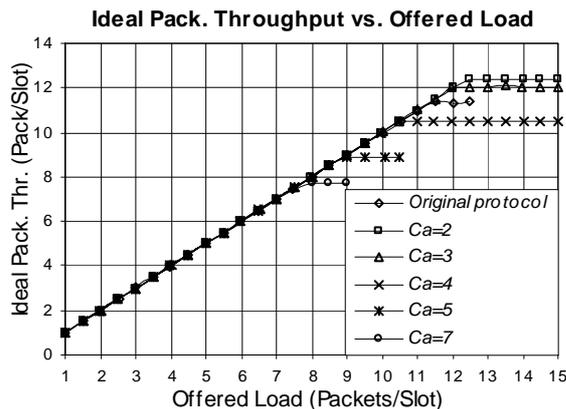


Figura 81. *Throughput* de paquetes para diferentes C_a

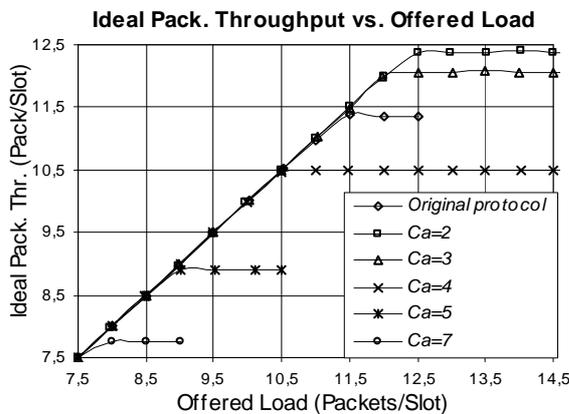


Figura 82. Detalle de la Figura 81

Observamos la mejora del rendimiento que tenemos en el caso $C_a=2$, incluso superior a la obtenida para $C_a=3$. También se observa que el resto de valores de C_a no generan ninguna mejora apreciable. Es más, incluso se puede observar como valores de C_a por encima de 3 son

inaceptables, puesto que el retardo comienza a divergir para valores de tráfico de entrada muy bajos. Por el contrario, $C_a=2$ y $C_a=3$ mantienen el retardo acotado hasta cargas de entrada de hasta 12 paquetes por slot, aproximadamente. Recordemos que todos los paquetes transmitidos son de longitud 1 paquete, por lo que este valor de carga representa una mejora respecto al diseño inicial (sin incluir la modificación propuesta).

De la Figura 80 podemos inferir, en una primera observación, que el valor óptimo se obtiene para $C_a=3$.

Las curvas de *throughput* (Figura 81 y Figura 82) confirman la mejora introducida, a pesar de que en este caso se observa que el valor $C_a=2$ ofrece un rendimiento ligeramente mejor que $C_a=3$. Para entender esta aparente contradicción entre el comportamiento del retardo y el caudal debemos apelar al modo de operar de DQRAP/CDMA.

En efecto, cuanto mayor es C_a más rápido se resuelven las colisiones y por tanto el número de posiciones ocupadas en CRQ es menor. Este hace que el caso $C_a=3$ ofrezca un mejor comportamiento en cuanto a retardo.

Por otro lado, debe tenerse en cuenta que cuanto menor sea el valor de C_a mayor es el número de colisiones que pueden tratar de resolverse cada slot de tiempo. Por consiguiente, incluso aunque la tasa de resolución de colisiones sea menor para el caso $C_a=2$, de este modo el sistema es capaz, en media, de dar servicio a un mayor número de usuarios cada slot. Por tanto, en el caso límite, puede soportar un tráfico ligeramente superior.

En cualquier caso, la pequeña diferencia en el comportamiento del *throughput* no compensa la pérdida de respuesta en retardo, por lo que concluimos que el valor más adecuado es:

- $C_a=3$.

4.4.2 Explotación de los códigos inactivos

La siguiente modificación propuesta trata de explotar el hecho de que, en determinados momentos, el número de secuencias código que serán usadas para los intentos de acceso es menor que el número máximo disponible. En el diseño inicial, los usuarios recién llegados al sistema sólo disponen del primer código 'libre' (no asignado a ninguna posición de CRQ) para realizar sus peticiones de acceso. Esta estrategia mantiene la ventaja de que la estación base únicamente debe 'escuchar' un subconjunto de los códigos disponibles en el sistema, ahorrando recursos de cómputo que pueden ser utilizados en otras tareas. Sin embargo, en este caso los usuarios fuera del sistema disponen de tan solo m minislots de control para realizar sus peticiones de acceso, con lo que la probabilidad de colisión en el primer intento de acceso crece notablemente en situaciones de carga de tráfico elevada.

Existen al menos dos alternativas para mejorar este aspecto del protocolo:

1. Permitir que los usuarios recién llegados tengan disponibles *todos* los códigos que dejan sin utilizar los que están en la cola de resolución de colisiones. Es decir, hacer que las peticiones de los nuevos usuarios puedan utilizar cualquier código desde el K_{RQ+1} hasta el K_K (todos los libres). La realización de esta propuesta tan sólo requiere modificar la regla RTR-1 del algoritmo del protocolo.

- Reservar un cierto número de secuencias código, que llamaremos C_r , de modo que estén siempre a disposición de los usuarios que realizan su primer intento de acceso. Esta estrategia evita que los usuarios recién llegados puedan encontrarse que no tienen códigos disponibles, pero por otro lado reduce el número máximo de códigos disponibles para los usuarios que están en CRQ. En particular, los usuarios en CRQ sólo dispondrán de $K - C_r$ códigos disponibles. Este esquema de asignación de códigos se muestra en la Figura 83.

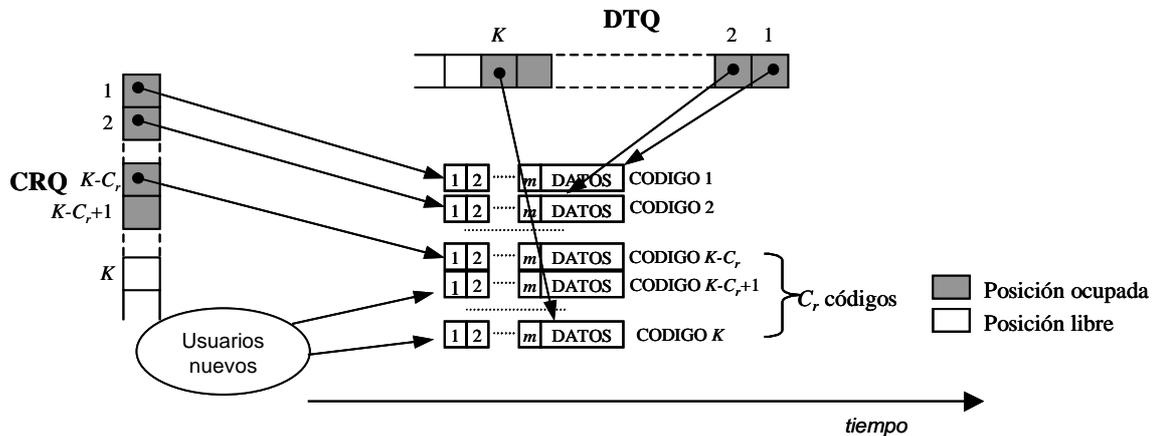


Figura 83. Asignación de códigos de acceso libres

A modo de ejemplo, en la Figura 83 se observa que la posición $K - C_r + 1$ de la cola de resolución de colisiones está ocupada. Siguiendo las reglas del protocolo, esta posición debería recibir servicio del sistema (una secuencia código para resolver la contienda), puesto que si $C_r > 1$ se cumple que $K - C_r + 1 < K$. Sin embargo, con la modificación propuesta, y manteniendo que $C_r > 1$, no quedarán códigos suficientes para que se le asigne uno y por tanto no podrá recibir esta asignación.

Estas dos alternativas son en realidad complementarias, y por tanto incluiremos ambas en el diseño del algoritmo del protocolo.

Las reglas que debe modificarse para implementar estas modificaciones son las dos que conforman las RTR. Estas reglas deben ser ahora:

- RTR-1:** Si un usuario cumple que $pTQ=0$ y $pRQ=0$ (usuario recién llegado) y tiene datos listos para ser transmitidos, selecciona aleatoriamente un código en el rango desde $K_{\min(RQ, K-C_r+1)}$ hasta K_K , de modo que la probabilidad de escoger cualquiera de estos códigos sea la misma para todos ellos. Después selecciona al azar uno de los m minislots y envía una petición de acceso en él haciendo uso del código escogido.
- RTR-2:** Si un usuario cumple que $0 < pRQ \leq K - C_r$, selecciona al azar uno de los m minislots, y transmite una petición de acceso en él haciendo uso del código K_{pRQ} .

Seguidamente se muestran las curvas del retardo y el throughput del sistema cuando se incluye estas nuevas reglas propuestas, en función del valor del parámetro C_r .

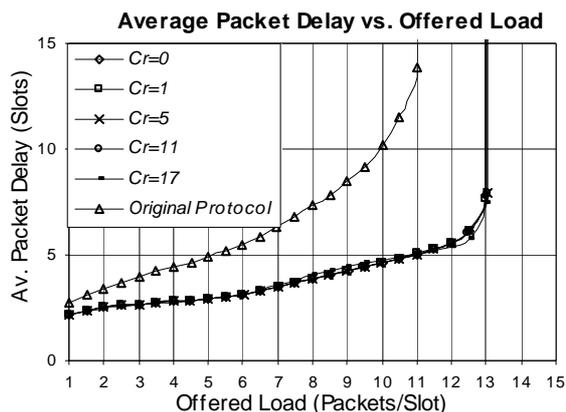


Figura 84. Retardo de paquetes en función de C_r

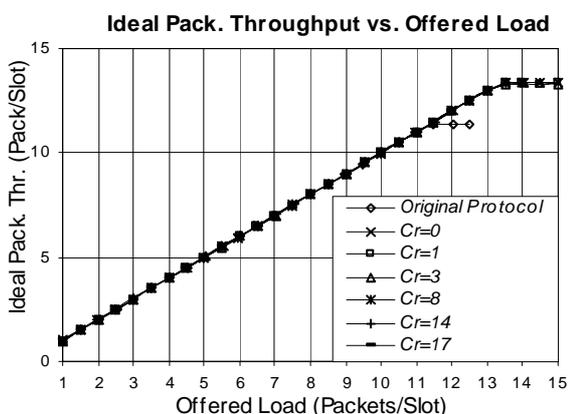


Figura 85. Throughput de paquetes en función de C_r

Puede observarse que el rendimiento obtenido al añadir la modificación propuesta, para todos los valores de C_r , mejora apreciablemente respecto al sistema inicial (sin la modificación). Por otro lado, también se observa que la sensibilidad de la eficiencia del protocolo a la variación del parámetro C_r es muy pequeña. Este hecho indica que la mejora obtenida se debe principalmente al hecho de aprovechar los códigos no utilizados, más que del hecho de ajustar con precisión el valor del parámetro C_r .

Nótese que, mientras que en el protocolo original el retardo de paquetes diverge para cargas de tráfico ofrecido de aproximadamente 11 paquetes por slot, haciendo uso de la modificación se alcanzan cargas máximas de unos 13'3 paquetes por slot.

Debemos resaltar un detalle que puede pasar desapercibido de estas curvas. Una de ellas corresponde al caso $C_r=17$, lo que se corresponde con el número total de códigos disponibles en el sistema. En esta situación, la totalidad de los códigos disponibles se reservan para las llegadas de los nuevos accesos. De este modo, no existe ninguna asignación de secuencias código a las posiciones de la cola de resolución de colisiones, lo que implica, para que el sistema sea estable, que no hacemos uso de esta cola. Si no fuera así, las colisiones producidas no podrían ser resueltas tras colocarse en la cola de resolución. Es decir, consideramos todas las peticiones de acceso como 'primeros' accesos. Este escenario se corresponde al caso genérico de un sistema de reserva puro en el que los usuarios que tienen datos a transmitir envían una petición de transmisión y simplemente se colocan de forma ordenada en una cola de transmisión, a medida que sus peticiones tienen éxito. Si una petición colisiona o no es detectada, el usuario implicado simplemente se mantiene fuera de la cola de transmisión y

vuelve a intentar enviar una petición de acceso en un minislot escogido al azar en el siguiente slot de tiempo.

De la observación de estos resultados no es evidente la elección de un valor de C_r que ofrezca las prestaciones óptimas. El estudio analítico de este problema queda fuera del ámbito de nuestro trabajo. Sin embargo, es razonable escoger un valor intermedio que evite las situaciones extremas al mismo tiempo que se aprovechan las ventajas de la modificación propuesta. De aquí en adelante, escogeremos el valor

- $C_r=K/2$, redondeado por defecto.

En nuestro caso, $C_r=8$. Como veremos en el siguiente punto, es posible encontrar un valor óptimo de este parámetro cuando la modificación se realiza en coordinación con la propuesta del punto 4.4.1

4.4.3 Solución Mixta

En este punto nos planteamos el estudio de las prestaciones de la solución mixta que combina las dos modificaciones propuestas en los puntos 4.4.1 y 4.4.2. Estas dos modificaciones tienen dos propósitos claramente definidos:

- La primera de ellas acelera el procedimiento de resolución de colisiones, es decir, el tiempo de servicio del subsistema de resolución de colisiones.
- La segunda reduce el número de colisiones que se producen en el primer intento de acceso de cualquier usuario, y como consecuencia reduce también la ocupación de la cola de resolución de colisiones.

Estas consideraciones muestran la complementariedad de ambas modificaciones. Se ha simulado el sistema con la inclusión de dichas modificaciones. Desde la Figura 86 hasta la Figura 89 se muestran los resultados de retardo y *throughput* obtenidos para diferentes valores de los parámetros C_a y C_r . Las gráficas se han dividido en dos grupos diferenciados, uno para $C_a=2$ y otro para $C_a=3$.

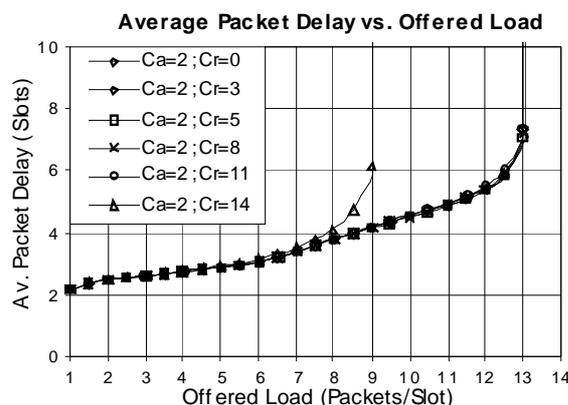


Figura 86. Retardo de paquetes para $C_a=2$ y diferentes valores de C_r

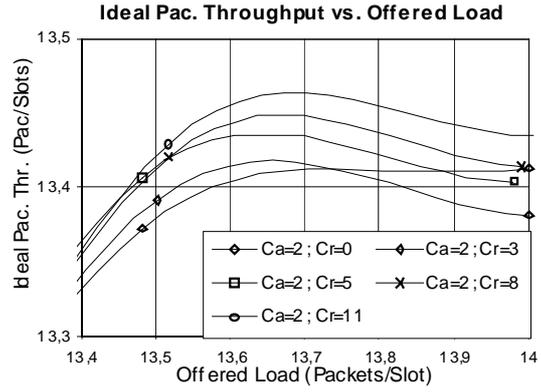
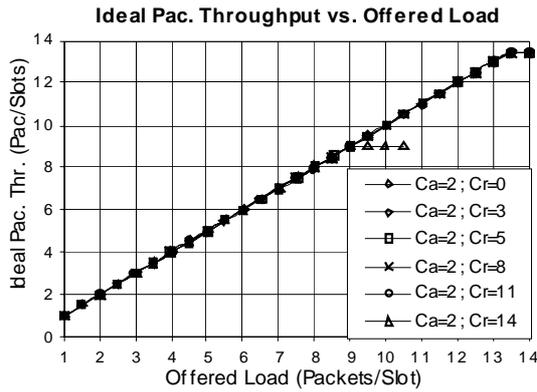


Figura 87. Throughput (y detalle) para $C_a=2$ y diferentes valores de C_r

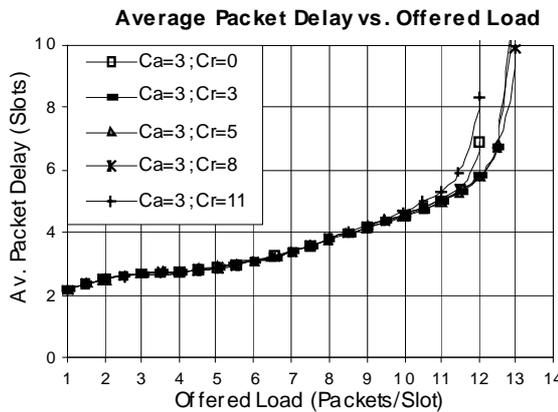


Figura 88. Retardo de paquetes para $C_a=3$ y diferentes valores de C_r

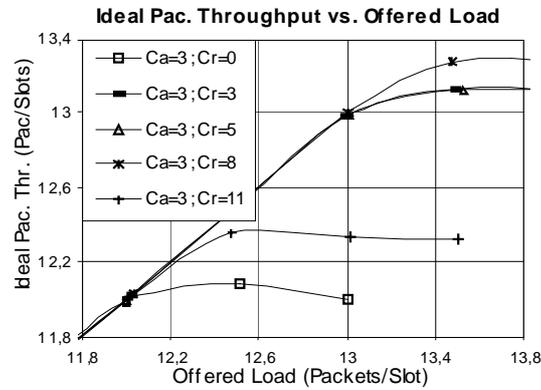
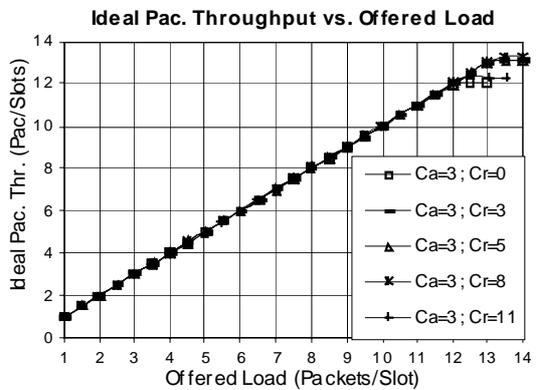


Figura 89. Throughput (y detalle) para $C_a=3$ y diferentes valores de C_r

De la observación de estas gráficas se puede inferir que:

- Todas las curvas obtenidas con $C_a=2$ son prácticamente equivalentes, excepto para el caso $C_r=14$. En esta situación, estamos reservando 14 de los 17 códigos disponibles para los accesos de los nuevos usuarios. Por tanto, se producirán muy pocas colisiones en los primeros accesos al sistema. Por el contrario, los usuarios situados en la cola de resolución de colisiones dispondrán tan solo de 3 códigos para resolver sus contiendas. Además, para el caso $C_a=2$ resulta que tan sólo la primera posición de la cola puede recibir servicio, es decir, sólo los usuarios en la cabeza de la cola CRQ tienen asignadas secuencias código para tratar de resolver sus colisiones. Se muestra de forma evidente que la mejora en la tasa de éxitos en el primer acceso no compensa

la pérdida en la velocidad de resolución de las colisiones ya producidas que se encuentran en la cola CRQ.

- Todos los gráficos obtenidos con $C_a=3$ muestran unas curvas que marcan un peor rendimiento general del sistema que para cualquiera de las curvas obtenidas con $C_a=2$. De entre estas curvas, la que presenta un mejor comportamiento se corresponde con $C_r=8$, pero en cualquier caso, peor que el obtenido con cualquiera de las correspondientes a $C_a=2$.
- Para cada combinación de valores de C_a y C_r puede calcularse el máximo *throughput* estable que el sistema es capaz de servir.

Haciendo uso de esta última consideración, se han realizado una serie de simulaciones para diferentes combinaciones de los parámetros C_a y C_r . Los resultados obtenidos para el máximo *throughput* en cada escenario se han representado en la Figura 90. Cada uno de los puntos de cada curva indica en el eje de ordenadas el valor del máximo caudal que es capaz de servir el sistema de manera estable (manteniendo acotado el retardo de los paquetes).

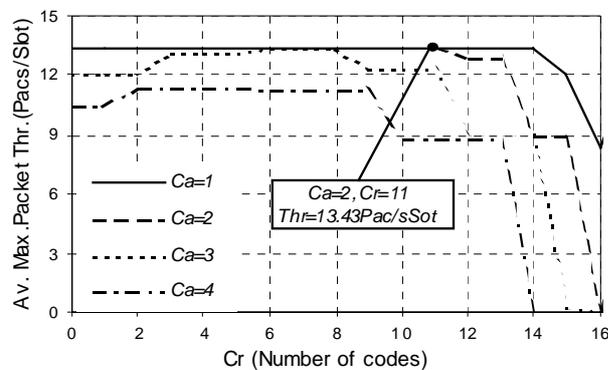


Figura 90. *Throughput* máximo para diferentes valores de C_a y C_r

Se ha indicado en la gráfica el máximo caudal alcanzado por esta solución mixta, que se produce para $C_a=2$ y $C_r=11$, en el que el sistema sirve un caudal estable de 13'4 paquetes/slot.

4.4.4 Comparación de rendimientos

Las modificaciones propuestas hasta en este momento afectan a la política de asignación de códigos a las posiciones de la cola de resolución de peticiones. En este punto vamos a analizar la comparativa de los resultados obtenidos de cara a decidir el más adecuado para mejorar el rendimiento y encontrar el mecanismo óptimo. La Figura 91 y la Figura 92 muestran los resultados obtenidos.

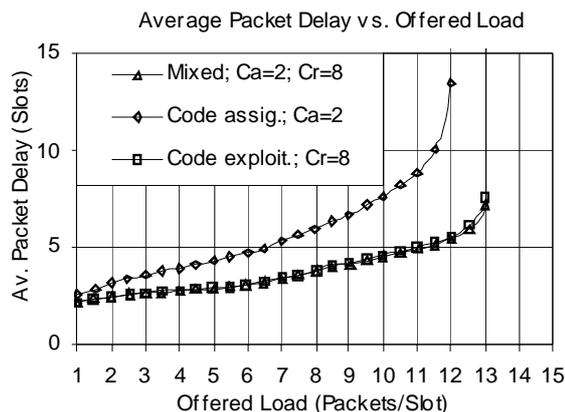


Figura 91. Comparativa de retardo con las modificaciones propuestas

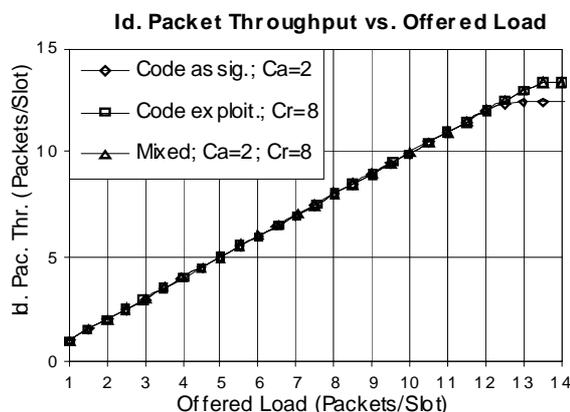


Figura 92. Comparativa de throughput con las modificaciones propuestas

Estos resultados muestran que el peor rendimiento se obtiene para el caso en el que simplemente se utiliza la propuesta de asignar más de una secuencia código a cada posición de la cola de resolución de peticiones, con el valor $C_a=2$. Las otras dos modificaciones ofrecen rendimientos muy similares.

- Observando detalladamente las curvas, podemos concluir que la solución mixta con los valores $C_a=2$ y $C_r=8$ es la que presenta un mejor comportamiento conjunto de retardo y throughput.

Esta solución se muestra como un buen compromiso entre la reducida probabilidad de colisión en el primer intento de acceso y la mejorada tasa de resolución de las colisiones situadas en la cola CRQ. La idea es reservar la mitad de los códigos disponibles para las llegadas de peticiones nuevas, mientras que se asignan dos secuencias código a cada una de las posiciones de la cola de resolución de colisiones.

Es interesante reseñar que las modificaciones propuestas presentan algunas ventajas adicionales para el funcionamiento del protocolo en determinadas situaciones. En efecto, reservar un cierto número de códigos para los primeros intentos de acceso de los usuarios recién llegados al sistema asegura una mayor robustez del protocolo a cambios bruscos de la carga de tráfico.

Analicemos el siguiente ejemplo: supongamos que el sistema está trabajando en condiciones de tráfico cerca del límite de capacidad de transmisión. Supongamos también que en ese momento se genera una llegada de un gran número de mensajes, de modo que se produce un

brusco incremento de la carga de tráfico cuya duración es limitada en el tiempo. Si no se reserva ningún código para las nuevas llegadas, los usuarios recién llegados deberán esperar para acceder. En el momento que algún código quede liberado, todos los usuarios que estaban esperando tratarán de enviar sus peticiones haciendo uso de esta única secuencia código, con lo que probabilidad de colisión es muy elevada. Por tanto, el hecho de reservar ciertos recursos para las nuevas llegadas, es decir, impedir que las colisiones en curso puedan copar todos los recursos disponibles, permite al sistema absorber mejor los cambios bruscos de carga de tráfico. Los nuevos usuarios son colocados rápidamente en las colas distribuidas, lo que redundará en una mejora en el rendimiento general del sistema.

4.5 Optimización con variación del número de minislots por código (m)

Las modificaciones propuestas en el punto 4.4.3 permiten obtener una mejora significativa en el rendimiento del protocolo a través de unos cambios muy simples de las reglas que definen el funcionamiento del mismo. En este punto vamos a analizar el efecto de la variación de otro de los parámetros del sistema: el número de minislots de acceso por slot de datos (m).

Para cada uno de los valores considerados de m , se ha realizado un estudio similar al del punto 4.4.3 de cara a buscar los valores óptimos de C_a y C_r en cada caso. Las siguientes gráficas muestran las curvas de rendimiento (retardo y *throughput*) obtenidas para los valores encontrados.

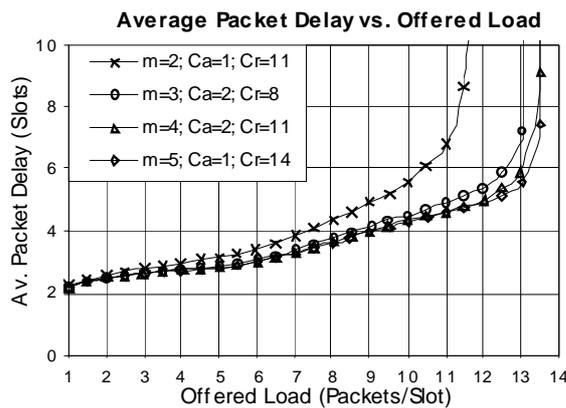


Figura 93. Retardo óptimo para diferentes valores de m .

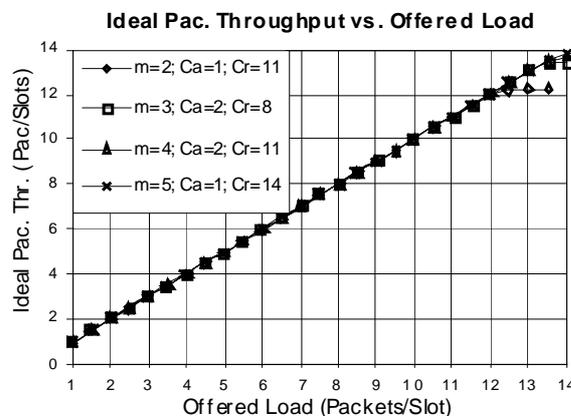


Figura 94. *Throughput* óptimo para diferentes valores de m

Se observa en estas curvas que el rendimiento general empeora sensiblemente para valores de m menores que 3, mientras que la mejora es muy poco significativa para valores mayores que

3. Esto es debido a que un mayor número de minislots de acceso redundante en un mayor tamaño de la cabecera de cada slot que no se utiliza para la transmisión de datos, lo que compensa en cierta manera la mejora obtenida en la tasa tanto de generación como de resolución de colisiones.

4.6 Rendimiento general cuasi-óptimo

Llegados a este punto, se ha descrito una visión detallada del rendimiento del protocolo para diferentes situaciones de tráfico. Es necesario puntualizar algunos aspectos:

- El protocolo, en su versión original, presenta un rendimiento muy próximo al óptimo teórico tan sólo cuando los mensajes generados tienen una longitud, medida en paquetes, mayor que la unidad. Cuando los mensajes son muy pequeños y frecuentes, el rendimiento decrece de manera acusada.
- Las modificaciones propuestas permiten recuperar el comportamiento cuasi-óptimo del protocolo incluso para las peores condiciones de tráfico posibles. En estas condiciones, el rendimiento óptimo se alcanza para un valor de $m=4$.
- Cuando los mensajes tienen una media de longitud de 10 paquetes (caso estándar), el óptimo se alcanza para $m=2$. En este caso se consigue incluso un rendimiento superior al protocolo original (sin las modificaciones) usando $m=3$.

Estamos pues en condiciones de escoger el valor de los parámetros que mejor se ajustan a un caso general de tráfico indeterminado. En principio, podemos suponer que las fuentes de tráfico reales que ofrecerán carga al sistema serán una mezcla heterogénea de mensajes cortos y racheados y otros largos y menos frecuentes. Por tanto, estudiando las gráficas mostradas en el análisis, realizaremos la elección de los valores que serán utilizados de aquí en adelante. El objetivo consiste en introducir el protocolo de acceso DQRAP/CDMA en un ámbito de tráfico heterogéneo en el que deseamos tratar de garantizar una cierta calidad de servicio. Estos valores son:

- Número de minislots por slot de datos: $m=3$
- Número de secuencias código asignadas a cada posición de CRQ, $C_a=2$
- Número de secuencias código reservadas para los primeros accesos, $C_r=8$

4.7 Análisis del rendimiento del protocolo optimizado

Finalmente, vamos a mostrar los resultados detallados obtenidos en las simulaciones realizadas haciendo uso de los parámetros escogidos para el protocolo, con la incorporación de las modificaciones propuestas. Comentaremos los aspectos más relevantes de estos resultados. Cabe reseñar que las gráficas que se muestran a continuación han sido obtenidas para el peor escenario posible de tráfico ofrecido, es decir, para mensajes de longitud determinista igual a la unidad (medida en paquetes) y con carga de tráfico variable. Se han considerado los parámetros de sistema detallados en el punto 4.6. En estas condiciones es cuando se visualiza de una manera más evidente la mejora aportada por las modificaciones introducidas.

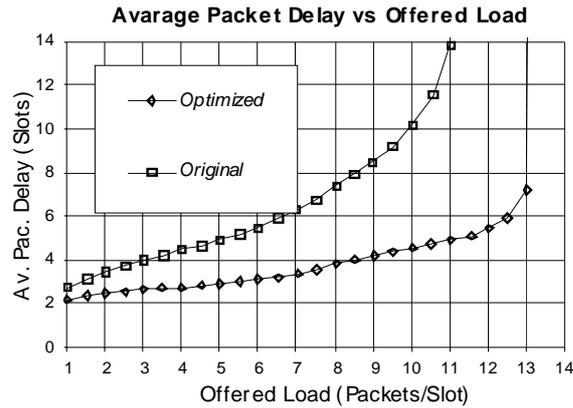


Figura 95. Comparativa de retardo de paquetes

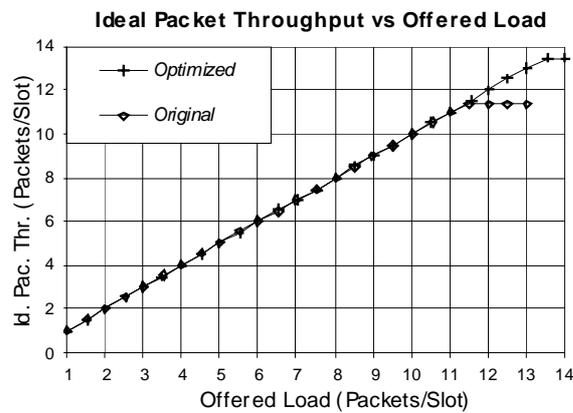


Figura 96. Comparativa de *throughput*

La Figura 95 muestra la mejora introducida por las modificaciones realizadas en el retardo medio de los paquetes. Se comprueba que el grado de mejora aumenta cuando la carga de tráfico es elevada, pero existe para todas las cargas. Por otro lado, la Figura 96 muestra que la máxima carga de tráfico estable soportada por el sistema se ha incrementado en aproximadamente 2 paquetes por slot, alcanzando un caudal máximo de 13,4 paquetes por slot. Este valor se sitúa de nuevo muy próximo al máximo absoluto teórico.

Otro modo de visualizar la mejora obtenida por las modificaciones introducidas es observar las curvas que se muestran en la Figura 97 y en la Figura 98. En estas curvas se representa la función densidad de probabilidad del retardo de los paquetes, tanto para la versión inicial del protocolo (sin las modificaciones) como para la versión optimizada. Las curvas corresponden a dos escenarios de tráfico ofrecido distintos, que corresponden respectivamente con 5 y 10 paquetes por slot.

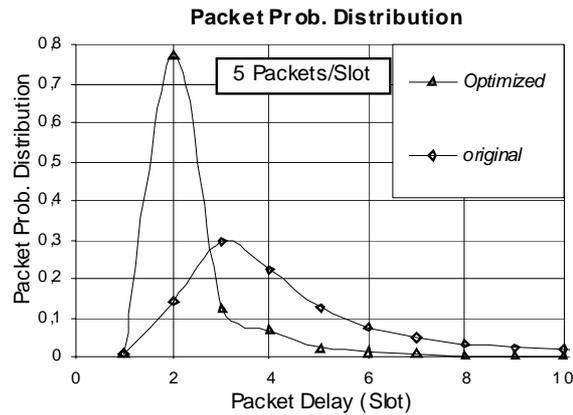


Figura 97. Función densidad de probabilidad del retardo para una carga de 5 p/s

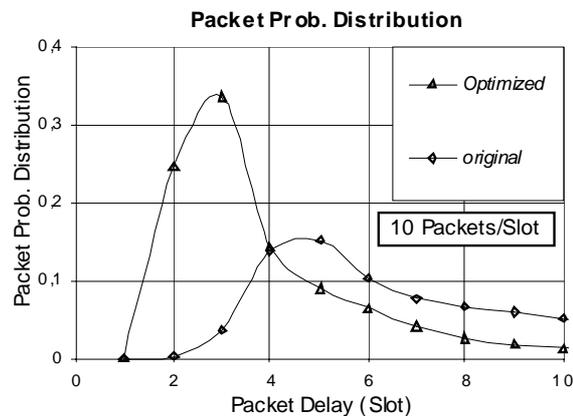


Figura 98. Función densidad de probabilidad del retardo para una carga de 10 p/s

Puede observarse que las funciones densidad de probabilidad de la versión optimizada del protocolo están mucho más ‘concentradas’ en valores más pequeños del retardo, esto es, la probabilidad de que cada paquete transmitido sufra un retardo pequeño es mucho mayor. Este hecho, además de redundar en la reducción de la media del retardo, muestra que la varianza del retardo sufrido sufre también una reducción importante, con la mejora que ello supone en la calidad de las transmisiones. Estas propiedades también pueden corroborarse haciendo uso de la función de distribución de probabilidad o función de probabilidad acumulada, que nos muestra directamente la probabilidad de que los paquetes lleguen con un cierto retardo máximo. Estas curvas corresponden a la Figura 99 y a la Figura 100. Comparando ambas figuras puede comprobarse de nuevo que la mejora introducida por el proceso de optimización es tanto más importante cuanto mayor es la carga de tráfico ofrecida al sistema.

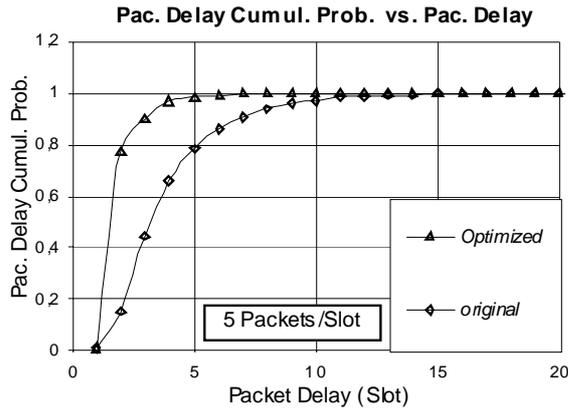


Figura 99. Función de distribución de probabilidad del retardo para una carga de 5 p/s

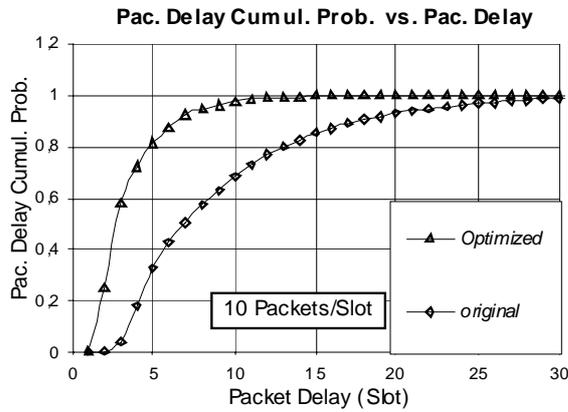


Figura 100. Función de distribución de probabilidad del retardo para una carga de 10 p/s

Seguidamente analizamos las curvas que muestran el comportamiento de otros dos parámetros importantes del sistema: la ocupación media de las dos colas distribuidas. En la Figura 101 se muestra la ocupación media de la cola de transmisión. Observando esta curva se comprueba que la versión optimizada del protocolo mantiene un comportamiento muy similar a la versión original para cargas de tráfico pequeñas. Sin embargo, cuando la carga se incrementa, el valor de saturación de esta ocupación se alcanza para cargas superiores, lo que permite que el sistema sea capaz de soportar en régimen permanente un caudal más elevado de manera estable.

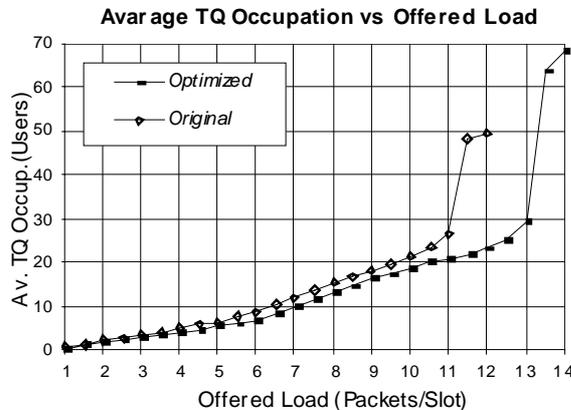


Figura 101. Ocupación media de la cola de transmisión

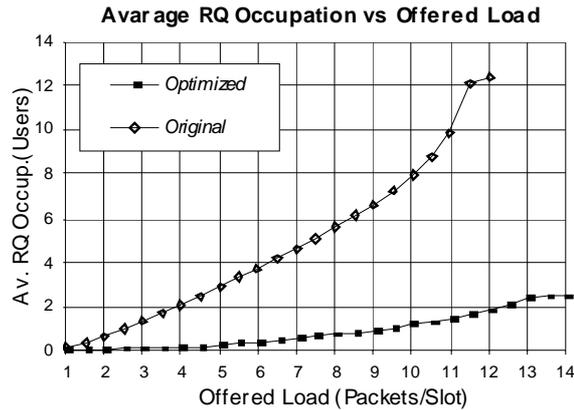


Figura 102. Ocupación media de la cola de resolución de colisiones

La observación de la curva de ocupación media de la cola de resolución de colisiones proporciona una clave primordial de la mejora en el rendimiento obtenido. La Figura 102 muestra este valor en función de la carga. El protocolo optimizado presenta una ocupación media de la cola CRQ seis veces inferior a la versión inicial del mismo. Este hecho muestra de un modo evidente que la estrategia modificada de asignación de recursos para los intentos de acceso mejora de forma acusada el número de colisiones producidas, así como la velocidad a la que éstas son resueltas con éxito.

4.8 Adaptación de velocidades de transmisión

El protocolo DQRAP/CDMA permite conocer a priori en cada momento cuántas transmisiones simultáneas de datos van a realizarse en cada uno de los slots. Esta información puede permitir a los usuarios adaptar sus velocidades de transmisión en función de la interferencia que van a sufrir, manteniendo la calidad medida en términos de probabilidad de error en el bit. En efecto, si el número óptimo de transmisiones simultáneas viene dado por K y el número de transmisiones que realmente se van a dar es menor que $K/2$, los usuarios pueden transmitir usando la mitad de la ganancia de procesamiento nominal (la que se usa cuando transmiten K usuarios) y mantener la calidad de la transmisión.

Recordemos que si usamos la hipótesis gaussiana y despreciamos el ruido térmico, en un sistema perfectamente sincronizado la probabilidad de error en el bit viene dada por:

$$P_b = \frac{1}{2} \operatorname{erfc} \left(\sqrt{\frac{G_p}{m-1}} \right) \quad (4.1)$$

donde G_p es la ganancia de procesamiento y m es el número de transmisiones simultáneas. Vemos por tanto que el cociente $G_p/(m-1)$ es el que marca la calidad de las transmisiones. En principio la ganancia de procesamiento suele variarse únicamente en potencias de dos, es decir, que puede multiplicarse o dividirse por dos. Por tanto, para mantener la calidad deberemos esperar que m también varíe en la misma proporción. Es evidente que si el número de transmisiones simultáneas se reduce a la mitad, éstas podrán realizarse también con la mitad de protección antes las interferencias, es decir, con la mitad de la ganancia de procesamiento.

4.8.1 Algoritmo de selección de velocidad

Supongamos que el sistema está diseñado inicialmente para que se realicen un máximo de K transmisiones simultáneas con una ganancia de procesamiento G_p . Los móviles conocen el número de transmisiones que van a realizarse en cada slot a partir del conocimiento del valor del parámetro TQ del protocolo DQRAP/CDMA. Cada vez que tienen que realizar una transmisión de un paquete de datos, escogen la ganancia de procesamiento que utilizarán según el siguiente algoritmo:

- Si $TQ > K/2$ la transmisión se realiza con ganancia de procesamiento G_p .
- Si $TQ \leq K/2$ pero $TQ > K/4$, la transmisión se realiza con ganancia de procesamiento $G_p/2$.
- Si $TQ \leq K/4$ pero $TQ > K/8$, la transmisión se realiza con ganancia de procesamiento $G_p/4$.
- Si $TQ \leq K/8$ pero $TQ > K/16$, la transmisión se realiza con ganancia de procesamiento $G_p/8$.
- Este proceso de selección de la ganancia de procesamiento se detiene cuando el valor de ganancia que se debería utilizar $G_p/2^n$ es menor que la mínima ganancia de procesamiento admitida en el sistema o bien si el valor de $K/2^n$ es ya menor que la unidad.

4.8.2 Ejemplo de elección de velocidades

Analicemos cómo podemos adaptar las velocidades de transmisión en el caso estándar estudiado hasta el momento. Supongamos que $K=17$ y que las transmisiones se realizan con $G_p=64$, con lo que cada 10 ms se transmiten paquetes de $L=600$ bits. Veamos el criterio con el que los usuarios ajustan la ganancia de procesamiento con la que realizan sus transmisiones:

- Si $TQ > 7.5$, es decir, puesto que TQ es un valor entero, si $TQ \geq 8$, la transmisión del paquete se realiza con ganancia de procesamiento $G_p=64$, es decir se envía un paquete de $L=600$ bits.
- Si $TQ < 7.5$ pero $TQ > 4.25$, es decir si $TQ \leq 7$ y $TQ \geq 5$, la transmisión del paquete se realiza con ganancia de procesamiento $G_p=32$, es decir se envía un paquete de $L=1200$ bits.
- Si $TQ < 4.25$ pero $TQ > 2.125$, es decir si $TQ \leq 4$ y $TQ \geq 3$, la transmisión del paquete se realiza con ganancia de procesamiento $G_p=16$, es decir se envía un paquete de $L=2400$ bits.
- Si $TQ < 2.125$ pero $TQ > 1.0625$, es decir si $TQ=2$, la transmisión del paquete se realiza con ganancia de procesamiento $G_p=8$, es decir se envía un paquete de $L=4800$ bits.
- En el caso $TQ=1$, la transmisión del paquete puede realizarse con cualquier ganancia de procesamiento ya que es la única transmisión simultánea. Se decide por simplicidad que se utilizará $G_p=8$ para limitar el número de posibles valores de G_p .

Con este criterio se han realizado unas simulaciones por ordenador para verificar el funcionamiento del algoritmo de adaptación propuesto. La Figura 103 muestra el retardo medio de los paquetes transmitidos en un sistema con $N=100$ usuarios que generan mensajes de longitud exponencial con llegadas de Poisson, donde los usuarios utilizan 4 velocidades de transmisión diferentes, para cada uno de los valores de G_p definidos en el anterior algoritmo.

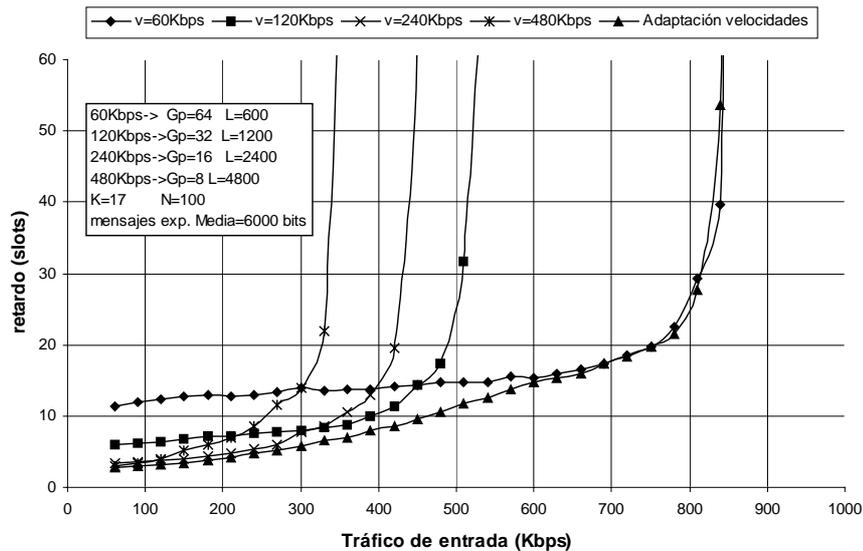


Figura 103. Retardo de paquetes con adaptación de velocidades

Se muestran cinco curvas. Las indicadas con una velocidad corresponden con escenarios donde todos los usuarios utilizan la misma velocidad en todas las transmisiones. La curva marcada con la etiqueta ‘Adaptación de velocidades’ corresponde al sistema en el que se ha incluido el algoritmo presentado en el punto anterior en el que cada paquete es transmitido con una velocidad que depende de la carga de tráfico en cada slot. El número de transmisiones simultáneas en todos los casos se ha mantenido constante igual a 17. Podemos observar como para cada una de las velocidades constantes el máximo de tráfico soportado alcanza los valores correspondientes a la máxima interferencia que es capaz de soportar cada usuario con la protección que le brinda la ganancia de procesamiento utilizada. Para cargas pequeñas se observa que el retardo mínimo de los paquetes también es inversamente proporcional a la velocidad utilizada en la transmisión.

Se comprueba como la técnica de adaptación de velocidades es capaz de seguir los valores mínimos de cada una de las velocidades, ya que va cambiando la ganancia de procesamiento en función de la carga del sistema. Por otro lado, en la Figura 104 se comprueba que el *throughput* que se obtiene es equivalente al caso óptimo para el que se eligió el valor de K . Por tanto, es una técnica que mejora sustancialmente el rendimiento en retardo del sistema manteniendo las propiedades respecto al *throughput*. A título informativo, la Figura 105 muestra cómo se distribuyen porcentualmente las diferentes velocidades de transmisión en función de la carga de tráfico. Se muestra con el valor $v=1$ la velocidad más lenta (60 Kbps), con $v=2$ la velocidad doble y así sucesivamente. Puede comprobarse que las velocidades altas de transmisión se utilizan en mayor medida cuanto menor es la carga de tráfico, mientras que cuando ésta aumenta, se usan las velocidades lentas, que tienen una mayor protección para las interferencias.

Por tanto, podemos concluir que la técnica de adaptación de velocidades propuesta se muestra como una estrategia eficiente en el aprovechamiento de los recursos del enlace radio en un sistema de comunicaciones que utilice el protocolo DQRAP/CDMA en el nivel MAC.

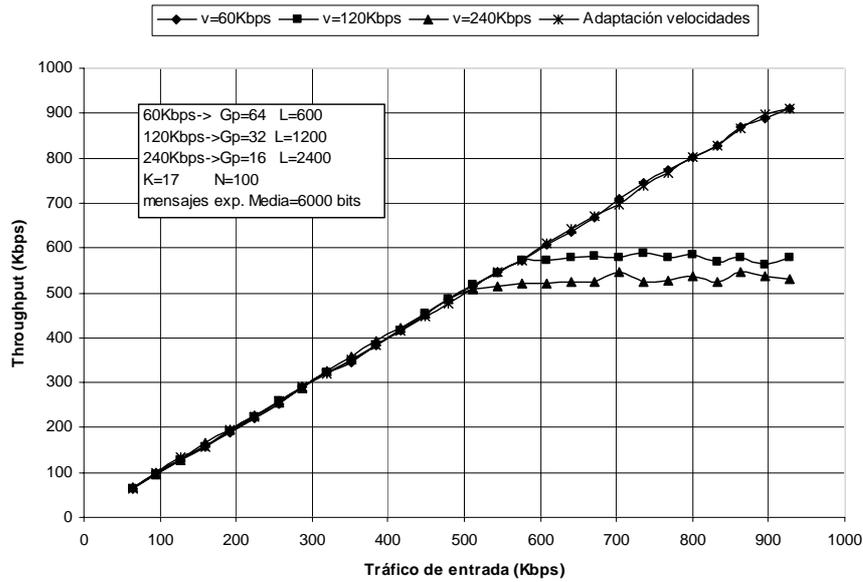


Figura 104. Throughput con adaptación de velocidades

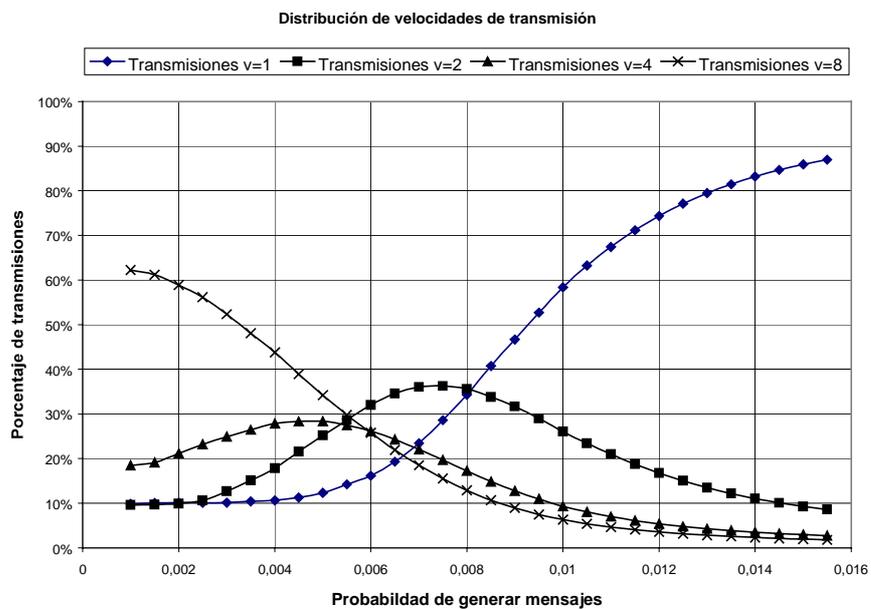


Figura 105. Distribución de velocidades de transmisión

4.9 Conclusiones

En este capítulo se ha realizado un estudio en profundidad del comportamiento de un sistema de transmisión basado en el protocolo DQRAP/CDMA de cara a encontrar los valores de los parámetros que optimizan el rendimiento para condiciones de tráfico genéricas. El análisis ha permitido encontrar estos valores de un modo razonado, a partir del estudio en detalle del funcionamiento del protocolo. Se ha mostrado como también puede utilizarse la información del protocolo para adaptar las velocidades de transmisión, variando las ganancias de procesamiento, de las comunicaciones, con lo que se obtiene un rendimiento global óptimo.

Finalmente, se ha analizado la respuesta del protocolo optimizado y se han mostrado los resultados obtenidos.

4.10 Anexo A

De cara a completar el análisis del rendimiento del protocolo optimizado, se muestran en este anexo algunos resultados complementarios de parámetros del sistema, obtenidos de las simulaciones. Resulta interesante realizar la comparativa de los resultados de las frecuencias relativas con las que se producen los diferentes tipos de error en la detección del estado de los minislots. Estos errores están definidos en el punto 3.10.2. Las gráficas desde la Figura 106 hasta la Figura 111 muestran las curvas correspondientes a los seis tipos de error posibles.

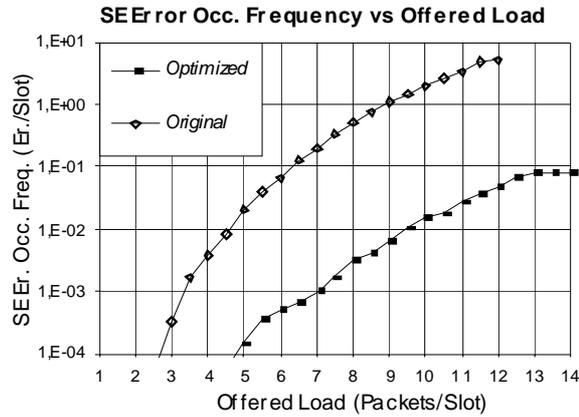


Figura 106. Comparativa de la frecuencia del error ($S/k=0$)

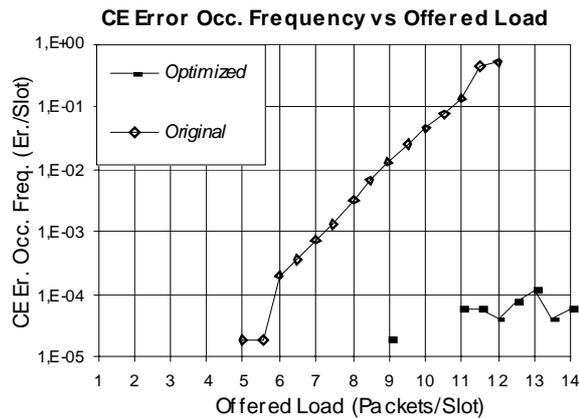


Figura 107. Comparativa de la frecuencia del error ($C/k=0$)

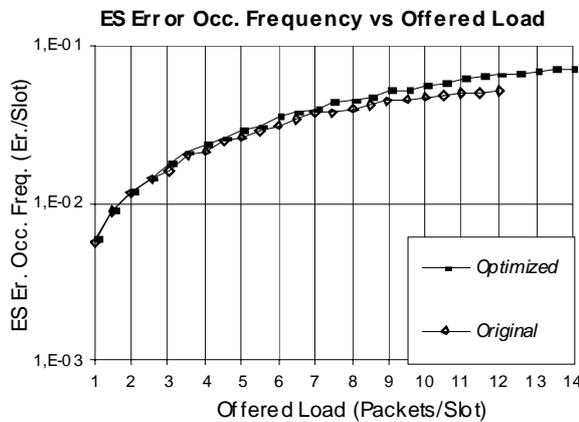


Figura 108. Comparativa de la frecuencia del error ($E/k=1$)

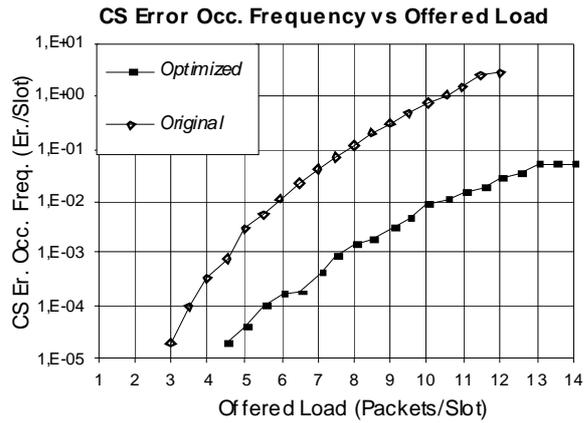


Figura 109. Comparativa de la frecuencia del error ($C/k=1$)

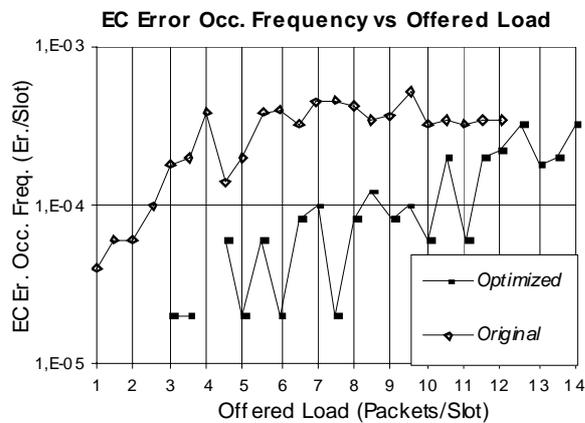


Figura 110. Comparativa de la frecuencia del error ($E/k>1$)

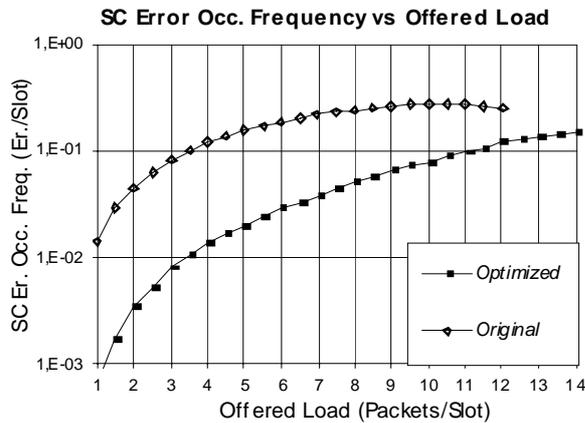


Figura 111. Comparativa de la frecuencia del error ($S/k>1$)

Para ilustrar el comportamiento de estas curvas, se muestran también los resultados obtenidos para otros parámetros complementarios. La Figura 112 muestra el porcentaje de errores del tipo ($S/k>1$) en los que tan sólo se vieron implicados 2 usuarios. Es decir, los éxitos detectados por error cuando más de dos usuarios han transmitido su secuencia de acceso. Este hecho se produce a causa de la no detección de alguna de las secuencias de acceso de alguno de ellos. Podemos observar que para el protocolo optimizado, este porcentaje se sitúa por encima del 99% para cualquier carga de tráfico ofrecida.

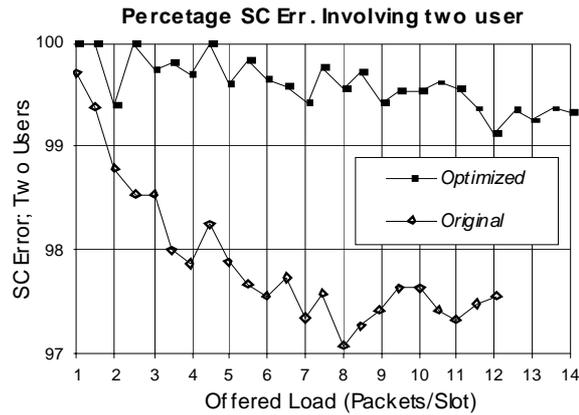


Figura 112. Porcentaje de errores ($S/k > 1$) en los que hay sólo 2 usuarios involucrados

Por otro lado, y para completar las gráficas complementarias, se muestran las frecuencias relativas por slot de la ocurrencia de los tres estados posibles *a priori* de los minislots de acceso. Es decir, cuántos slots están en principio vacíos (ningún usuario transmitió en ellos ninguna petición), cuántos tuvieron un solo acceso (éxito) y cuántos recibieron más de una petición (colisión), en función de la carga ofrecida. Podemos comprobar que las diferencias fundamentales entre las dos versiones del protocolo se encuentran en el número de minislots vacíos y en el número de colisiones.

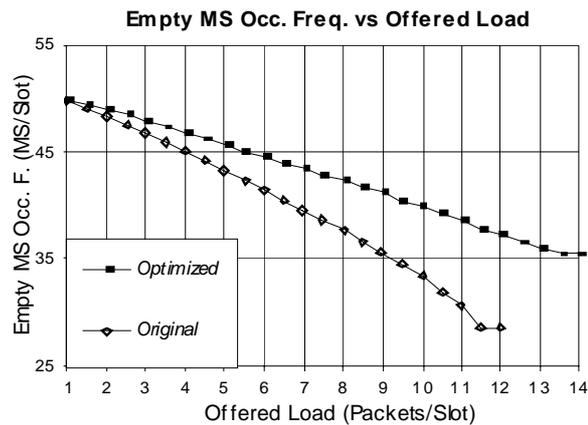


Figura 113. Número medio de minislots vacíos por slot

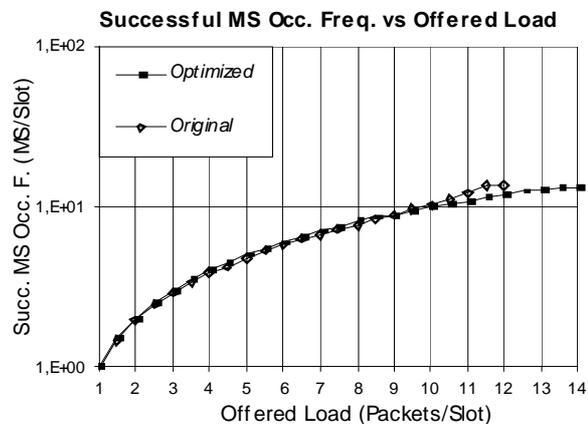


Figura 114. Número medio de minislots con éxito por slot

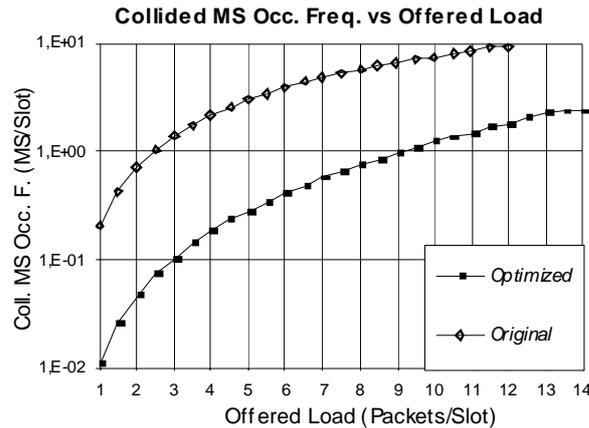


Figura 115. Número medio de minislots con colisión por slot

Seguidamente, y a la vista de todas las curvas presentadas, vamos a pormenorizar el análisis de cada una de ellas y del efecto que producen de los diferentes tipos de errores en la detección del estado de los minislots.

- *Error (S/k=0)*: Las modificaciones introducidas en la optimización del protocolo reducen la frecuencia con la que se produce este error en aproximadamente dos órdenes de magnitud. Este hecho es debido a la reducción del número de peticiones simultáneas que pueden generarse en cada slot, con la consiguiente descarga del subsistema de resolución de colisiones. Por tanto, la probabilidad de que la interferencia generada produzca una falsa alarma en el banco de filtros de detección de secuencias es muy pequeño.
- *Error (C/k=0)*: Este error es el que presenta la mejora más significativa con la optimización del protocolo. La menor carga del subsistema de resolución de colisiones permite disminuir la frecuencia con la que se produce este error en más de cuatro órdenes de magnitud.
- *Error (E/k=1)*: En una primera observación, este error parece tener un comportamiento contradictorio con lo esperado. La frecuencia con la que se produce es mayor en el protocolo optimizado que en la versión inicial. Recordemos, sin embargo, que la frecuencia de estos errores están definidas como sucesos condicionados al número de peticiones simultáneas que se dan en un cierto minislot. La menor carga ofrecida al subsistema de resolución de colisiones, con el consiguiente menor número de colisiones, hace que la probabilidad de que se produzca el suceso condicionante ($k=1$) es mucho mayor para el protocolo optimizado. Además, dado que el nivel de interferencia en el minislot es menor, resulta más probable no detectar las secuencias de acceso a que la interferencia pueda generar falsas alarmas.
- *Error (C/k=1)*: Este error es muy similar al error ($S/k=0$), puesto que en ambos casos se ha producido al menos una falsa alarma. En ambos casos la frecuencia de que se produzca el error se reduce en un factor 100 en el protocolo optimizado respecto al diseño inicial.
- *Error (E/k>1)*: De nuevo, la menor carga ofrecida al subsistema de resolución de colisiones hace que el suceso condicionante ($k>1$ en este caso) se produzca con mucha menos frecuencia en el protocolo optimizado. Este hecho, junto con el menor nivel de

interferencia global, hace que la frecuencia condicionada de este error sea menor en el diseño inicial del protocolo.

- *Error ($S/k > 1$):* Podemos aplicar de nuevo el mismo razonamiento para explicar el comportamiento de este error. La menor carga de peticiones sobre los minislots reduce el número de colisiones *a priori* e incrementa la probabilidad de que tan sólo se produzca la no detección de alguna de las secuencias de acceso, con el consiguiente aumento de la frecuencia con la que se produce este error.