

6 Capítulo 6. CALIDAD DE SERVICIO (QoS) GARANTIZADA

6.1 Introducción

Tal y como se ha reseñado en el capítulo 1, la llegada de las comunicaciones móviles de tercera generación, conocida bajo el concepto IMT-2000 dentro de ITU y como UMTS en el ámbito europeo, tiene como una de sus directrices principales la adopción de las denominadas técnicas de transmisión en modo paquete en sustitución de las actuales soluciones basadas en conmutación de circuitos. Algunos de los argumentos esgrimidos a favor de la incorporación de estos mecanismos de transmisión fueron detallados también en el capítulo 1.

Por otro lado, la demanda de servicios multimedia de características cada vez más heterogéneas está sufriendo un enorme auge, sobretodo debido a la fuerza que está adquiriendo Internet como red de comunicación global para todos los ámbitos de la economía y la sociedad. En el terreno del acceso móvil a Internet, las comunicaciones móviles de tercera generación representan un eje crucial para permitir que el punto de acceso a la red sea tan móvil como los posibles usuarios finales. Por tanto, el acceso radio de dichas comunicaciones debe soportar y dar servicio a un número indeterminado de aplicaciones multimedia con características de tráfico variadas y variables. Estas aplicaciones incluyen el comercio electrónico, las transacciones bancarias seguras, el vídeo bajo demanda, los sistemas de localización, la navegación por la red, la consulta de bases de datos, el envío de fotografías, música, etc. Todas estas aplicaciones ya están proyectadas en la actualidad, pero deben tenerse en cuenta también todas las posibles que puedan sugerirse en un futuro más o menos cercano.

En el capítulo 2 se vio que de cara a asegurar que todas las aplicaciones cumplan sus funciones y requerimientos deben arbitrarse los mecanismos adecuados para garantizar un cierta calidad de servicio (QoS) a cada una de las conexiones asociadas a cada aplicación. Así, para que el usuario final pueda recibir correctamente una secuencia de vídeo, es necesario que la tasa de transmisión esté por encima de un cierto umbral. Asimismo, la descarga de un cierto fichero de información importante puede requerir un tiempo máximo de transmisión. Evidentemente, los parámetros de esta calidad necesaria pueden ser tan variados como los son las mismas aplicaciones. Por ejemplo, dos conexiones, una para transferir mensajes de correo electrónico y otra para escuchar una secuencia de audio, no requerirán el mismo caudal ni la misma cota de retardo de transmisión. Es necesario en primer lugar, por tanto, definir correctamente los criterios de calidad que le vamos a exigir al sistema para cada conexión. Estos criterios pueden basarse en medidas objetivas o subjetivas sobre la percepción de calidad que tendrá el usuario final.

En este capítulo vamos a enfocar los aspectos descritos en el capítulo 2 desde el punto de vista concreto de un sistema de comunicaciones móviles de tercera generación.

A la hora de abordar el problema del diseño de un sistema que pueda ofrecer una cierta calidad de servicio garantizada, deben considerarse una serie de aspectos novedosos. En particular, aparecen algunas nuevas tareas a cumplir que no eran aplicables en los sistemas de comunicaciones móviles de segunda generación, o tan sólo podían considerarse superficialmente debido a su naturaleza orientada a circuito. Entre estos aspectos cabe destacar:

- *Protocolo de Acceso al Medio (MAC)*: Especifica la manera en la que los usuarios acceden al sistema para iniciar la transmisión de un conjunto de paquetes de datos. Aunque hasta la fecha se han estudiado un gran número y variedad de protocolos de este tipo, todos ellos tienen en común que contienen un cierto grado de aleatoriedad en el acceso. Cuando mayor es este grado de aleatoriedad, mayor es su flexibilidad, pero peor es su comportamiento cuando tratan de preservar unos ciertos requisitos de retardo máximo en el acceso. Por tanto, cuando una cierta calidad de servicio debe ser garantizada, el grado de aleatoriedad debe ser reducido en la medida de lo posible. Como veremos a continuación, el protocolo de acceso al medio DQRAP/CDMA (capítulo 3) presenta unas características intrínsecas muy adecuadas para su posible utilización como soporte en el acceso radio a este tipo de servicios de calidad garantizada.
- *Algoritmo de gestión de recursos (scheduling)*: Especifica el instante en el que un usuario que ya ha ganado acceso al sistema a través del MAC puede comenzar la transmisión de su información. También indica qué cantidad de recursos puede utilizar en esta transmisión. En sistemas CDMA esto incluye la ganancia de procesamiento y la potencia transmitida. Este mecanismo requiere de la definición de algún tipo de regla de priorización entre los usuarios así como de un algoritmo para distribuir los recursos entre ellos y así garantizar la calidad de servicio. En el capítulo 2 se describieron los principales algoritmos propuestos hasta la fecha para sistemas genéricos de comunicaciones.
- *Asignación dinámica de canales*: Este mecanismo es el responsable de decidir qué recursos (slots, secuencias de códigos, etc.) en concreto deben ser utilizados para realizar las transmisiones. Deberá trabajar en estrecha colaboración con el algoritmo de gestión de recursos (*scheduling*).
- *Control de admisión*: Es el encargado de decidir cuántos usuarios y de qué tipo pueden ser aceptados y en consecuencia pueden tratar de acceder al sistema. Este control de admisión es por tanto inherentemente dependiente del protocolo MAC y del algoritmo de gestión de recursos utilizado. Cuánto mejor trabajen estos dos mecanismos, mayor será el número de usuarios que podrán ser admitidos en el sistema. Los criterios de admisión de usuarios deben ser los adecuados para garantizar que todos los usuarios cumplan sus requerimientos de calidad de servicio. El número de usuarios máximo que pueden aceptarse en el sistema, teniendo en cuenta que puede haber usuarios con diferentes patrones de tráfico y requisitos de calidad, es lo que define la llamada *región de admisión*.

En el presente capítulo se pretende presentar y analizar una propuesta novedosa de algoritmo de gestión de los recursos (*scheduling*) radio de manera que, en un sistema basado en el acceso mediante el protocolo MAC DQRAP/CDMA, se puedan ofrecer diferentes servicios genéricos con calidades garantizadas variables. Se integrarán dentro del algoritmo tanto los mecanismos del MAC como los de gestión de recursos. Todo ello permitirá tener una plataforma en la que puedan definirse los criterios de control de admisión.

En cuanto a los servicios a tratar, se hará especial hincapié en los más usuales basados en el protocolo de Internet (IP), dada la importancia que están cobrando en la actualidad estas aplicaciones, aunque se tratará de definir una plataforma lo más genérica posible que pueda permitir albergar el máximo abanico posible de tipos de fuentes de tráfico y de servicio.

En primer lugar, vamos a definir algunos conceptos en los términos en los que serán usados de aquí en adelante como base de las técnicas que permiten garantizar calidad de servicio.

6.1.1 Clases de Servicio

Definimos Clase de Servicio (CS) como el conjunto de parámetros de calidad de transmisión que delimitan las características de un cierto flujo de información. En principio, cada una de las aplicaciones posibles (cuyo número es en principio indefinidamente grande), tendrá asociada una o más CS. Cada uno de los flujos de información generados y que deben ser transmitidos por la aplicación tendrá asignada una CS. Las conexiones asociadas a una CS generarán información siguiendo un cierto *patrón de tráfico*.

Definimos *patrón de tráfico* como la estadística con la que una cierta conexión genera paquetes de información. Los modelos de generación de tráfico se caracterizan por variables aleatorias y procesos estocásticos. Por tanto, el conocimiento completo de un cierto patrón de tráfico implica la definición de todos los momentos estadísticos de las variables aleatorias que definen el patrón.

Es evidente que el planteamiento de un sistema de comunicaciones que pueda dar cabida a la infinidad de patrones de tráfico posibles, cada uno de ellos con sus requerimientos de calidad, representa una tarea poco menos que inabordable. Es por ello que debe arbitrarse una solución viable y aplicable a un entorno real. Una solución posible consiste en la definición de un conjunto acotado de CS a las que deban acogerse todas las conexiones activas en el sistema y sus correspondientes aplicaciones. Este conjunto debe ser lo suficientemente amplio como para abarcar, de un modo suficientemente preciso, a la práctica totalidad de las conexiones que puedan requerir servicio del sistema, y a su vez lo suficientemente restringido como para simplificar en lo posible la implementación real de los mecanismos de gestión de recursos.

Por tanto, cuando una aplicación quiera ser servida por el sistema de transmisión, deberá analizar cada una de sus conexiones o flujos de información activos. Este análisis debe permitir decidir cuál de las CS definidas en el sistema resulta más adecuada a las necesidades de calidad de transmisión y se ajusta mejor al patrón de tráfico de cada conexión. Esta decisión deberá hacerse siempre basándose en un análisis conservador de los requerimientos, para asegurar así el cumplimiento de los requisitos necesarios para todas y cada una de las conexiones activas. En este sentido, el uso de conformadores de tráfico [50] para ‘suavizar’ el tráfico ofrecido al sistema es una técnica frecuentemente utilizada en los sistemas actuales.

De este modo, el sistema de comunicaciones verá las conexiones de todas las aplicaciones como un conjunto acotado y determinado de Clases de Servicio, cuyas características son conocidas a priori, lo que permite una planificación eficiente de la gestión del tráfico. Tanto el dimensionado de los accesos, enlaces y redes de comunicaciones, como la gestión de todo el funcionamiento del sistema para poder garantizar la calidad de servicio de las conexiones activas resultan realizables con un grado de complejidad abordable.

6.1.2 Parámetros de Calidad de Servicio (QoS)

De cara a la definición de las CS soportadas por el sistema, se hace necesario establecer los parámetros de transmisión que delimitan la calidad del servicio. Estos parámetros deberán tener una relación directa con la percepción que el usuario final (no necesariamente un ser humano) deba tener de la calidad de la conexión. Como usuario final se entiende cualquier nivel superior del sistema de comunicaciones que tenga unas necesidades de calidad

determinadas. Así por ejemplo, para el caso de una aplicación de transmisión de voz en tiempo real, deberán establecerse relaciones entre los parámetros medibles de la transmisión (retardo máximo de los paquetes, diferencia máxima entre retardos de paquetes, tasa máxima de paquetes perdidos, tasa máxima de errores en los bits de los paquetes, etc.) y la percepción subjetiva de inteligibilidad del habla.

Será por tanto el tipo de aplicación o usuario al que se deba dar servicio lo que condicionará el tipo y los valores de los parámetros que marcarán la definición de cada CS. Por cuestiones de simplicidad, en los trabajos de análisis de calidad de servicio realizados hemos considerado un conjunto acotado de parámetros de servicio, que se detallan a continuación:

- *Retardo medio de los paquetes de información:* se entiende por retardo de cada paquete el tiempo transcurrido desde que el bloque de información llega a la capa MAC hasta que es transmitido correctamente por la capa física.
- *Varianza del retardo de los paquetes de información:* el retardo de cada paquete, tal y como se ha definido en el punto anterior, es una variable aleatoria de la que podemos obtener su desviación típica y su varianza.
- *Retardo máximo de los paquetes de información:* se puede establecer un cierto tiempo de vida máximo de los paquetes, de tal modo que cuando el retardo de uno de ellos es superior a este tiempo de vida, el paquete es descartado.
- *Tasa máxima de paquetes perdidos:* porcentaje de paquetes descartados a causa de que su retardo ha superado el tiempo de vida prefijado para ellos.
- *Tasa de error media en los bits de información:* puede definirse antes o después de codificación, e indica el número relativo de bits erróneos (medido en porcentaje o como una probabilidad) que pueden admitirse por la aplicación.
- *Velocidad media de transmisión garantizada:* normalmente medida en Kbps, indica la velocidad media de transmisión para intervalos ‘largos’ de tiempo. Un intervalo largo se define como un número suficientemente grande de unidades de tiempo del sistema. Este número deberá ser grande en comparación con el tiempo en el que pueden variar las condiciones del tráfico ofrecido.
- *Velocidad mínima instantánea de transmisión:* también medida en Kbps, indica la velocidad mínima de transmisión de datos, si la hay, que se le garantiza a una conexión determinada. Si este valor es mayor que cero, indica que se está reservando una cierta cantidad de recursos de transmisión mínimos en exclusiva para la conexión, independientemente de la carga restante del sistema.
- *Velocidad máxima instantánea de transmisión:* indica la máxima velocidad de transferencia que le es permitida a una cierta conexión. Este valor puede usarse para impedir que una única conexión pueda copar una cantidad excesivamente grande de recursos del sistema, y evitar los problemas que de este hecho pudieran derivarse.

Normalmente, una CS estará definida por un subconjunto de estos parámetros, así como los valores correspondientes para cada uno de ellos.

6.2 DQRAP/CDMA con calidad garantizada

6.2.1 Características de DQRAP/CDMA

Algunas de las características del protocolo DQRAP/CDMA lo hacen especialmente indicado para ser utilizado como estrategia MAC para dar servicio a aplicaciones que generan tráfico heterogéneo multimedia y que requieren de una cierta calidad garantizada. Además, la estructura del protocolo permite integrar el algoritmo de gestión de recursos (*scheduling*) dentro del mismo protocolo. En efecto, el sistema de acceso basado en DQRAP/CDMA consiste en una serie de colas que gestionan tanto las solicitudes de recursos (posiciones de las colas distribuidas) como las transmisiones de los paquetes de datos. Este hecho hace que sea posible, en base a definir una gestión adecuada de estas colas, conseguir de forma directa establecer unos criterios de prioridad necesarios para ofrecer la garantía de calidad requerida.

Por todo ello, nos planteamos la definición de una propuesta novedosa de gestión de los recursos de transmisión que permita asegurar una cierta calidad de servicio en un sistema de comunicaciones móviles con transmisión en modo paquete que se base en el protocolo MAC DQRAP/CDMA. Esta propuesta presenta un enfoque distinto a los esquemas presentes en la literatura, en tanto que realiza una gestión distribuida en los terminales móviles. Esta estructura distribuida minimiza la cantidad de señalización necesaria en el sistema y mejora por tanto el rendimiento general del mismo. Es además un enfoque novedoso que plantea un punto de vista de diseño que presenta características muy interesantes en algunos entornos de trabajo.

6.2.2 Colas independientes para servicios distintos

DQRAP/CDMA es un protocolo de acceso al medio basado en dos sistemas de colas distribuidas, una para la resolución de las peticiones de transmisión y otra para la transmisión de los datos. Por tanto, la propia estructura del protocolo plantea la posibilidad de adaptarlo a un entorno en el que puedan soportarse distintos servicios de calidades y características diferentes.

La adaptación propuesta consiste en crear tantas colas distribuidas como tipos o clases de servicio, y asignar cada una de ellas a su correspondiente servicio.

La Figura 132 muestra el esquema del modelo teórico de funcionamiento del protocolo DQRAP/CDMA clásico, en el que se observan explícitamente los dos subsistemas de colas distribuidas en cascada que funcionan simultáneamente. Los terminales de usuario utilizan únicamente dos números enteros para llevar el registro de cada una de las colas.

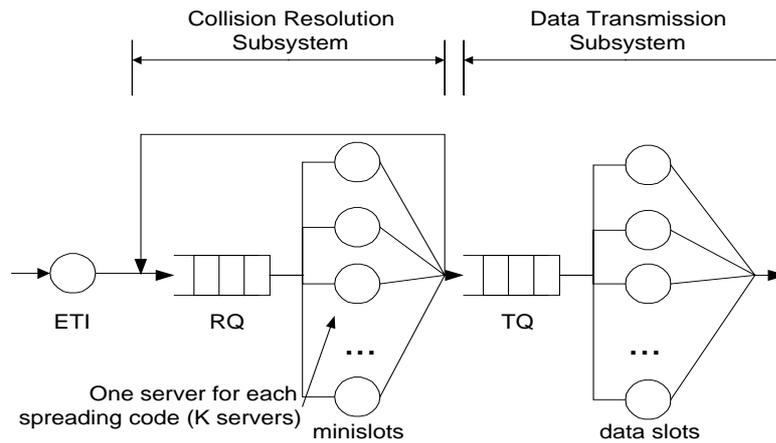


Figura 132. Modelo de DQRAP/CDMA clásico

En el caso de que las fuentes de tráfico de los usuarios tengan requerimientos de calidad diferentes, deberemos clasificar estas fuentes en clases de servicio (CS). Todas las fuentes que tengan los mismos requerimientos de calidad, o que puedan ajustarse a los parámetros de una cierta CS definida en el sistema, serán las que generarán el tráfico correspondiente a esa CS. En realidad, como ya se ha comentado con anterioridad, una clase vendrá definida por los parámetros de calidad que deban satisfacerse. Por otro lado, un mismo usuario físico puede tener varias conexiones lógicas pertenecientes a diferentes CS.

La Figura 133 muestra el esquema genérico propuesto para el caso de un sistema con acceso MAC del tipo DQRAP/CDMA, que ofrece servicio a un número indeterminado de fuentes de tráfico con distintas CS (cada una de ellas con sus requerimientos de QoS).

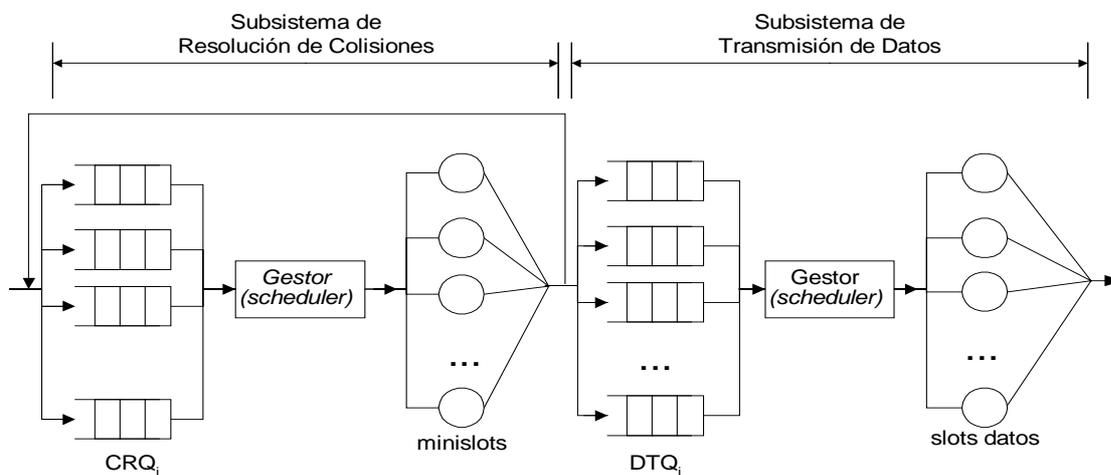


Figura 133. Modelo de DQRAP/CDMA con QoS garantizado

En primer lugar, las colas CRQ y DTQ del modelo clásico han sido sustituidas por sendas baterías de colas, una de ellas para la resolución de peticiones y otra para la transmisión de los datos. Denotaremos a cada una de las colas de cada batería respectivamente por CRQ_i y DTQ_i . Habrá tantas en cada grupo como diferentes clases de servicio tengamos en el sistema. Cada valor del subíndice i corresponde a cada uno de los servicios admitidos. Se mantiene la estructura de dos subsistemas en cascada trabajando de forma simultánea, pero en cada uno de ellos aparece un elemento nuevo: el gestor o *scheduler*. Este elemento es el encargado de

decidir cuál o cuáles de los elementos contenidos en las colas CRQ_i o DTQ_i (según el caso) deben recibir servicio y qué recursos serán asignados a dicho servicio. Cada uno de los servidores del modelo representa una cierta cantidad de recursos (potencia de transmisión, secuencias código) del sistema, que a su vez puede corresponder, en el eje de tiempos, a minislots de control o a slots de datos (según el subsistema). Por ejemplo, en el caso del subsistema de resolución de colisiones, cada servidor representa a cada uno de los receptores de la estación base que detecta la información enviada con una cierta secuencia código específica. La tarea del gestor será entonces la de asignar uno o más servidores (los recursos disponibles) a los elementos presentes en las colas, de tal modo que puedan cumplirse los requerimientos de tráfico de cada una de las clases.

En el caso del subsistema de transmisión de datos, debe reseñarse que el hecho de asignar de forma dinámica los recursos de transmisión a las diferentes posiciones de las colas DTQ_i comporta que el sistema realiza una adaptación de las tasas de transmisión de cada una de las conexiones activas. En efecto, puesto que cada uno de los servidores del modelo representa una de las secuencias códigos del sistema, el hecho de asignar más de uno de estos servidores a una cierta conexión permite aumentar su tasa efectiva de transmisión. Este mecanismo realiza la variación de la velocidad de transmisión haciendo uso de la técnica llamada de multicódigo (MC-CDMA) [51]. Esta adaptación de velocidades es la que permitirá que el sistema pueda garantizar una cierta calidad de servicio a las conexiones.

El gestor del subsistema de resolución de colisiones y el del subsistema de transmisión funcionan de modo independiente.

A la hora de implementar un sistema como el propuesto basado en DQRAP/CDMA, debe tenerse en cuenta que al desdoblarse las colas distribuidas (una por cada CS), deben realizarse algunos cambios a la hora de ejecutar los tres conjuntos de reglas del algoritmo del protocolo MAC. En efecto, cada una de las colas CRQ_i y DTQ_i actúan de modo independiente, haciendo uso de las reglas del protocolo DQRAP/CDMA descritas en el punto 3.4. Sin embargo, debe tenerse en cuenta que cada uno de los terminales deberá ejecutar dichas reglas para todo el conjunto de colas de todas las clases de servicio. Esto es necesario para mantener el conocimiento global del estado del sistema. Por otro lado, y para evitar problemas en el proceso de funcionamiento, es necesario desactivar la regla DTR-1 en todos los terminales, es decir, inhibir los accesos libres a la transmisión de datos.

A continuación se va a analizar uno de los conceptos básicos que permite introducir el uso de DQRAP/CDMA en un sistema con calidad garantizada y que puede mejorar el rendimiento respecto de otros esquemas de gestión de recursos.

6.2.3 Gestión de los recursos (scheduling) distribuida

Uno de los objetivos de cualquier protocolo de acceso al medio es minimizar la cantidad de información de señalización que debe circular entre la estación base y los terminales móviles y, al mismo tiempo, maximizar la eficiencia en la utilización de los recursos radio en el enlace ascendente.

La estructura de DQRAP/CDMA permite plantear el esquema de señalización de un modo totalmente diferente al clásico. En los sistemas convencionales, la estación base mantiene un registro de todas las peticiones de los distintos usuarios, realiza el algoritmo de gestión de recursos definido en el sistema y va otorgando los correspondientes permisos para transmitir en el momento y con el formato adecuados a cada usuario de un modo explícito. En cada

confirmación de asignación de recursos deberá indicar con detalle cuáles son estos recursos, generando el mensaje de control correspondiente. Por tanto, si se desea que esta asignación sea dinámica según van llegando las peticiones de acceso de cada clase de servicio, la implementación puede resultar muy compleja y conllevar un tráfico de control de un volumen elevado.

Haciendo uso de DQRAP/CDMA podemos introducir un nuevo concepto: la gestión de recursos distribuida o *scheduling* distribuido. El funcionamiento del protocolo requiere que la estación base envíe a los usuarios por el canal descendente el resultado de lo ocurrido en todos los minislots de acceso. Con esta información, los terminales conocen el estado global del sistema y su posición en las colas distribuidas. De este modo, si los usuarios conocen las reglas de gestión de recursos definidas en el sistema, podrán ejecutar de modo local el algoritmo de asignación de recursos. Así, serán capaces de calcular cuándo deben transmitir y con qué recursos deben hacerlo, sin necesidad de que la estación base se lo indique explícitamente con un mensaje de señalización. Este hecho disminuye la cantidad de tráfico de control necesario en el sistema.

Del mismo modo, en el caso de un sistema con varias colas en paralelo, si los móviles conocen el estado de las peticiones de acceso de cada una de las clases de servicio, también podrán aplicar la misma filosofía de funcionamiento. Esto es, cada uno de los móviles ejecuta localmente el algoritmo de gestión de recursos definido para el sistema, actualiza el estado de su copia local de las colas, y decide si puede transmitir o no, y en su caso con qué recursos. Son por tanto los móviles de forma distribuida los que deciden cuándo y cómo realizar la transmisión de los datos, eliminando la necesidad de enviar mensajes explícitos de señalización para este cometido. En cierto modo, los dos gestores mostrados en la Figura 133 están lógicamente distribuidos entre todos los usuarios presentes en el sistema.

Sin embargo, para que los terminales puedan realizar el *scheduling* distribuido es necesario que dispongan de una información adicional que en principio no pueden obtener de la capa MAC. El simple estado de los minislots de acceso correspondientes a cada clase de servicio no es suficiente para realizar un esquema de gestión de los recursos que pueda garantizar algún tipo de calidad de servicio. Los usuarios necesitan saber cuál es el tiempo transcurrido desde que un mensaje es generado por otro usuario hasta que este mensaje llega al sistema de acceso (la capa MAC) por primera vez. Este valor será necesario sea cual sea la técnica de gestión empleada para poder evaluar las prioridades y realizar el reparto de recursos. El hecho de que la calidad del servicio que debe garantizarse esté definida con un tipo concreto de parámetros (de los definidos en el punto 6.1.2) no afecta a la necesidad de disponer de esta información.

Es posible arbitrar algunas soluciones para abordar esta necesidad del sistema. A continuación se detallan tres propuestas concretas:

- Hacer que los minislots de acceso transporten información explícita, en la que se incluya un valor numérico que indique el tiempo de vida (o retardo acumulado) del paquete que está accediendo, desde el momento de su generación (desde que llegó al sistema desde capas superiores) hasta que inició la solicitud de recursos de transmisión en la capa MAC. De este modo, los minislots de acceso dejan de ser una simple secuencia de chips sin contenido explícito de datos.
- Hacer que los códigos utilizados en los minislots de acceso estén ordenados y clasificados de forma que, cuando un terminal deba transmitir una petición de acceso haciendo uso de

uno de ellos, no escoja de forma aleatoria entre los códigos disponibles, sino que lo haga seleccionando uno de entre un cierto subconjunto. Este subconjunto corresponderá a los códigos asignados al retardo o tiempo de vida asociado al mensaje que tiene que transmitir. La recepción de una cierta secuencia de chips llevará por tanto de forma implícita la información del retardo previo al acceso al sistema del paquete de datos correspondiente.

- Estimar el retardo acumulado de cada paquete a partir del conocimiento de la estadística con la que generan sus paquetes todas las conexiones, es decir, a partir del conocimiento a priori del patrón de tráfico de cada conexión (ver 6.1.1). Todos los usuarios podrán estimar el retardo de los paquetes generados por el resto de usuarios calculándolos respecto a los instantes de referencia en los que ‘deberían’ haberse generado los paquetes de cada conexión. Estas estimaciones serán tanto más precisas cuanto más ‘disciplinadas’ sean las conexiones activas, esto es, cuanto más fielmente sigan su patrón de tráfico definido.

La primera de las soluciones es la que permite conocer con más exactitud el retardo de los paquetes en cada momento, pero al mismo tiempo elimina la característica principal de los minislots de acceso utilizados en DQRAP/CDMA: su nulo contenido de información explícita. Esta característica es la que permite reducirlos a un tamaño mínimo, con la consiguiente mejora en la eficiencia del sistema.

Por otro lado, la segunda de las alternativas presenta un grado de complejidad mayor en la estructura del receptor, pero al mismo tiempo mantiene la simplicidad de los minislots de acceso. La filosofía subyacente en esta propuesta es la de tratar de que en el sistema global esté definida la mayor cantidad de información de forma inherente al mismo. Es decir, cuanto más información esté preestablecida entre la estación base y los móviles, menos señalización será necesaria intercambiar entre ellos. Si hacemos que el mero hecho de escoger una secuencia para transmitir un petición de acceso lleve de forma implícita cierta información, estamos eliminando la necesidad de transmitirla explícitamente. Sin embargo, para que el sistema sea factible, deberemos establecer la definición de los conjuntos de secuencias código de modo que cada uno de ellos corresponda a un cierto intervalo de retardos o tiempos de vida. Sería inviable requerir la existencia de una estación base con tantos receptores como valores diferentes del retardo pueden darse. Esto conduce a una cierta pérdida de precisión en el conocimiento del retardo exacto de cada mensaje que hay en el sistema.

La tercera de las alternativas se basa en el conocimiento a priori de la estadística del tráfico ofrecido al sistema. Si todas las conexiones siguen un patrón de tráfico conocido (conocemos la estadística con la que generan su información), es posible mejorar sustancialmente la mecánica de la gestión de los recursos. En particular, se hace posible estimar el retardo acumulado de todos los paquetes de información sin necesidad de indicar su valor de forma explícita en un mensaje de control. Sin embargo, de nuevo tenemos un compromiso entre complejidad y precisión. Necesitamos estimar el retardo acumulado desde que un paquete es generado hasta que se inicia la petición de recursos para su transmisión. En tanto en cuanto simplifiquemos la manera de realizar esta estimación, perderemos precisión en la misma. Este hecho resta necesariamente eficiencia a la gestión global del sistema.

- Por su equilibrio entre sencillez de implementación y eficiencia, de cara a la implementación del sistema propuesto, se ha escogido el uso de la segunda de las opciones presentadas. En lo sucesivo, por tanto, haremos uso de ella estableciendo conjuntos de códigos que se corresponden con diferentes valores del retardo de cada

paquete. En consecuencia, se realiza la suposición de que el número de conjuntos de códigos es lo suficientemente grande como para conocer con la precisión adecuada el retardo acumulado de los paquetes antes de llegar al sistema.

6.2.4 Gestión de los recursos para el subsistema de transmisión

En este punto vamos a describir el mecanismo de gestión de recursos propuesto que se aplicará a la batería de colas DTQ.

El esquema propuesto de sistema que garantiza una cierta calidad es genérico. El modelo presentado en la Figura 133 no presupone la existencia de ninguna estructura de trama en tiempo, ni de ningún tipo de acceso en particular. Basado en este modelo, es perfectamente posible utilizar cualquier propuesta de gestión de recursos con cualquier mecanismo de MAC o estructura de canal físico.

En la literatura podemos encontrar algunas propuestas de algoritmos de gestión de los recursos para la transmisión de paquetes con calidad garantizada [52], [53]. Algunas de ellas han sido descritas en el capítulo 2 a modo de antecedentes. Sin embargo, todas estas propuestas fueron diseñadas para entornos de transmisión con acceso TDMA. Resulta difícil encontrar propuestas que estudien el caso de un sistema con acceso CDMA y transmisión en modo paquete. La propuesta más cercana la encontramos en [54], donde se estudia un sistema mixto CDMA/TDMA en el que se establece una cierta estructura de trama en tiempo y se procura que los paquetes que tienen requerimientos de calidad iguales se transmitan de forma simultánea.

Nos planteamos por tanto la propuesta de un método nuevo que permita ofrecer calidad de servicio en una transmisión de datos basada en acceso CDMA y en el protocolo MAC DQRAP/CDMA. La propuesta elaborada es la que se detalla a continuación.

6.2.4.1 Un nuevo concepto: Crédito de Servicio

Podemos encontrar en la literatura algunas propuestas de gestión de recursos que se apoyan en un concepto conocido como Servicio Pendiente [55]. Todas ellas están basadas en un concepto usado en sistemas de red fija conocido como *'leaky bucket'* o *'cubo agujereado'* (ver 2.4.2). Este mecanismo consiste en asignarle a cada conexión unos testigos o *tokens* de forma regular en el tiempo. Estos *tokens* son unidades enteras abstractas que representan una cierta capacidad de transmisión pendiente de ser servida. La tasa de generación de estos testigos para cada conexión depende por tanto de su tasa de transmisión pactada en el momento de entrada en el sistema. A medida que cada conexión va recibiendo servicio del sistema, se le van retirando *tokens* proporcionalmente al servicio recibido. El número de testigos acumulados por una conexión mide la *'deuda'* de servicio que tiene el sistema con ella. Por otro lado, una conexión sólo podrá recibir servicio cuando tenga testigos disponibles, es decir, cuando el sistema le *'deba'* servicio. Analicemos en más detalle la filosofía del mecanismo.

Cada una de las conexiones a las que se debe prestar servicio con una cierta calidad garantizada debe solicitar una cierta calidad de servicio del sistema. Esta calidad, establecida en la negociación inicial entre el sistema y el usuario, estará definida con algunos de los parámetros detallados en el punto 6.1.2. Esta negociación se produce cuando los usuarios acceden por primera vez al sistema, o bien cuando estaban inactivos y reanudan una conexión. En ese momento, y según los parámetros solicitados por el usuario, el sistema debe decidir si puede admitir o no la nueva conexión. Para admitir una nueva conexión deberá asegurarse de

que podrá cumplir sus requerimientos de calidad así como no degradar la calidad de las conexiones que ya se están sirviendo de los usuarios que están en el sistema. Esta decisión es lo que hemos llamado control de admisión.

Supongamos que a una cierta conexión se le asegura una tasa media de transmisión de 0'1 paquetes de datos por cada unidad de tiempo del sistema. En este caso, y haciendo uso del mecanismo de testigos, cada 10 unidades de tiempo el sistema le asignará a esa conexión un testigo o *token*. Cada vez que la conexión reciba recursos de transmisión, es decir, cada vez que un paquete de esa conexión sea transmitido con éxito, se 'gastará' uno de estos testigos. Por tanto, dichos testigos se irán acumulando si dicha conexión no recibe servicio del sistema. En el caso de que durante un cierto período de tiempo el número de paquetes transmitidos de esa conexión sea tal que 'agote' los testigos disponibles, esa conexión quedará inhibida (desde el punto de vista de transmisión) y no ocupará más recursos del sistema, que podrán ser usados por otras conexiones.

La nueva propuesta realizada de gestión de recursos consiste en extender el concepto de los testigos a un entorno como el estudiado hasta este momento (sistema CDMA con transmisión en modo paquete). Para ello, redefinimos los testigos y utilizamos en su lugar un nuevo concepto: el Crédito de Servicio (de ahora en adelante CrS).

Para la definición y el cálculo del CrS es necesario establecer una serie de conceptos que serán utilizados de aquí en adelante.

- *Usuario*: Llamaremos usuario a la entidad abstracta que genera información y necesita transmitirla por el canal de comunicaciones a través del sistema objeto de diseño. Representa todas las capas superiores (por encima de la capa física, el MAC y la gestión de recursos) del sistema.
- *Conexión*: Llamaremos conexión a cada uno de los flujos de información que genera un usuario. Los usuarios generan diferentes flujos de información, provenientes de distintas aplicaciones con necesidades de calidad variadas. Diremos que una conexión está activa cuando es susceptible de generar información que deba ser transmitida por el sistema. Las transmisiones del sistema se realizan en modo paquete. Por tanto, una conexión puede permanecer activa incluso mientras no está generando o transmitiendo información. El tiempo durante el cual una conexión está activa suele llamarse sesión. Se dice que una conexión activa tiene una sesión *abierta*.
- *Tasa básica de transmisión de una conexión*: Velocidad mínima (en bits/s) con la que se transmiten los paquetes de datos de una conexión por el canal de comunicaciones. Esta velocidad se refiere a la tasa con la que se envían los bits de los paquetes de la conexión cuando se realiza la transmisión en la capa física.
- *Tasa mínima garantizada para una conexión*: Velocidad mínima (en bits/s) de transmisión que el sistema garantiza a una conexión, promediando los períodos de actividad con los de inactividad de la misma. Es decir, esta velocidad se refiere a la tasa de transmisión *media* mínima que una conexión tiene garantizada. Todas las conexiones que se vuelven activas en el sistema *deben* tener una tasa mínima garantizada.
- *Servicio recibido*: Diremos que una conexión ha recibido servicio del sistema cuando se realizado con éxito la transmisión de un paquete de información de dicha conexión.

6.2.4.1.1 Definición de CrS

El CrS es un número real, no necesariamente entero, asociado a una conexión activa del sistema. No tiene unidades físicas. Mide la diferencia entre el servicio que una conexión ha contratado con el sistema (definido por la tasa mínima garantizada de dicha conexión) y el servicio que ha recibido. Su valor numérico se calcula según las reglas que se detallan en el punto 6.2.4.1.2. Su valor se utiliza para comparar el servicio pendiente de diferentes conexiones y por tanto el valor en sí tiene únicamente un sentido relativo al que tienen asociado las demás conexiones activas del sistema. Cuanto mayor es su valor, más prioridad tendrá la conexión asociada en la asignación de recursos de transmisión. Se puede establecer por consiguiente una correspondencia directa entre CrS y prioridad.

Cada una de las conexiones activas en el sistema tendrá su propio valor de CrS, independiente del que tienen el resto de las conexiones. El valor del CrS de cada conexión deberá actualizarse cada cierto tiempo según las reglas que se detallan en el siguiente punto.

6.2.4.1.2 Mecanismo de cálculo del CrS

El valor del CrS de cada conexión debe actualizarse periódicamente en el tiempo. Esta periodicidad depende de la unidad de tiempo de transmisión del sistema. Según los sistemas, este intervalo se corresponderá a un slot, una trama, una multitrama o cualquier unidad de tiempo. En el caso de un sistema basado en el protocolo DQRAP/CDMA el tiempo de actualización es el slot. Cada vez que deba actualizarse el valor del CrS debe seguirse la siguiente expresión:

$$\text{CrS}_{\text{nuevo}} = \text{CrS}_{\text{anterior}} + (t_{\text{mín_garantizada}}/t_{\text{básica}}) - \text{num_paq_trans_con_éxito} \quad (6.1)$$

Donde $t_{\text{mín_garantizada}}$ es la tasa mínima garantizada para la conexión, $t_{\text{básica}}$ es la tasa básica de transmisión de la conexión y $\text{num_paq_trans_con_éxito}$ es el número de paquetes de datos de la conexión transmitidos con éxito durante el último intervalo de tiempo del sistema.

Es decir, por un lado el valor del CrS debe aumentarse en una cantidad igual al cociente entre la tasa mínima garantizada para la conexión y su tasa básica de transmisión. Este valor es una medida proporcional al servicio que debería haber recibido la conexión asociada en el período de actualización. Por ejemplo, si una conexión tiene una tasa mínima garantizada de 1 Kbps y transmite su información en paquetes de 500 bits en slots de 10 ms (tasa básica de transmisión de 5 Kbps), el valor de CrS deberá aumentarse por tanto en $1\text{Kbps}/5\text{Kbps}=0,2$ unidades cada 10 ms.

Por otro lado, cada vez que la conexión recibe servicio del sistema, es decir, cuando un paquete de dicha conexión es transmitido con éxito por la capa física, su valor de CrS se decrementa en una unidad. En el ejemplo, cada 5 slots de tiempo el valor acumulado del CrS aumenta en una unidad. Si en ese período se transmite un paquete de datos, la suma total del CrS será nula, lo que indica que el servicio ha cumplido sus requisitos de calidad, es decir, ha transmitido a su tasa mínima garantizada.

De este modo, e independientemente de la estadística con la que se genera la información de cada conexión, se tiene en todo momento una medida de la 'deuda' del sistema con la conexión. El valor obtenido para el CrS está normalizado a la tasa básica de transmisión de cada conexión, de modo que permite comparar la 'deudas' de conexiones distintas (diferentes CS). Valores positivos de CrS suponen que la conexión ha recibido hasta ese momento menos

servicio del sistema del que pactó. Cuanto mayor sea CrS, tanto más servicio debe darle el sistema a la conexión para ponerse ‘al día’. Por el contrario, un valor negativo de CrS supone que el sistema le ha dado más servicio a la conexión del estrictamente acordado, lo que puede ser perfectamente posible para cargas bajas de tráfico.

A la hora de asignar recursos a las conexiones (*scheduling*), se deberán ordenar éstas según su valor CrS. Se dará prioridad en la asignación a aquellas que tengan un CrS mayor. Los valores negativos de CrS suponen caso especial que debe abordarse en particular. En esta situación caben dos posibilidades:

- Estrategia *Work-conserving*: Mientras haya recursos disponibles, se da servicio a todas las conexiones activas posibles, incluso aunque su CrS sea negativo.
- Estrategia *Non-work-conserving*: Aunque existan recursos disponibles para la transmisión, no se da servicio a aquellas conexiones que tienen un CrS negativo. Esta estrategia, que podría parecer ineficiente en cuanto a aprovechamiento de los recursos, es especialmente adecuada para sistemas con acceso CDMA, en los que una asignación conservadora de los recursos permite reducir el nivel de interferencias del sistema y mejorar las prestaciones de las transmisiones en curso.

Existen, sin embargo, algunas situaciones que deben tratarse de un modo especial cuando se hace uso de una estrategia de *scheduling* basada en el CrS:

Si por alguna razón una conexión permanece inactiva durante un período largo de tiempo, ésta puede acumular un valor muy grande de CrS. En el momento en el que la conexión vuelva a la actividad, su valor anormalmente grande de CrS podría llevar al sistema a asignarle una cantidad demasiado elevada de recursos de transmisión y dejar a las demás conexiones en estado de ‘inanición’. Para solventar este tipo de situaciones existen al menos dos estrategias posibles:

Establecer un valor máximo de CrS admisible para cualquier conexión. Este límite puede ser dependiente del tipo de servicio ofrecido, e incluso puede variar con el tiempo. En el caso de que una conexión alcance este valor máximo, ya no se le incrementa más su CrS. Este límite define de algún modo la máxima prioridad que se puede otorgar a una cierta conexión. Al mismo tiempo, también determina la tolerancia del sistema a la variación en el tiempo de la tasa de generación de información de cada conexión. Si el límite es un valor pequeño, las conexiones que generen tráfico muy racheado pueden verse perjudicadas. Esto es debido a que su valor de CrS alcanzará repetidas veces este límite y por tanto estas conexiones perderán con frecuencia parte del crédito que les corresponde. Sin embargo, con un valor excesivamente grande del mismo estaríamos permitiendo que las conexiones racheadas tuvieran muy alta prioridad en la asignación de recursos al volver a la actividad tras un período de reposo, lo que podría perjudicar al resto de las conexiones.

Establecer un máximo de recursos asignables por el sistema a una conexión. De este modo, aunque una conexión tenga un valor inadecuadamente grande de CrS, nunca podrá dejar sin recursos a las demás conexiones a causa de que el sistema le asigne todos los recursos disponibles. En consecuencia, las conexiones que han estado durante un largo tiempo inactivas recuperarán su servicio pendiente de un modo más progresivo y sin causar perjuicio al resto de conexiones activas.

Es importante reseñar que para sistemas de transmisión por radio, especialmente en entornos móviles, sólo se debe reducir el valor del CrS cuando se ha recibido con éxito la notificación de que se ha servido correctamente a una conexión. Es decir, no debe actualizarse el valor del CrS sencillamente cuando se realiza la transmisión de un paquete de datos. Esto es así ya que en muchas ocasiones, y debido a la variabilidad e imprevisibilidad de la respuesta del canal radio y de las interferencias, la probabilidad de que un paquete de datos no llegue correctamente al destino es relativamente elevada.

Para poder realizar esta verificación es necesario que exista una cierta comunicación entre la capa encargada de la gestión de recursos y las capas inferiores (MAC, física). Esta comunicación es especialmente necesaria en entornos de transmisión por canal móvil.

6.2.4.2 Propuesta de mecanismo de gestión de recursos

Una vez definido el concepto de CrS y su método de cálculo, pasamos a detallar de qué manera el sistema asigna recursos de transmisión a las diferentes conexiones en función de su CrS. Por tanto, el objetivo del mecanismo de gestión es el reparto adecuado de los *recursos* disponibles de cara a conseguir un sistema con calidad de servicio garantizada.

- Llamaremos *recurso de transmisión* a cada uno de los códigos CDMA con los que cada conexión podrá realizar sus transmisiones a la velocidad correspondiente a su tasa básica de transmisión.

Se desea que el sistema permita la adaptación de velocidades de transmisión de cada una de las conexiones activas. Esto es necesario ya que en un momento dado, es posible que resulte conveniente que una conexión transmita su información a una tasa mayor que la correspondiente a su tasa básica. Ello será posible siempre que la calidad de las conexiones en curso no se vea degradada. En el sistema propuesto, la variación de la tasa de transmisión de las conexiones se realiza mediante técnica de transmisión multicódigo. Es decir, si a una conexión se le asignan dos códigos de transmisión podrá transmitir su información al doble de velocidad, usando ambos códigos al mismo tiempo. Si se le asignan tres códigos transmitirá al triple de velocidad y así sucesivamente. Esta técnica es totalmente equivalente a utilizar una variación de la ganancia de procesado en un sistema DS-CDMA. De este modo, la velocidad de transmisión de una conexión será siempre un múltiplo de su tasa básica de transmisión.

Pasamos ahora a describir el algoritmo de asignación de recursos basado en los valores de CrS de cada conexión. El objetivo final de este algoritmo es crear una tabla de asignación en la que se concretan cuántos códigos de transmisión puede utilizar cada una de las conexiones activas para transmitir sus paquetes en el slot inmediatamente posterior a la ejecución del algoritmo. Por tanto, cada vez que se indica que se realiza una asignación de recursos, quiere decir que se está asignando un código de transmisión a una determinada conexión. El resultado final del algoritmo es una tabla con la lista de conexiones activas y el número de códigos asignados a cada conexión.

Algoritmo de asignación de recursos: El algoritmo consiste en una serie de pasos que deben realizarse en cada uno de los ciclos de asignación de recursos del sistema. Estos ciclos pueden corresponder a slots, tramas, o cualquier intervalo de tiempo de transmisión del sistema. El valor del CrS de cada conexión debe actualizarse una vez en cada ciclo, siguiendo la expresión descrita en 6.2.4.1.2. En DQRAP/CDMA este ciclo lo constituye un slot de transmisión. Los pasos del algoritmo deben ser seguidos por todos los usuarios (gestión distribuida) antes de comenzar cada ciclo de asignación de recursos. Supondremos que el

tiempo de proceso es nulo, es decir, que todo el algoritmo se ejecuta en un tiempo despreciable situado entre el final de un intervalo de transmisión (en DQRAP/CDMA se corresponde con el slot de datos) y el inicio del siguiente. Todas las conexiones activas deben pertenecer a una de las CS del sistema y deben tener definidos dos parámetros que serán usados en el proceso, que son:

- *Tasa mínima de transmisión garantizada*: Se corresponde directamente con la tasa mínima definida en 6.2.4.1.
- *Nivel de interferencia máxima tolerada*: Este valor viene condicionado por el máximo BER admitido por la conexión.

El algoritmo de asignación de recursos se ilustra en la Figura 136. Los pasos a seguir en el mismo son los siguientes:

1. Se crea una tabla de asignaciones en la que se listan todas las conexiones activas, cada una de ellas con el número de códigos que tiene asignados. Inicialmente todas las conexiones tienen 0 códigos asignados. El proceso del algoritmo modificará los valores de esta tabla. Al terminar el proceso el resultado final será el contenido de la misma. Las conexiones activas realizarán las transmisiones de sus paquetes siguiendo la asignación de códigos que indique la tabla.
2. Se calculan los valores de CrS de todas las conexiones activas, siguiendo la regla de cálculo detallada en 6.2.4.1.2.
3. Se crea una copia de la tabla de valores de CrS que será utilizada para el resto del proceso de asignación. El objetivo es salvar los valores originales para mantenerlos intactos y realizar los cálculos del algoritmo sobre esta tabla temporal.
4. Se ordenan las conexiones activas según su CrS, de mayor a menor, respetando la prioridad que puedan tener las CS a las que pertenecen.
5. Se selecciona la conexión que ha quedado en primer lugar en la ordenación anterior y se realiza una asignación provisional de recursos a la misma. Esto es, en la tabla de asignaciones de salida (punto 1) se asigna un código de transmisión a la conexión seleccionada. Esta asignación se ejecuta sumando una unidad al valor correspondiente en la tabla. Como veremos más adelante, es posible que no sea la primera vez que se ejecuta este paso del algoritmo, y por tanto puede darse el caso de que no sea el primer código asignado a esta conexión. En esta situación, el valor en la tabla podrá resultar ser mayor que uno y por consiguiente la tasa de transmisión de la conexión será un múltiplo de su tasa básica de transmisión (adaptación de velocidades con la técnica multicódigo).
6. Se calcula la interferencia total que se generaría en el sistema con la asignación de recursos provisional realizada hasta este momento. Se compara este nivel de interferencia con el que pueden admitir todas las conexiones con recursos asignados. En caso de que la interferencia total esté por debajo de *todas* las cotas definidas para todas las conexiones (se cumplen los requisitos de calidad de todas ellas), se guarda la tabla de asignación de recursos realizada hasta este punto y se registra como la última asignación provisional válida. En caso negativo (la interferencia total está por encima de *alguna* de las cotas de las conexiones) se recupera la última asignación registrada

como válida y se finaliza el proceso de asignación de recursos. Esta tabla de asignación será la definitiva que se obtiene como resultado del algoritmo de gestión de recursos. Por tanto, será la asignación con la que se realizarán las transmisiones en el siguiente slot.

7. Si el proceso de asignación no ha terminado en el punto anterior, se recalcula el valor de CrS de la conexión seleccionada en el punto 5. A estos efectos se supone que ha realizado una transmisión con éxito y se resta una unidad a su valor de CrS.
8. Se vuelve al punto 4.

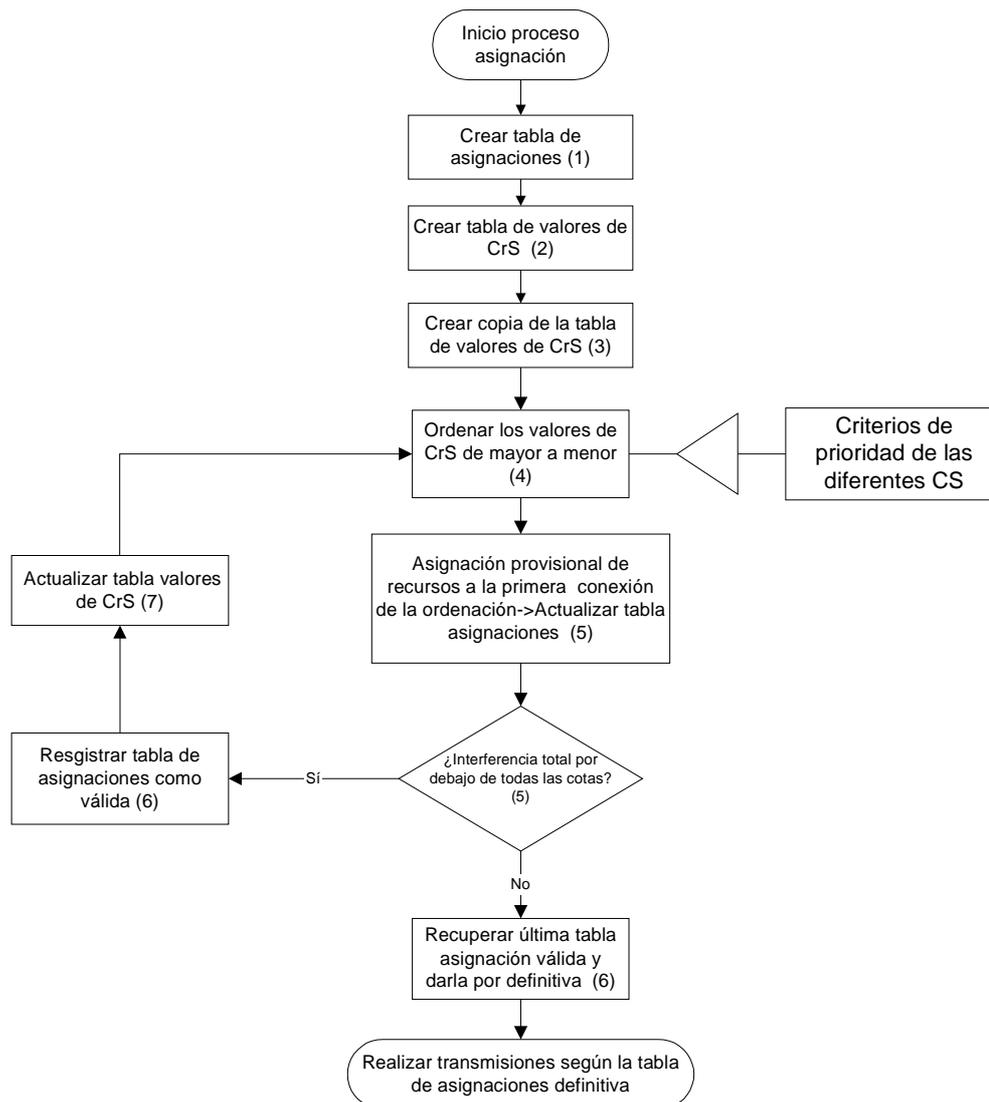


Figura 134. Proceso de asignación de recursos basado en CrS

En la Figura 134 se ha indicado con un número entre paréntesis el punto del algoritmo que representa cada cuadro del diagrama de flujo.

Con este algoritmo, si una conexión acumula suficiente prioridad como para que se le puedan asignar más recursos que los que le corresponderían para transmitir a su tasa básica de transmisión, el algoritmo presentado le dará todos aquellos recursos que sea posible asignarle. Esto ocurre gracias a que una misma conexión puede ser la que resulte seleccionada en el

punto 5 del algoritmo en más de una ocasión. En efecto, si el valor de CrS de una conexión es muy grande, puede ocurrir que incluso después de haberle asignado recursos y calculado su nuevo valor de CrS, este nuevo valor siga siendo el mayor de todos los correspondientes a las demás conexiones y de nuevo quede a la cabeza de la ordenación en el punto 4.

De cara a la evaluación de la interferencia multiusuario, cuando a un usuario se le asignan n veces los recursos mínimos de transmisión, se considerará el caso equivalente de tener n usuarios activos de dicha conexión.

Es importante recalcar que el algoritmo presentado en realidad no presupone ningún método concreto para el cálculo del CrS. Es decir, aunque en el punto 2 se detalla que deben calcularse los valores de CrS según la regla mostrada en 6.2.4.1.2, el mismo algoritmo es válido para cualquier otro método de cálculo que se utilice para el CrS, siempre que se mantenga el criterio de que las conexiones con más prioridad en la asignación de recursos tengan siempre valores más grandes de CrS. A su vez, el mismo algoritmo también es compatible con cualquiera de los métodos de control (como los mencionados en el punto 6.2.4.1) para evitar que comportamientos anómalos de algunas conexiones puedan producir situaciones indeseadas.

6.2.4.3 Ejemplo de asignación

En la Figura 135 se muestra un ejemplo de funcionamiento del algoritmo de gestión de recursos propuesto. En este diagrama se representa el algoritmo correspondiente a un cierto slot de tiempo en el que se asignan recursos de transmisión a seis conexiones activas, que representan a otros tantos usuarios. Cada usuario ejecuta el mismo algoritmo y de este modo conoce la asignación global de recursos del sistema. El ejemplo mostrado es un sistema con 3 clases de servicio (CS) diferentes. Estas clases de servicio tienen las características que se muestran en la Tabla 8. En esta tabla se especifica los siguientes valores:

- *Tasa mínima de transmisión garantizada (en Kbps)*: indica la tasa mínima garantizada por el sistema para las conexiones de cada CS.
- *Bits de información por slot (10 ms)*: tamaño en bits de los paquetes de información transmitidos por cada conexión en un slot del sistema. Indica la tasa básica de transmisión de las conexiones de cada CS.
- *Incremento de CrS por slot (10 ms)*: Valor que se añade al CrS de cada conexión por unidad de tiempo de transmisión (slot). Este valor se calcula a partir de los dos valores anteriores.
- *Retardo máximo por paquete*: retardo máximo que pueden admitir los paquetes de cada conexión, medido en ms. Este valor limita el CrS máximo que puede asignarse a una conexión. Los paquetes que superan este retardo son descartados.
- *Factor de ponderación interferencia*: Es el coeficiente que debe aplicarse a la interferencia que generan los usuarios de una conexión a causa de que transmiten al canal de comunicaciones con una potencia mayor que la de otros usuarios del sistema. Para mantener constante la energía por bit en todas las conexiones, las transmisiones con menor ganancia de procesamiento deben realizarse con una potencia mayor. Este incremento de potencia debe tenerse en cuenta en la evaluación de las interferencias

utilizando este factor de ponderación. La potencia que genera cada usuario debe multiplicarse por este valor.

- *Máxima interferencia tolerada (norm)*: es el valor máximo de interferencia, medido en usuarios que transmiten a la velocidad más lenta, que puede admitir una transmisión para que la calidad (BLER) de sus transmisiones cumpla los requisitos mínimos exigidos por la aplicación final.

Tabla 8. Características de las Clases de Servicio usadas en el ejemplo de asignación

	Clase 0 (CS0)	Clase 1 (CS1)	Clase 2 (CS2)
Tasa mínima garantizada (Kbps)	7.5	15	30
Bits de información por slot (10 ms)	75	150	300
Incremento de CrS por slot (10 ms)	1	1	1
Retardo máximo por paquete (ms) [valor máximo de CrS]	50	1000	1000
Factor ponderación interferencia	1	2	4
Máxima interferencia tolerada (norm)	15	12	10

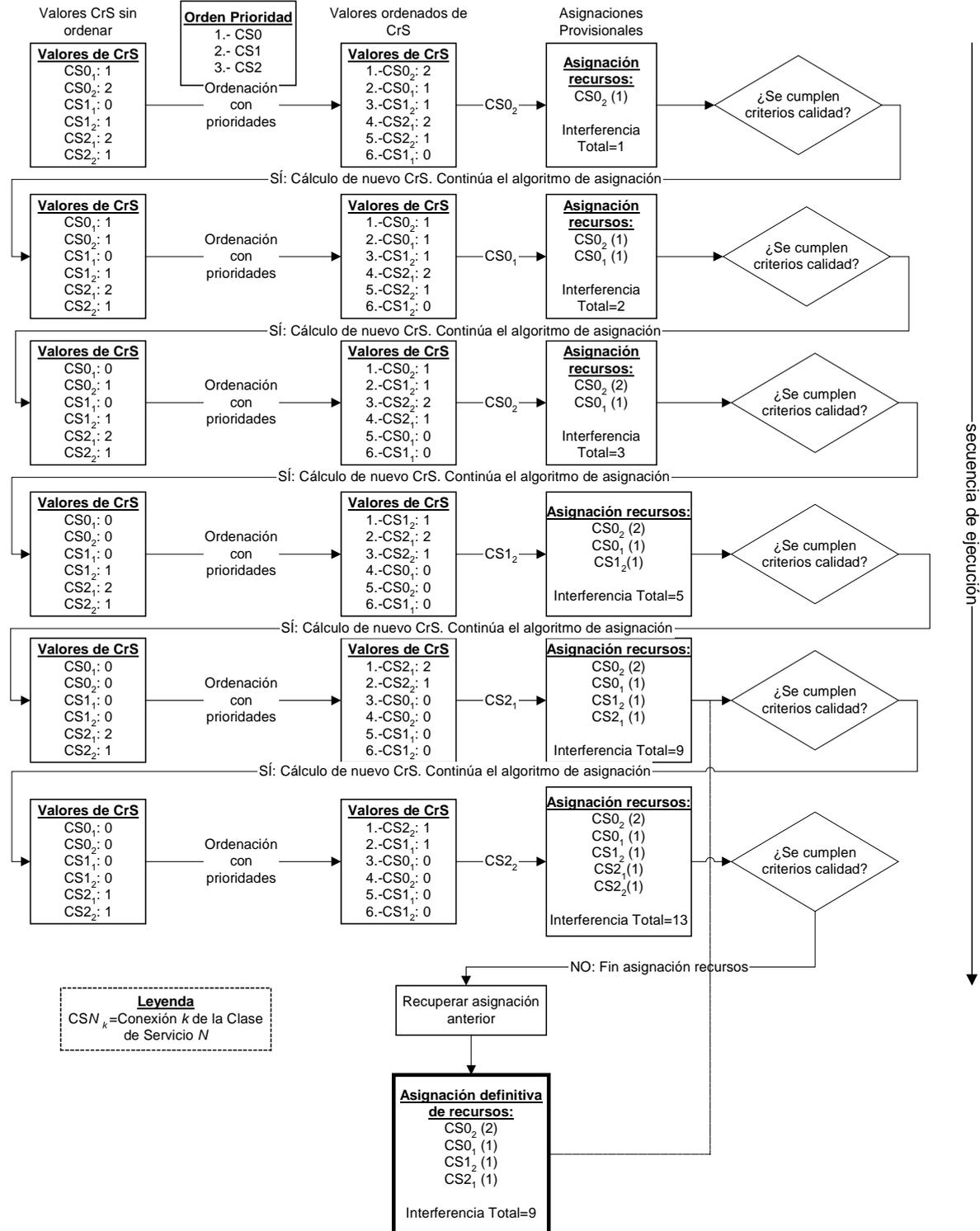


Figura 135. Ejemplo de asignación de recursos

En el ejemplo se ha supuesto que tenemos 2 conexiones de cada una de las CS. La conexión k de la clase de servicio N se denota con las siglas CSN_k . Cada una de ellas tiene un valor de CrS acumulado que se muestra en la figura en el primer cuadro (arriba a la izquierda). Por simplicidad, se ha supuesto que todos los valores de CrS son enteros y que cada clase de servicio aumenta en una unidad su CrS por cada slot de tiempo (tasa mínima garantizada igual a la tasa básica de transmisión de cada conexión). Los valores máximos de CrS de la Tabla 8 son indicativos, puesto que en el ejemplo se han utilizado valores de CrS muy por debajo de este máximo. Tal y como se indica en la figura, las clases de servicio están ordenadas de manera que la clase 0 es la más prioritaria y la clase 1 es la siguiente en prioridad. Cada una de las filas del diagrama de la Figura 135 representa un ciclo del algoritmo de asignación (ver Figura 134).

En cada uno de estos ciclos se realizan las acciones descritas en el punto 6.2.4.2.:

- En la primera columna se muestran los valores de CrS de cada una de las conexiones activas.
- En la segunda columna se ordenan las conexiones según su valor de CrS de mayor a menor, teniendo en cuenta la prioridad de la CS a la que pertenecen. Las conexiones que tienen un valor de CrS igual a 0 se sitúan siempre al final de la lista.
- En el siguiente paso se escoge la conexión que, después de la ordenación, ha quedado en la primera posición de la lista. Se realiza entonces una asignación provisional de recursos a esta conexión. La tercera columna muestra la tabla de asignaciones. Entre paréntesis, detrás de cada conexión, está el número de veces que cada una de ellas ha recibido una asignación de recursos. Este número indica la velocidad relativa con la que dicha conexión podrá enviar su información. Esta velocidad está referida a su tasa básica de transmisión. Se indica sobre la flecha la conexión escogida a la que se le asignan recursos en este paso.
- También en la tercera columna, en la parte inferior de cada tabla de asignaciones, se indica el cálculo de la interferencia total que habría en el sistema con la asignación provisional realizada hasta ese punto, ponderando cada conexión según el factor correspondiente mostrado en la Tabla 8.
- Se verifica entonces si el valor de interferencia total (provisional) está por debajo o por encima del límite permitido para todas las conexiones. El algoritmo compara la potencia interferente calculada con cada uno de los límites de las conexiones activas. Si el valor de interferencia está por debajo de todos los límites se inicia otro ciclo del algoritmo, guardando la asignación de recursos realizada hasta este momento e iniciándose una nueva fila del diagrama.
- Cuando la interferencia total llega a 13, puesto que este valor está por encima del límite para las conexiones de la CS2, que es 10, el sistema decide que la asignación actual ya no es válida. Se recupera la última asignación válida que se realizó en el algoritmo y se finaliza el proceso. Esta asignación final es la definitiva que será utilizada por las conexiones para realizar sus transmisiones.

6.2.5 Gestión de los recursos para el subsistema de acceso

En este punto vamos a proponer una estrategia de gestión de recursos que puede aplicarse al gestor (*scheduler*) correspondiente al subsistema de resolución de colisiones (ver Figura 133). En este sentido cabe destacar que no se encuentran en la literatura referencias que estudien el problema de gestionar el servicio en un sistema de colas en las que los elementos contenidos en ellas son peticiones de acceso de un sistema de transmisión.

De cara a abordar el problema, se plantea una propuesta de esquema de asignación de recursos análoga a la utilizada para los recursos de transmisión. Los recursos a repartir son secuencias código del acceso CDMA. La asignación consiste en reservar unos ciertos códigos para las conexiones de cada CS. En este caso, cada conexión realizará las transmisiones de sus peticiones de acceso utilizando su código asignado.

El diseño incluye una cierta capacidad de adaptación, tanto del algoritmo como de la asignación de códigos, en función de la carga de tráfico ofrecido al sistema. Esta capacidad de adaptación se realiza para aprovechar la capacidad inherente de DQRAP/CDMA de obtener una estimación precisa de la carga de tráfico que cada tipo de servicio está ofreciendo al sistema. El mecanismo de gestión de recursos propuesto se describe a continuación:

Dados los códigos disponibles que podrán ser utilizados en las peticiones de acceso, el algoritmo de asignación de recursos se basa en la definición de unos límites máximos de códigos asignables a las conexiones de cada CS. Estos límites de asignación deberán respetarse en todo momento, para el total de las conexiones de cada CS, excepto cuando dos límites fuesen incompatibles, en cuyo caso se dará prioridad al límite perteneciente a la CS más restrictiva. La asignación de códigos se realiza por orden de prioridad de las CS y de forma que se cumplan los citados límites.

Antes de describir de un modo formal el algoritmo de asignación de códigos, y de cara a ilustrar la filosofía de su funcionamiento, vamos a explicar un ejemplo genérico de funcionamiento del esquema de asignación:

Sea un sistema con tres clases distintas de servicio, que denotaremos por A, B y C. Sea A el servicio más restrictivo en cuanto a tiempo de acceso. El servicio B tiene unas condiciones de retardo intermedias, mientras que el servicio C no tiene restricción alguna en este aspecto. Supongamos que tenemos 6 códigos de acceso (recursos de acceso). Supongamos que la cola de peticiones del servicio B, que denotaremos por CRQ_B , tiene 4 elementos. Es decir, hay 4 colisiones de usuarios del servicio B que deben resolverse. En la notación usual del protocolo, $RQ_B=4$. Por otro lado, y siguiendo con la misma notación, supongamos que $RQ_A=1$ y $RQ_C=1$. En principio sería perfectamente posible que todas las peticiones de todos los servicios fueran atendidas. Por tanto, el gestor o *scheduler* podría asignar el primer código al servicio A, los 4 siguientes al servicio B y el siguiente al servicio C. Este reparto, sin embargo, no dejaría ningún código disponible para los nuevos accesos que llegan al sistema (los que aún no están en las colas CRQ pero que entran en el slot actual). En el caso de servicios muy sensibles al retardo, en el ejemplo el servicio A, puede resultar adecuado traspasar códigos del servicio C al A de forma que queden códigos libres para los nuevos accesos de A. Esto redundaría en una menor probabilidad de demora en el primero acceso de los paquetes de A. Se decide que las clases de servicio tengan los límites de asignación de códigos detallados en la Tabla 9. Estos límites son:

- *Máximos códigos asignables*: Es el número máximo de códigos que pueden asignarse para resolver colisiones de la clase de servicio correspondiente. Este valor representa una cota máxima de recursos para evitar que una carga elevada de tráfico en una clase de servicio más prioritaria deje sin recursos al resto de clases.
- *Máximos códigos nuevos accesos*: Es el número máximo de códigos que pueden reservarse para las nuevas peticiones de la clase de servicio correspondiente. Este valor también es una cota que permite asegurar que los códigos que no se utilizan para la resolución de colisiones se repartan de una manera equitativa entre las clases de servicio, aunque manteniendo la prioridad.

El servicio C es *best effort*, y por tanto podemos dar preponderancia a las conexiones del servicio A, que es más prioritario, y traspasar un código hacia este servicio. De este modo, los 6 códigos se repartirían 2 para el servicio A (uno para los nuevos accesos) y 4 para el servicio B.

Tabla 9. Límites de asignación para las CS

Clase de servicio	A	B	C
Máximos códigos asignables	6	5	3
Máximos códigos nuevos accesos	2	2	2

La Figura 136 ilustra el ejemplo propuesto. Podemos observar que los umbrales de asignación de recursos pueden alterarse de forma dinámica en función de las condiciones de tráfico del sistema. Este mecanismo permite una mayor eficiencia en el uso de los recursos de acceso.

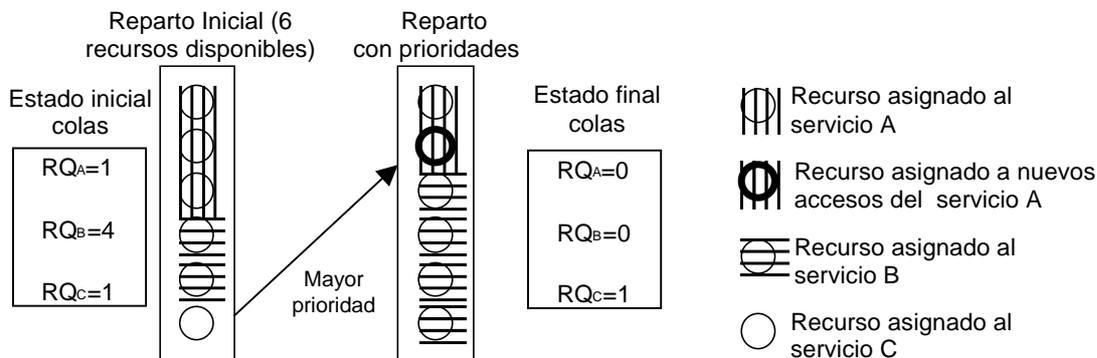


Figura 136. Ejemplo de reparto de recursos de acceso

En el caso en el que las peticiones contenidas en las colas CRQ no ocupen todos los recursos disponibles (todos los códigos), algunos de los códigos quedarán libres. Estos códigos podrán repartirse para ser usados por las nuevas peticiones de todas las clases de servicio. Este reparto se realizará con la misma mecánica utilizada para los códigos asignados a las colisiones de las colas CRQ.

6.2.5.1 Algoritmo de asignación

Glosando la definición, el diseño del sistema que se utilizará sigue el siguiente algoritmo de funcionamiento:

1. Llamaremos K al número total de códigos disponibles para realizar peticiones nuevas y resolver colisiones. Se define un límite máximo de asignación de códigos para las

peticiones de acceso correspondientes a la resolución de colisiones de las conexiones de cada CS. Para la CS_i (donde i indica el número de la CS), llamaremos $R_{i_{\max}}$ al número máximo de códigos asignables para resolver las colisiones de las conexiones de dicha clase.

2. Se define también un límite máximo de asignación de códigos para las nuevas peticiones de acceso de las conexiones de cada CS. Llamaremos $N_{i_{\max}}$ al número máximo de códigos asignables para los nuevos intentos de acceso de las conexiones de la clase CS_i .
3. Se selecciona la CS de más alta prioridad en la asignación. Denotaremos esta clase como CS_0 . Llamaremos CRQ_0 a la cola de resolución de colisiones correspondiente a la CS_0 . Por tanto, siguiendo la notación usual del protocolo, RQ_0 será el número de elementos contenidos en esta cola.
 - a. Si $RQ_0 \leq R_{0_{\max}}$ se asigna un código a todas las colisiones contenidas en CRQ_0 , de modo que a la colisión que tiene la primera posición se le asigna el primer código disponible, a la que tiene la segunda posición el segundo código disponible, y así sucesivamente.
 - i. Si $R_{0_{\max}} - RQ_0 \geq N_{0_{\max}}$ se reservan $N_{0_{\max}}$ códigos para las nuevas peticiones de acceso de la clase CS_0 . Los $K - RQ_0 - N_{0_{\max}}$ códigos restantes no asignados quedan a disposición del resto de las clases de servicio.
 - ii. Si $R_{0_{\max}} - RQ_0 < N_{0_{\max}}$ se asignan $R_{0_{\max}} - RQ_0$ códigos a las nuevas peticiones de acceso de la clase CS_0 . Los $K - R_{0_{\max}}$ códigos restantes no asignados quedan a disposición del resto de las clases de servicio.
 - b. Si $RQ_0 > R_{0_{\max}}$ se asigna un código únicamente a las $R_{0_{\max}}$ primeras colisiones de la cola CRQ_0 . Esta asignación se realiza del mismo modo, es decir, a la colisión que tiene la primera posición se le asigna el primer código disponible, a la que tiene la segunda posición el siguiente código disponible, y así sucesivamente. Los $K - R_{0_{\max}}$ códigos restantes quedan a disposición del resto de clases de servicio.
4. Se selecciona la siguiente CS con más prioridad. Supongamos que es la clase CS_i .
 - a. Si $RQ_i \leq R_{i_{\max}}$ se asigna un código a todas las colisiones contenidas en CRQ_i , de modo que a la colisión que tiene la primera posición se le asigna el primer código disponible de los que quedan libres, que son: $K - \min(RQ_0 + N_{0_{\max}}, R_{0_{\max}})$. A la que tiene la segunda posición el siguiente código disponible, y así sucesivamente. En el caso de que, en un momento dado, ya no queden códigos disponibles para realizar la asignación, el proceso del algoritmo finaliza. Llamamos ahora K_d a los códigos que aún no han sido asignados a ninguna clase de servicio.
 - i. Si $R_{i_{\max}} - RQ_i \geq N_{i_{\max}}$ se reservan $N_{i_{\max}}$ códigos para las nuevas peticiones de acceso de la clase CS_i . Si no quedan suficientes códigos libres como para hacer la asignación, se asignan todos los restantes y se

finaliza el algoritmo. Los $K_d - RQ_i - Ni_{\max}$ códigos restantes quedan a disposición del resto de las clases de servicio.

- ii. Si $Ri_{\max} - RQ_i < Ni_{\max}$ se asignan $Ri_{\max} - RQ_i$ códigos a las nuevas peticiones de acceso de la clase CS_i . Si no quedan suficientes códigos libres como para hacer la asignación, se asignan todos los restantes y se finaliza el algoritmo. Los $K_d - Ri_{\max}$ códigos restantes no asignados quedan a disposición del resto de las clases de servicio.
 - b. Si $RQ_i > Ri_{\max}$ se asigna un código únicamente a las Ri_{\max} primeras colisiones de la cola CRQ_i . Esta asignación se realiza del mismo modo, es decir, a la que tiene la primera posición en la cola se le asigna el primer código disponible, a la que tiene la segunda posición el siguiente código disponible, y así sucesivamente. En cualquier momento, si se agotan los códigos disponibles, se detiene el proceso de asignación y el algoritmo finaliza. El resto de códigos que aún no han sido asignados quedan a disposición del resto de clases de servicio.
5. Si quedan CS que aún no han recibido asignación de recursos, se vuelve al punto 4.
 6. Si aún quedan códigos disponibles, se selecciona la clase con más prioridad (CS0) y se intentan reservar hasta NO_{\max} códigos para las nuevas peticiones de acceso de las conexiones de esta clase. Si no quedan suficientes códigos disponibles, se asignan todos los que quedan y se finaliza el algoritmo.
 7. Se selecciona la siguiente clase más prioritaria. Supongamos que es la CS_i . Se intentan reservar hasta Ni_{\max} códigos para las nuevas peticiones de acceso de las conexiones de esta clase. Si no quedan suficientes códigos disponibles, se asignan todos los que quedan y se finaliza el algoritmo.
 8. Si aún quedan códigos disponibles, se vuelve al punto 7. Cuando en este punto ya no quedan clases de servicio por escoger, el algoritmo finaliza.

6.2.5.2 Ejemplo de asignación

Supongamos que tenemos un sistema con tres CS diferentes, por orden decreciente de prioridad: CS0, CS1 y CS2. El sistema dispone de $K=17$ códigos disponibles para realizar las peticiones de acceso. En la Tabla 10 se definen los límites de asignación de códigos y los valores de ocupación de las colas de cada una de las CS. La ejecución del algoritmo sigue los siguientes pasos:

Tabla 10. Límites de asignación para las CS

CS	CS0	CS1	CS2
Ri_{\max}	6	5	2
Ni_{\max}	4	2	2
RQ_i	5	3	5

La definición de la Tabla 10 incluye la ejecución de los puntos 1 y 2 del algoritmo.

En el punto 3 se selecciona la clase CS0. Dado que $RQ_0 < R0_{\max}$, se asignan los primeros $RQ_0=5$ códigos a las correspondientes primeras colisiones por resolver de la clase CS0. Quedan por tanto $K=17 - 5=12$ códigos disponibles por asignar.

En el mismo punto, dado que $R0_{max}-RQ_0 < N0_{max}$ se asignan únicamente $R0_{max}-RQ_0=1$ códigos a las nuevas peticiones de la clase CS0. Quedan por tanto $K=11$ códigos libres.

Se pasa al punto 4 y se selecciona la CS1. Puesto que $RQ_1 < R1_{max}$ se asignan los siguientes $RQ_1=3$ códigos disponibles a las colisiones de la clase CS1. Quedan por tanto $K=8$ códigos libres.

En este mismo punto, puesto que $R1_{max}-RQ_1=N1_{max}$ ($5-3=2$), se asignan $N1_{max}=2$ códigos a las nuevas peticiones de la clase CS1. Quedan por tanto $K=6$ códigos libres.

El algoritmo llega al punto 5. Dado que la clase CS2 aún no ha recibido recursos, el algoritmo regresa al punto 4. Se selecciona la CS2 y puesto que $RQ_2 > R2_{max}$ se asignan $R2_{max}=2$ códigos para las colisiones de la clase CS2. Quedan por tanto $K=4$ códigos disponibles.

En el punto 5 esta vez no se regresa al punto 4, puesto que ya hemos recorrido todas las CS del sistema. La ejecución pasa al punto 6. En este punto se intentan reservar $N0_{max}=4$ códigos para las nuevas peticiones de la CS0. Puesto que ya se ha asignado un código a estas peticiones, se realiza la asignación de 3 códigos más y aún quedan por tanto $K=1$ códigos libres.

En el punto 7, se intentan reservar $N1_{max}=2$ códigos para las nuevas peticiones de la CS1. Puesto que esta asignación ya se ha hecho, no se vuelve a realizar.

En el punto 8, como aún quedan códigos disponibles, se regresa al punto 7 del algoritmo. En este punto se intentan reservar $N2_{max}=2$ códigos para las nuevas peticiones de la clase CS2. Puesto que sólo queda 1 disponible, se asigna este código y se finaliza el algoritmo.

El resultado final de la asignación se muestra en la Tabla 11. Puede comprobarse que se han repartido los 17 códigos disponibles inicialmente.

Tabla 11. Asignación final del códigos obtenida del algoritmo

CS	CS0	CS1	CS2
Códigos para resolver colisiones	5	3	2
Códigos para nuevos accesos	4	2	1

En el siguiente punto se muestran los resultados obtenidos en las simulaciones de un sistema que utiliza este esquema de asignación de recursos. En la descripción se indican los valores utilizados para los límites de asignación códigos.

6.2.6 Simulaciones y resultados

Usando el esquema de sistema presentado hasta este punto, es decir, con las propuestas de gestión de recursos presentadas en los puntos 6.2.4 y 6.2.5, se han realizado un serie de simulaciones por ordenador para comprobar y validar el funcionamiento de lo que representa un sistema de comunicaciones móviles basado en DQRAP/CDMA que pueda ofrecer una cierta calidad de servicio garantizada. En este apartado se presentan los resultados obtenidos.

El sistema simulado consiste en una estación base que se comunica con un conjunto de usuarios que tienen conexiones activas que pertenecen a su vez a un conjunto de varias CS. Cada una de las conexiones genera tráfico (paquetes de datos) siguiendo diferentes modelos que se detallan en el punto 6.2.6.1. Estas conexiones se ordenan según un criterio de prioridad que depende del tipo de aplicación que soportan, es decir, de la CS a la que pertenecen. Así,

tal y como se ha hecho hasta este momento, llamaremos CS_i a la CS de orden i en el orden de prioridad, donde CS_0 será la clase con mayor prioridad, CS_1 la siguiente más prioritaria, y así sucesivamente. Normalmente, las CS que tienen unos parámetros de calidad más restrictivos deberán tener asignado un orden superior en la tabla de prioridades.

Por consiguiente, los gestores correspondientes al subsistema de resolución de colisiones y al subsistema de transmisión (ver Figura 133) incorporan los algoritmos de gestión de recursos presentados respectivamente en los puntos 6.2.4 y 6.2.5.

6.2.6.1 Modelos de tráfico

En este punto vamos a detallar los modelos de tráfico de las CS que se han utilizado en las simulaciones de sistema realizadas. El objetivo de estas simulaciones es la verificación del correcto funcionamiento de los mecanismos de gestión de recursos expuestos en los puntos anteriores. Por simplicidad, y sin pérdida de generalidad del estudio, utilizaremos cuatro CS diferentes. Cada una de ellas corresponde a una de las aplicaciones más representativas de los sistemas de comunicaciones digitales. Se presentan en el mismo orden aplicado para decidir la prioridad de cada una de ellas en el sistema de asignación de recursos.

6.2.6.1.1 Voz (ON-OFF)

Se utiliza este modelo para representar a usuarios que generan paquetes de datos correspondientes a un flujo de audio vocal. El modelo utilizado es un modelo clásico ON-OFF representado por el diagrama de estados mostrado en la Figura 137. Los usuarios parten de un estado OFF o de silencio, en el que no generan información. Cada cierto tiempo tienen una probabilidad α de pasar al estado ON o activo, en el que generan paquetes de información a tasa constante (son los llamados *talkspurts* o ráfagas vocales). En este estado tienen a su vez una probabilidad β de volver al estado OFF. Por tanto, el tiempo medio de estancia en cada estado es una variable aleatoria exponencial de media $1/\alpha$ y $1/\beta$ respectivamente. El factor de actividad, o probabilidad de encontrar al usuario en el estado activo, es $\alpha/(\alpha+\beta)$ e indica la tasa media de transmisión de cada conexión, que no coincide con a la tasa de generación de información en el estado activo.

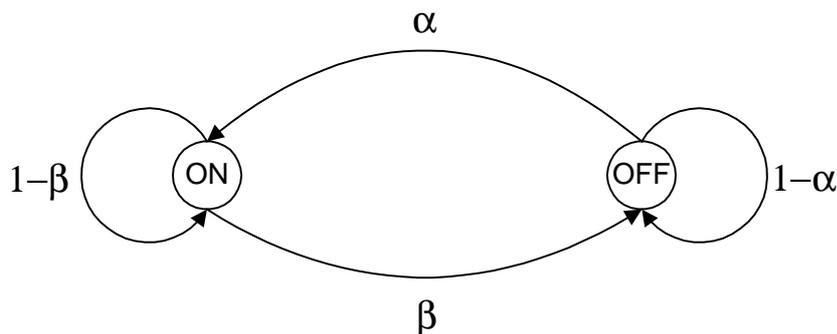


Figura 137. Modelo de tráfico ON-OFF

Las características de tráfico de la CS correspondiente a este tipo de usuarios se definen haciendo uso de los parámetros detallados en el punto 6.1.2. Los valores de los parámetros usados para este tipo de tráfico se detallan en la Tabla 12.

Tabla 12

Número Servicio	0 (Voz)
Modelo de generación tráfico	ON-OFF
Factor de actividad	44%
Duración media de los períodos ON	1 s
Duración media de los períodos OFF	1'35 s
Tasa mínima para cálculo CrS (Kbps)	6'6
Tasa básica (Kbps)	15
Ganancia de procesado	256
Retardo máximo paquetes	30 ms
Porcentaje máximo paquetes perdidos	1%
Tasa de codificación	1/2
BER máximo	10^{-3}

Debe reseñarse que la ganancia de procesado que se especifica en la tabla corresponde a la tasa básica de transmisión de la Clase de Servicio. Según los criterios del algoritmo de gestión de recursos adoptado y el número de códigos de transmisión asignados a estas conexiones, es perfectamente posible que se transmita con una con una tasa mayor que la básica. Recordemos que esta adaptación de velocidades se realiza mediante la técnica de transmisión multicódigo. Por otro lado, y dado el estricto requerimiento en cuanto a retardo de los paquetes, en este tipo de conexiones no se hace uso de ninguna técnica ARQ para el control del flujo de los paquetes.

6.2.6.1.2 Datos interactivos (VBR)

Este modelo representa aquellas aplicaciones de datos que generan información de manera racheada, cuya tasa de generación de información es variable en el tiempo y que tienen unos requerimientos de retardo ajustados. Algunos ejemplos típicos del tipo de aplicaciones que se pretenden emular son las aplicaciones de datos interactivas, como por ejemplo conversaciones en línea (*chats*) o juegos en red. Se ha modelado este tipo de usuarios haciendo uso de una estructura MMPP (*Markov Modulated Poisson Process*) consistente en un diagrama de estados como el mostrado en la Figura 138.

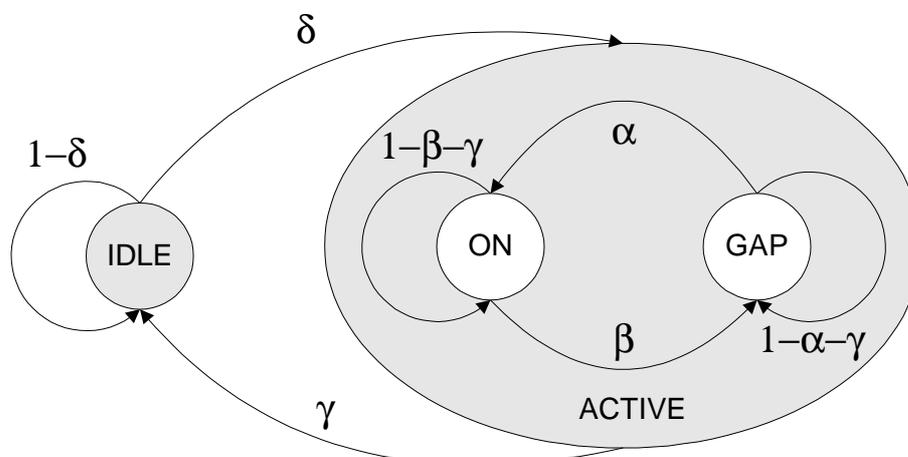


Figura 138. Diagrama de estados del modelo de tráfico VBR

Los usuarios están inicialmente en un estado de reposo o silencio (*IDLE*) en el que no generan bits de información. Con una cierta probabilidad (δ) pueden pasar a un estado activo (*ACTIVE*) en el que se distinguen a su vez dos nuevos estados, uno de silencio (*GAP*) en el que tampoco se genera información, y otro de actividad (*ON*) en el que se genera información a tasa constante. Visto de otro modo, el proceso de generación consiste en un modelo ON-OFF en el que la generación de paquetes de datos en los períodos ON no es constante sino que sigue un proceso de Poisson con longitud de mensajes exponencial, es decir, otro modelo ON-OFF insertado dentro del estado ON del primero. Los estados *ON* y *GAP* representan las pequeñas ráfagas de información o de silencio que se generan dentro del estado activo. Esto es, podemos considerar que dentro de los períodos en el estado *ACTIVE* la tasa de generación es variable y se caracteriza por una cierta tasa media, que sería el equivalente al factor de actividad del modelo ON-OFF contenido en este estado.

Los parámetros que definen esta CS de cara a las simulaciones de sistema realizadas se detallan en la Tabla 13.

Tabla 13

Número Servicio	1 (datos interactivos)
Modelo de generación tráfico	MMPP
Factor de actividad global	50 %
Tasa de llegada de mensajes en <i>ACTIVE</i> (α)	0,02 (por usuario)
Longitud media de mensajes en <i>ACTIVE</i> ($1/\beta$)	1000 bits (por usuario)
Tasa mínima para cálculo CrS (Kbps)	1
Tasa básica (Kbps)	30
Ganancia de procesado	128
Retardo máximo paquetes	500 ms
Porcentaje máximo paquetes perdidos	0,1%
Tasa de codificación	1/2
BER máximo	10^{-6}

Cabe recalcar de nuevo que la ganancia de procesado indicada en esta tabla corresponde a la tasa básica de transmisión de la CS, pero que es posible transmitir a velocidades múltiples de este valor si el resultado de la asignación de recursos de transmisión así lo indica.

Es importante reseñar también que en la simulaciones realizadas se ha incluido un cierto control a nivel de enlace que implementa una estrategia ARQ para conseguir que las transmisiones sean fiables. Por simplicidad, se ha utilizado una estrategia ARQ simple del tipo Stop&Wait. Por tanto, y a diferencia de lo que ocurre con las transmisiones de paquetes de voz, en el rendimiento del sistema se deben tener en cuenta las retransmisiones de aquellos paquetes que han resultado erróneos debido a la interferencia multiusuario.

6.2.6.1.3 Datos no interactivos (IP)

Con este modelo de tráfico se pretende representar a todas aquellas transmisiones de datos no en tiempo real que sin embargo sí tienen unas ciertas restricciones de retardo máximo que deben garantizarse. El modelo se basa en las estadísticas empíricas realizadas por el grupo de estandarización IEEE 802.14 sobre tráfico de Internet (IP) y consiste en la generación de mensajes con llegadas de Poisson (tiempo entre llegadas exponencial) cuyos tamaños siguen

la distribución empírica que se muestra en la Figura 139. Es un modelo equivalente al utilizado en el punto 3.10.4.

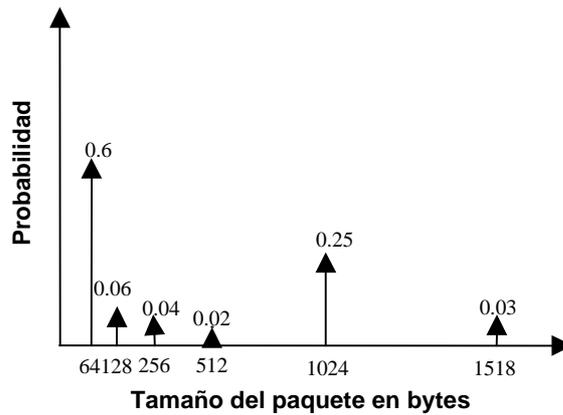


Figura 139. Distribución del tamaño de los mensajes IP

Esta distribución está basada en el hecho de que los *datagramas* del protocolo IP no pueden tener longitudes arbitrarias, y por tanto la distribución probabilística de las longitudes de los mensajes es discreta.

La Tabla 14 muestra los parámetros de tráfico que definen la CS correspondiente.

Tabla 14

Número Servicio	2 (datos no interactivos, IP)
Modelo de generación tráfico	Basado en IEEE 802.14
Tasa mínima para cálculo CrS (Kbps)	0,5
Tasa básica (Kbps)	60
Ganancia de procesado	64
Retardo máximo paquetes	1 s
Porcentaje máximo paquetes perdidos	0,01%
Tasa de codificación	1/2
BER máximo	10^{-6}

Al tratarse de una transmisión de datos, se ha incluido una estrategia ARQ simple del tipo Stop&Wait en el control del enlace de cara a dotar de fiabilidad a las transmisiones de paquetes de este servicio.

6.2.6.1.4 Datos genéricos (best-effort)

En esta CS se engloban todos los servicios y conexiones que no tienen ningún requerimiento de calidad específico y que por tanto quedan relegados al final del orden de prioridades. Son las llamadas comunicaciones *best-effort*. Estas conexiones deberán utilizar los recursos que dejen libres las demás CS una vez realizada la gestión de recursos correspondiente. En cualquier caso, puede arbitrarse la reserva de unos mínimos recursos para este servicio de cara a evitar que un excesivo tráfico general en el sistema deje totalmente bloqueadas sus conexiones de manera demasiado continuada en el tiempo.

El modelo de generación de tráfico utilizado consiste en hacer uso de llegadas con estadística de Poisson y suponer que cada mensaje generado tiene una longitud cuyo valor es una

variable aleatoria exponencial. Los parámetros que definen esta CS se muestran en la Tabla 15.

Tabla 15

Número Servicio	3 (<i>best-effort</i>)
Modelo de generación tráfico	Poisson
Tasa de llegada de mensajes (λ)	0,01 (por usuario)
Longitud media de los mensajes ($1/\mu$)	6000 bits
Tasa mínima para cálculo CS (Kbps)	NA
Tasa básica (Kbps)	120
Ganancia de procesado	32
Retardo máximo paquetes	No
Porcentaje máximo paquetes perdidos	0%
Tasa de codificación	1/2
BER máximo	10^{-6}

Nótese que al ser un servicio *best-effort* no se realiza ningún cálculo del valor de CrS que tiene cada conexión. Las conexiones de este servicio quedan siempre relegadas al final de la lista de prioridades y tan sólo reciben recursos cuando las demás CS dejan recursos libres. Se ha incluido también un mecanismo ARQ simple del tipo Stop&Wait para el control del enlace.

6.2.6.2 Codificación de canal

Uno de los elementos necesarios de todo sistema de comunicaciones digitales es la codificación de canal. Las técnicas englobadas dentro de este concepto son las encargadas de dotar de fiabilidad a las transmisiones digitales de datos. Todas ellas consisten en añadir una cierta redundancia a la información transmitida que permita detectar o corregir los posibles errores que haya introducido el canal de transmisión.

Hasta este momento, en todos los sistemas propuestos se ha incluido únicamente la existencia implícita de algún código de detección que permita saber qué bloques de información contienen errores y de ese modo solicitar las retransmisiones necesarias haciendo uso de un mecanismo de ARQ simple. Sin embargo, de cara a poder evaluar las prestaciones de un sistema en el que tenemos tráfico heterogéneo proveniente de diversas fuentes (CS) es necesario tener en cuenta algún mecanismo de codificación de canal de mayor complejidad. Básicamente, existen dos tipo de codificación de canal:

- Códigos bloque
- Códigos convolucionales

En las transmisiones en modo paquete, dada la propia estructura de transmisión, la codificación tipo bloque es la más ampliamente usada. Sin embargo, cuando la transmisión es continua normalmente se hace uso de una codificación tipo convolucional.

Cae fuera del ámbito de nuestro estudio el entrar en el detalle del estudio del tipo de codificación más apropiado para el entorno de trabajo utilizado hasta el momento. Por otro lado, el grado de codificación usado también suele ser dependiente del servicio que es soportado por el sistema de comunicaciones. Es más, en muchas ocasiones, suele utilizarse simultáneamente más de un nivel de codificación, con frecuencia combinando los dos tipos

mencionados anteriormente. En este sentido, los Turbo Códigos [56], que recientemente han experimentado un auge de cierta consideración, son un ejemplo paradigmático de la combinación de esquemas simples de codificación para obtener mejores resultados. Sin embargo, parece evidente que, puesto que el entorno de trabajo en el que nos movemos es un sistema de comunicaciones cuya transmisión se realiza en modo paquete, la codificación tipo bloque será la que de forma natural será utilizada.

En cualquier caso, debemos recordar que nuestro objetivo es evaluar las prestaciones de un sistema de gestión de recursos radio, y para ello tan solo es necesario conocer las prestaciones del esquema de codificación que queramos ‘simular’ que hemos incluido en el mismo. En general, estas prestaciones son muy dependientes del código concreto utilizado. Sin embargo, existen en la literatura cotas matemáticas que, dada la tasa de redundancia incluida en la información, nos dan un valor mínimo de la capacidad tanto detectora como correctora de cualquier código utilizado. En particular nos interesa la cota llamada de Varsharmov-Gilbert [57], que dice lo siguiente:

Siempre existe un código lineal (n,k) cuya distancia mínima al menos vale d_{\min} y que satisface la siguiente desigualdad:

$$2^{n-k} \leq \sum_{i=1}^{d_{\min}} \binom{n}{i} \quad (6.1)$$

Por otro lado, es importante recordar que, dado un código lineal con distancia mínima d_{\min} y llamando t al número de errores que será capaz de corregir, entonces se cumple que:

$$d_{\min} \geq 2t + 1 \quad (6.2)$$

Lo que nos permite calcular la capacidad correctora de un código lineal cualquiera a partir del conocimiento de su distancia mínima. Además, el número de errores que se podrán detectar con total seguridad es igual a $d_{\min}-1$, con lo que la distancia mínima también determina la capacidad de detección de un código lineal cualquiera. En principio estas cotas se aplican a códigos bloque lineales, ya que en caso de los códigos convolucionales el rendimiento es muy dependiente del código en particular que se utilice. De todos modos, es posible establecer la aproximación de que un código convolucional de tasa $1/r$ tiene al menos las mismas capacidades que un código bloque $(r \cdot k, k)$ equivalente, donde la tasa de redundancia es la misma en ambos casos [58].

Por ello, y sin perder generalidad, en el estudio de nuestro sistema de comunicaciones supondremos que tenemos un cierto código donde por cada k bits de información generamos n bits de palabra código sin especificar si es una codificación bloque o convolucional. Haremos uso de (1) para evaluar la capacidad detectora y correctora de dicho código. De este modo, realizaremos el estudio independiente del tipo concreto de codificación usada y por tanto las conclusiones serán más generales.

6.2.6.3 Resultados

Teniendo presente la descripción del sistema presentada en este capítulo, se han realizado una serie de simulaciones por ordenador para comprobar el funcionamiento de la estructura de gestión de recursos propuesta y su adecuación a las necesidades planteadas.

A continuación se muestran las gráficas obtenidas para el sistema con diferentes combinaciones de números de usuarios de cada una de las CS descritas anteriormente (en los puntos desde el 6.2.6.1.1 hasta el 6.2.6.1.4). Se muestran las gráficas de las funciones de distribución (PDF) del retardo de cada uno de los paquetes de información. Estas gráficas permiten conocer, para cada CS, el porcentaje de paquetes que se reciben dentro de un cierto intervalo de retardo. Por ejemplo, si para un retardo de 30 ms el valor de la PDF es 0'86, esto indica que el 86% de los paquetes de la conexión se entregan en menos de 30 ms. En todos los casos y para todas las clases de servicio se ha mantenido el tiempo de trama (tiempo de transmisión de cada paquete) en 10 ms. Los parámetros generales de las simulaciones así como los de la gestión de recursos basada en el protocolo DQRAP/CDMA y en el criterio de Crédito de Servicio se detallan en la Tabla 16. La descripción de los datos de esta tabla es la siguiente:

- *Duración de sistema de cada simulación*: En combinación con la duración de las tramas, indica el número de tramas simuladas para calcular cada uno de los puntos de cada gráfica.
- *Tiempo de trama (todos los servicios)*: Duración de cada una de las tramas en las que cada servicio transmite sus paquetes de datos.
- *Número de minislots de acceso (m)*: Número de minislots de acceso por cada slot de datos transmitido.
- *Tamaño del minislot de acceso*: Número de chips de la secuencia de acceso enviada por los usuarios en cada minislot. Define la probabilidad de detectar correctamente una petición de acceso.
- *Máx. núm. códigos para el acceso (Servicio i)*: Número máximo de códigos asignables para realizar las peticiones de acceso de la clase de servicio correspondiente. Se corresponde con el valor de $R_{i_{max}}$ definido en el punto 6.2.5.1. En este sentido, este valor también se ha hecho corresponder con el valor de $N_{i_{max}}$.
- *Número max. de transmisiones simultáneas (Servicio i)*: Número máximo de transmisiones de datos simultáneas que pueden realizarse para cada una de las clases de servicio.

Tabla 16

Duración de sistema de cada simulación	5000 s. (83'3 min)
Tiempo de trama (todos los servicios)	10 ms
Número de minislots de acceso (<i>m</i>)	3 (por slot)
Tamaño minislot de acceso	256 chips
Máx. núm. códigos para el acceso: Servicio 0	40
Máx. núm. códigos para el acceso: Servicio 1	30
Máx. núm. códigos para el acceso: Servicio 2	30
Máx. núm. códigos para el acceso: Servicio 3	25
Número max. de transmisiones simultáneas: Servicio 0	170
Número max. de transmisiones simultáneas: Servicio 1	123
Número max. de transmisiones simultáneas: Servicio 2	65
Número max. de transmisiones simultáneas: Servicio 3	34

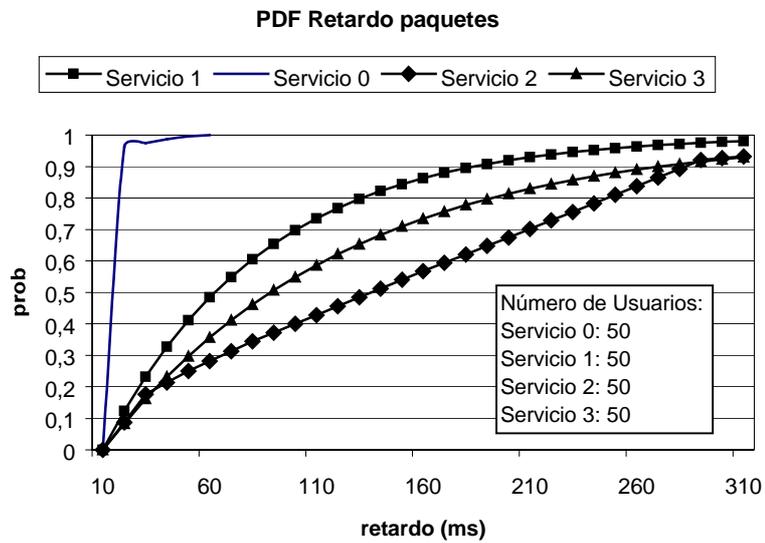


Figura 140. PDF retardo paquetes

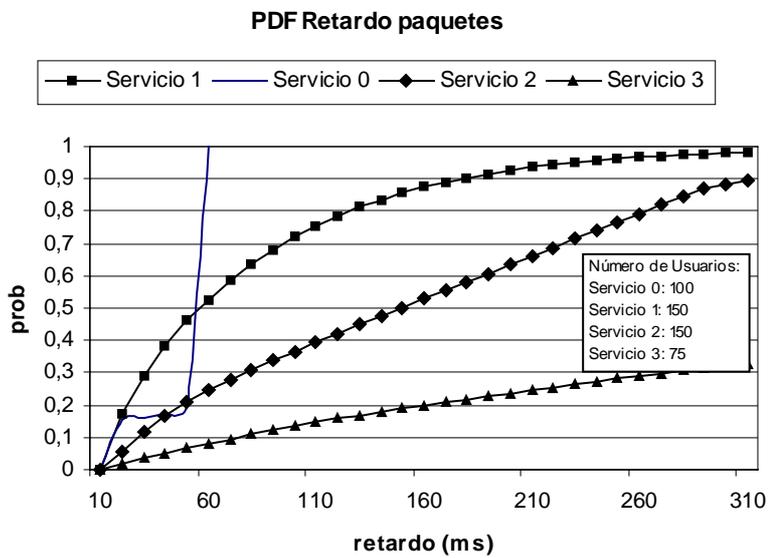


Figura 141. PDF retardo paquetes

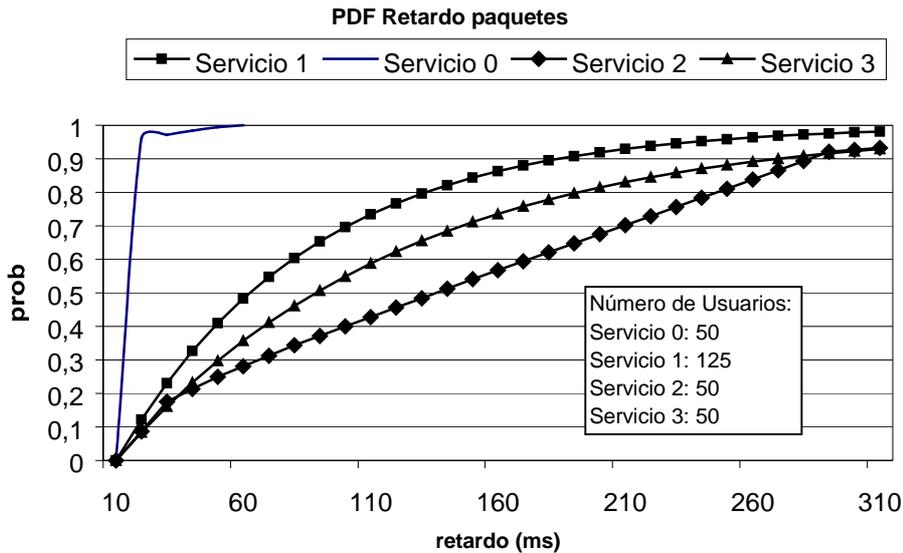


Figura 142. PDF retardo paquetes

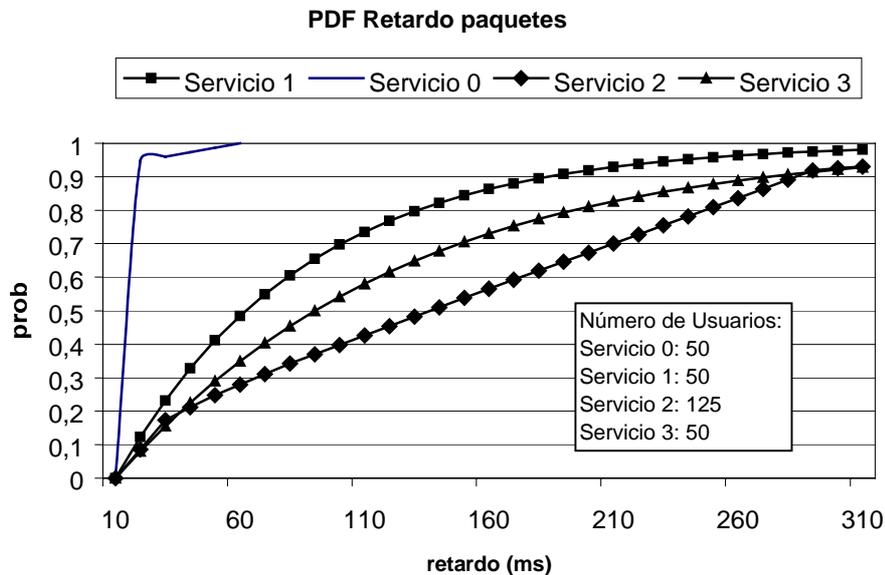


Figura 143. PDF Retardo paquetes

En las dos primeras gráficas (Figura 140 y Figura 141) podemos comparar el rendimiento del sistema para dos situaciones de carga muy diferentes. En la Figura 140 los resultados mostrados son de un sistema que trabaja con una carga de tráfico relativamente baja, con 50 usuarios de cada una de las clases de servicio. Se puede observar la diferencia de estos resultados con los obtenidos en el caso de que la carga ofrecida sea mucho más elevada (Figura 141). Los resultados que muestra esta gráfica fueron obtenidos con un sistema con 100 usuarios de voz, 150 de datos VBR e interactivos, y 50 usuarios *best-effort*. En ambas situaciones se mantienen los criterios de prioridad para las diferentes CS. La Tabla 17 muestra los retardos garantizados en el 90% de los paquetes de cada CS de forma comparativa entre las dos situaciones. Vemos como la carga del segundo escenario está por encima del máximo tolerado para poder dar el servicio requerido por el servicio de voz. Es por tanto un caso de sobrecarga que debería ser evitado por el sistema de control de admisión.

Tabla 17

Retardo garantizado en el 90% casos	Baja carga (Figura 140)	Alta carga (Figura 141)
CS0	20 ms	60 ms
CS1	180 ms	180 ms
CS2	260 ms	310 ms
CS3	280 ms	900 ms

Se comprueba a su vez que para cargas crecientes de tráfico (mayor número de usuarios) la curva que resulta más afectada es la correspondiente al servicio 3, es decir, al servicio *best-effort*. Este hecho viene a corroborar que está funcionando correctamente el mecanismo de prioridades que asegura que sólo parte de la capacidad ‘sobrante’ en el sistema es asignada a los usuarios de este tipo.

Por otro lado, las gráficas de las Figura 142 y Figura 143 muestran una comparativa de la respuesta del sistema a dos situaciones en las que el número de usuarios total es el mismo pero distribuidos de forma distinta entre CS. Se puede observar como el sistema se comporta de forma similar en ambas situaciones. Se cumple la calidad requerida para todas las conexiones y puesto que la carga equivalente total es similar, el comportamiento del sistema también es muy parecido. Se sigue manteniendo la prioridad de servicio para las conexiones y las curvas de retardo acumulado mantienen las mismas características.

Un ejemplo significativo de la mejora que introduce el esquema de gestión de recursos propuesto es el que se muestra en la Figura 144 y la Figura 145. En ellas se podemos observar una comparativa entre las curvas de probabilidad acumulada de retardo para el mismo escenario (en cuanto a número de usuarios de cada CS) pero haciendo uso o no del sistema de gestión de recursos propuesto. Podemos ver claramente la mejora que introduce el sistema propuesto respecto a su no utilización. Por citar un ejemplo numérico, se observa que cuando no se hace uso de la gestión de los recursos propuesta tan sólo un 50% (aprox.) de los paquetes del servicio 1 se entregan en menos de 300 ms. Sin embargo, haciendo uso de dicha gestión este porcentaje se eleva por encima del 96%. Y esta mejora se da para todos los servicios, lo que demuestra el buen comportamiento del esquema global.

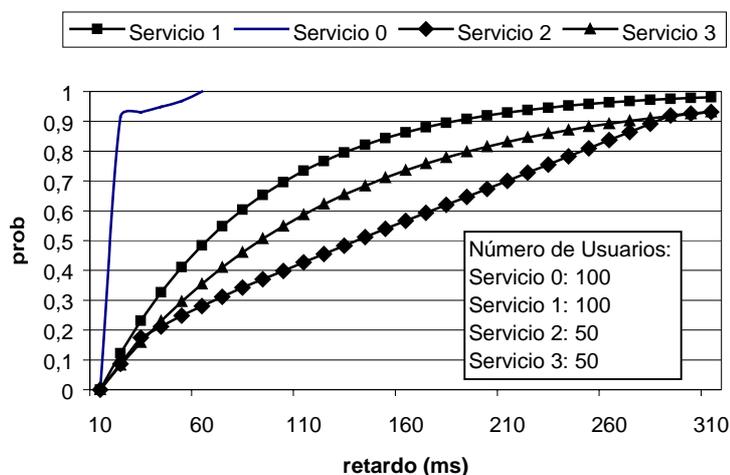


Figura 144. PDF Retardo paquetes (CON scheduling)

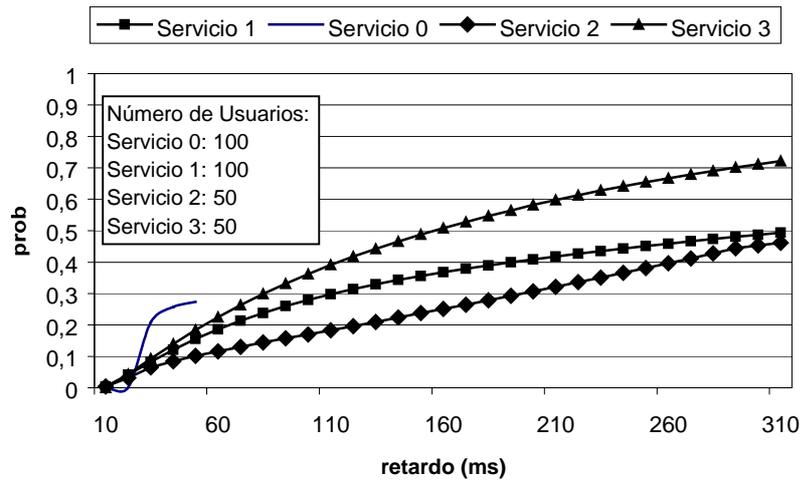


Figura 145. PDF Retardo paquetes (SIN scheduling)

Las siguientes figuras (Figura 146 y Figura 147) permiten observar otro de los aspectos importantes del rendimiento de la gestión de recursos propuesta. Para el cálculo de estas curvas se ha mantenido el número de usuarios de tres de los servicios en valores de carga relativamente elevados (75 usuarios de los servicios 1 y 2, y 50 usuarios del servicio 3) y se ha variado la carga de usuarios de voz (servicio 0). En la Figura 146 se representa la evolución del porcentaje de paquetes perdidos de este servicio en función de la carga de usuarios del mismo. Podemos comprobar que este porcentaje de pérdidas se estabiliza en torno al 0'7%, por debajo del umbral del 1% definido. Esto se debe a que el criterio de prioridades definido en el sistema va otorgando los recursos necesarios a las transmisiones de este servicio para que pueda seguirse manteniendo la calidad requerida. El precio que se paga es empeorar el retardo de los servicios menos prioritarios. En particular, podemos comprobar este hecho en la Figura 147, donde podemos observar la evolución del retardo de los paquetes del servicio *best-effort* (servicio 3) para las mismas condiciones de tráfico de la Figura 146.

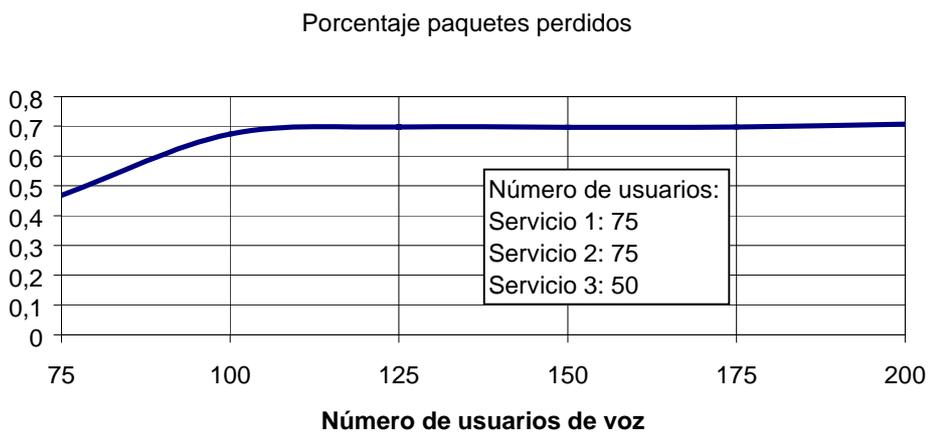


Figura 146. Porcentaje de paquetes perdidos del servicio de voz

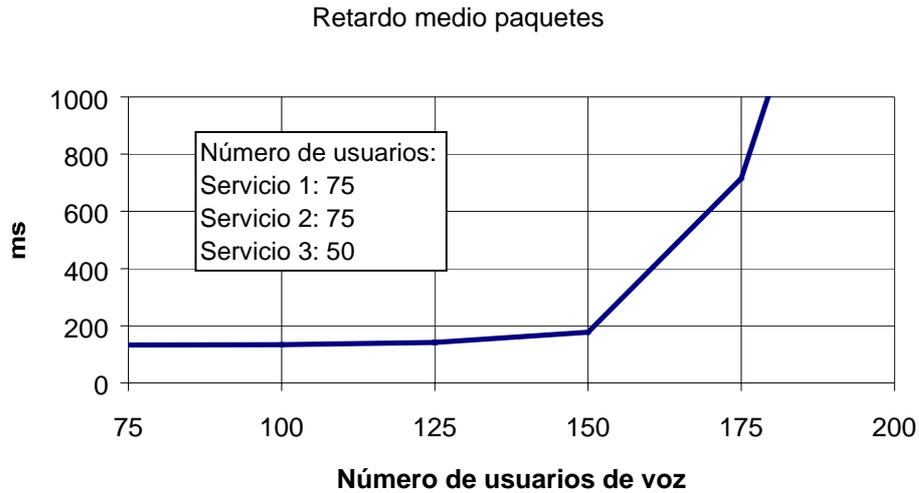


Figura 147. Retardo de los paquetes de datos *best-effort*.

Se comprueba de forma explícita que el mantenimiento de la calidad de las transmisiones de voz se consigue a costa de que, en el momento de sobrepasar un umbral aproximado de 150 usuarios, los recursos disponibles para las transmisiones *best-effort* se reducen hasta el punto de disparar el retardo de los paquetes de este servicio.

Debe recalcar que, en todos los casos, el mecanismo de gestión de recursos debe limitar la interferencia total presente en el sistema. De este modo se preserva la calidad de todas las transmisiones. En concreto, este proceso se realiza en el punto 4 del algoritmo descrito en el apartado 6.2.4.2.

6.3 Conclusiones

Se ha realizado una propuesta de gestión dinámica de recursos para poder garantizar una cierta calidad de servicio a un conjunto de Clases de Servicio en un sistema de transmisión por paquetes para entornos móviles. La propuesta utiliza como protocolo de acceso MAC el protocolo DQRAP/CDMA, y define unos mecanismos de gestión distribuida tanto de los recursos de peticiones de acceso como de la transmisión de datos. Se ha estudiado su comportamiento mediante simulaciones por ordenador y se ha comprobado la bondad de su funcionamiento y el cumplimiento de los objetivos marcados. Sus virtudes más importantes son la sencillez de implementación, la reducción del tráfico de señalización y de control necesarios para su funcionamiento, la facilidad de establecer criterios variables de prioridad, así como los buenos resultados obtenidos en cuanto al aprovechamiento de los recursos del sistema.

Por todo ello, concluimos que puede ser una buena propuesta a tener en cuenta para su puesta en práctica en los futuros sistemas de tercera generación.

