

Gestión Distribuida con Calidad de Servicio Garantizada para Sistemas de Transmisión por Paquetes en Entornos CDMA.

Luis Alonso Zárate, Ramón Agustí Comes, Oriol Sallent Roig
Grupo de Comunicaciones Radio

Departamento de Teoría de la Señal y Comunicaciones
Universitat Politècnica de Catalunya, UPC

c) Jordi Girona, 1-3, Campus Nord, Edifici D4, Barcelona.-08034

e-mail: [luisg, ramon, oriol]@tsc.upc.es

Teléfono: 93 401 71 95, Fax: 93 401 72 00

Área Temática 1: Sistemas y Tecnologías de Radiocomunicaciones

Gestión Distribuida con Calidad de Servicio Garantizada para Sistemas de Transmisión por Paquetes en Entornos CDMA.

Luis Alonso Zárate, Ramón Agustí Comes, Oriol Sallent Roig
Grupo de Comunicaciones Radio
Departamento de Teoría de la Señal y Comunicaciones
Universitat Politècnica de Catalunya, UPC
c) Jordi Girona, 1-3, Campus Nord, Edifici D4, Barcelona.-08034
e-mail: [luisg, ramon, oriol]@tsc.upc.es; Teléfono: 93 401 71 95, Fax: 93 401 72 00

Resumen: En esta comunicación se presenta una estrategia de gestión distribuida para poder garantizar una cierta calidad de servicio en conexiones cuya transmisión se realiza en modo paquete. El entorno de trabajo son los sistemas de comunicaciones móviles multimedia con acceso múltiple CDMA. El esquema global se basa en un protocolo de acceso al medio del tipo DQRAP/CDMA y en una gestión de recursos basada a su vez en un criterio de Crédito pendiente de Servicio y un mecanismo de prioridades por servicio. Se muestra el modelo del sistema y los resultados obtenidos mediante simulaciones por ordenador de cara a validar el esquema propuesto.

1.- INTRODUCCIÓN

El presente trabajo se enmarca dentro del estudio de técnicas de gestión eficiente que permiten garantizar una cierta calidad de servicio en sistemas de transmisión por paquetes basados en la técnica de acceso DS-CDMA (*Direct Sequence–Code Division Multiple Access*). Con toda seguridad, esta estrategia constituirá la base de los futuros sistemas móviles de tercera generación, puesto que es la escogida en la propuesta de UTRA del 3GPP [1] para el acceso radio en UMTS [2].

La creciente demanda registrada en los últimos años en lo que a servicios multimedia se refiere, particularmente en todo lo que envuelve al mundo de Internet, precisará de sistemas flexibles para integrar fuentes de tráfico de diferente naturaleza. Dentro de este contexto, y para el manejo de tráfico a ráfagas, las estrategias de conmutación de paquetes sobre canal común se presentan como más adecuadas frente a otras estrategias de transmisión por canal dedicado. En las primeras, los diferentes usuarios pueden compartir dicho canal común gracias a la multiplexación estadística originada en las propias fuentes, sin requerir de una asignación previa de recursos. Por contra, en las estrategias de canal dedicado se precisa de un cierto tiempo inicial de reserva del canal, que en ocasiones es comparable al tiempo en que dicho canal va a ser utilizado posteriormente. Por otro lado, para una gestión adecuada de los recursos compartidos (códigos y niveles de potencia en el entorno CDMA), es necesario el uso de protocolos de acceso al medio que garanticen una utilización equitativa y eficiente de los mismos. Finalmente, una vez se han establecido los mecanismos de acceso, los paquetes de información multimedia deben ser gestionados y transmitidos de forma inteligente para que las diferentes aplicaciones soportadas puedan cumplir con sus requerimientos de calidad de servicio. Estas técnicas de gestión o *scheduling* constituyen un puntal básico para las futuras comunicaciones móviles multimedia de tercera generación.

En particular, en esta comunicación nos planteamos el estudio de una técnica de gestión en modo distribuido que ofrece una eficiencia superior a los esquemas clásicos de prioridades. Es una técnica basada en el concepto de crédito pendiente de servicio [3] y en el protocolo de acceso al medio DQRAP/CDMA [4]-[5], del que se aprovechan algunas de sus ventajas características de cara a la precisa estimación del tráfico instantáneo de cada clase de servicio. El sistema a estudiar consiste en una estación base que se comunica con un número elevado de usuarios mediante un canal de transmisión común con acceso CDMA en modo paquete. Cada uno de estos usuarios mantiene un cierto número de conexiones activas con diferentes requerimientos de calidad de servicio. Se definen unas clases de servicio (conjunto de parámetros de calidad de servicio) a las que se adaptan las conexiones activas. Por su parte, la estación base transmite a los usuarios, mediante canales de control, la cantidad mínima de información necesaria para que éstos sepan cuándo y con qué recursos deben realizar sus transmisiones en modo paquete. El protocolo de acceso al medio escogido permite a todos los

móviles conocer en todo momento el número de conexiones activas de cada clase de servicio y ello permite a su vez optimizar la señalización necesaria para establecer los mecanismos de gestión de los recursos radio. El uso de colas distribuidas consigue minimizar el número de colisiones en los intentos de acceso y eliminar las colisiones en la transmisión de datos, lo que mejora significativamente la eficiencia del sistema. Por otro lado, el concepto de crédito de servicio permite adaptar la gestión de los recursos radio a las condiciones cambiantes y a menudo hostiles e impredecibles de un canal móvil, con lo que también se mejora el rendimiento general, lo que redundará en una mayor capacidad del sistema.

Esta ponencia se estructura del siguiente modo. En el punto 2 se define el concepto de Clase de Servicio que será usado en el sistema. En el punto 3 se detalla la estructura de la propuesta realizada. El punto 4 muestra los resultados obtenidos de las simulaciones de validación. Finalmente, el punto 5 está dedicado a las conclusiones.

2.- CLASES DE SERVICIO

Definimos Clase de Servicio (CS) como el conjunto de parámetros de calidad de servicio que delimitan la transmisión de un cierto flujo de información. Asimismo, cada CS incluye también los valores que limitan estos parámetros de calidad. En principio, cada aplicación tendrá asociada una o más CS. Cada uno de los flujos de información que genera y que deben ser transmitidos por la aplicación tendrá asignada su CS. 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.

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 y decidir cuál de las CS definidas en el sistema resulta más adecuada a sus necesidades y se ajusta más a su patrón de tráfico. En este sentido, el uso de conformadores de tráfico (*traffic shaping*) [7] para ‘suavizar’ el tráfico ofrecido al sistema es una técnica frecuentemente utilizada en los sistemas actuales.

3.- DESCRIPCIÓN DEL SISTEMA

3.1.- Descripción del mecanismo de acceso

La propuesta de nuestro estudio se basa en el protocolo de acceso al medio DQRAP/CDMA. Algunas de las características de este protocolo lo hacen especialmente indicado para dar servicio a aplicaciones que generan tráfico heterogéneo multimedia y que requieren de una cierta calidad garantizada. En efecto, el sistema de acceso basado en este protocolo consiste en unas colas distribuidas que gestionan tanto las solicitudes de recursos como las transmisiones de los paquetes de datos. Este hecho hace que sea posible, definiendo una gestión adecuada de las colas, conseguir directamente establecer unos criterios de prioridad necesarios para ofrecer la garantía de calidad requerida.

DQRAP/CDMA consiste en dos subsistemas que funcionan en paralelo: el de resolución de colisiones y el de transmisión de datos. Cuando un usuario tiene datos para transmitir, utiliza unos minislots de control para solicitar su transmisión. Si su petición colisiona con la de otro usuario, aquél se sitúa (a nivel lógico) en una cola distribuida de resolución de colisiones. Siguiendo la política de esta cola, en un momento dado le llegará su turno para tratar de resolver esta colisión. Cuando esto ocurre y la petición tiene éxito, el usuario pasa a la cola de transmisión de datos, en la que esperará su turno para realizar dicha transmisión, pero ya sin

contención y sin posibilidad de colisión. Una explicación más detallada de este protocolo puede verse en [4].

La adaptación del protocolo propuesto consiste en crear tantas colas distribuidas como tipos o clases de servicio, tanto en el subsistema de resolución de colisiones como en el de transmisión de datos, y asignar cada una de ellas a su correspondiente servicio. La Figura 1 muestra el esquema genérico del sistema de acceso para el caso de tener el protocolo DQRAP/CDMA dando servicio a un número indeterminado de fuentes de tráfico con distintas QoS.

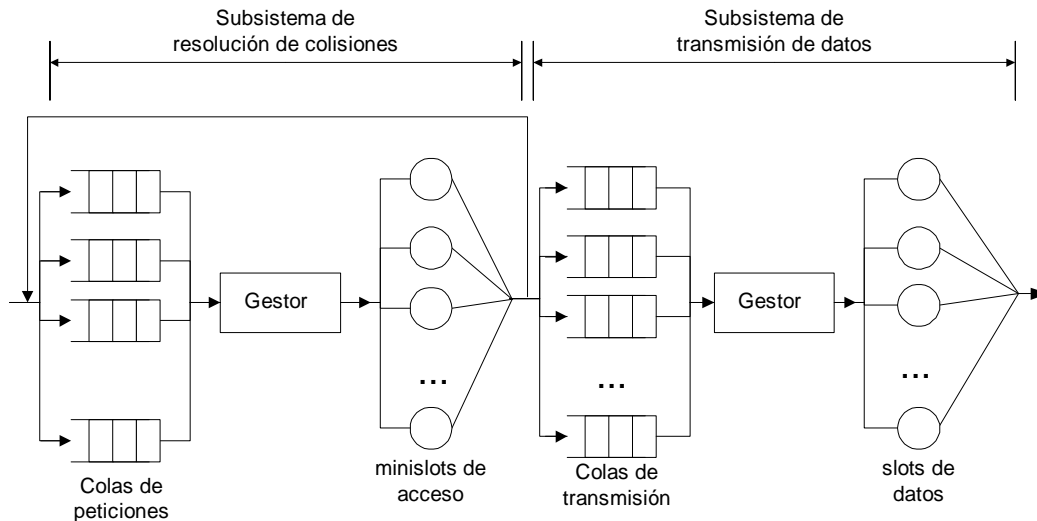


Figura 1. Modelo de gestión del sistema

Podemos observar los dos subsistemas descritos para el protocolo DQRAP/CDMA. La línea de realimentación en el subsistema de resolución de colisiones indica el flujo de las peticiones que colisionan y deben entrar de nuevo en la cola correspondiente.

Cada una de las CS presentes en el sistema tendrá asignada una de las colas de peticiones y una de las colas de transmisión. Por otro lado, 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 a minislots de control o a slots de datos (según el subsistema). Por ejemplo, cada servidor puede representar 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 es la de asignar uno o más de estos servidores (los recursos disponibles) a los elementos presentes en las colas, de tal modo que puedan cumplirse los requerimientos de calidad de cada una de las CS.

3.2.- Concepto de gestión distribuida

Uno de los objetivos de cualquier sistema de comunicaciones móviles es minimizar la cantidad de información de señalización que debe circular entre la estación base y los terminales móviles, al mismo tiempo que se maximiza la eficiencia en la utilización de los recursos radio.

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 definido para el acceso y va otorgando los correspondientes permisos para transmitir en el momento adecuado a cada usuario. Sin embargo, la asignación dinámica de recursos para las peticiones de acceso de cada clase de servicio puede resultar muy compleja. Asimismo, el tráfico de señalización necesario para realizar las notificaciones explícitas de cada asignación puede llegar a tener un volumen considerable.

Haciendo uso de DQRAP/CDMA podemos introducir el concepto de gestión distribuida o *scheduling* distribuido. Para el funcionamiento del protocolo, incluso en su versión más sencilla, la estación base tiene que enviar a los usuarios por el canal descendente el resultado de lo ocurrido en los minislots de acceso. Con esta información, los terminales conocen el estado global del sistema y pueden calcular su posición en las colas distribuidas. De ese modo, conocen 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.

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 pueden aplicar la misma filosofía de funcionamiento. Esto es, cada uno de los móviles ejecuta el algoritmo de gestión definido para el sistema, actualiza el estado de sus copias locales de las colas, y decide si debe transmitir o no, y en su caso con qué recursos. Se mantiene el criterio de que son los móviles de forma distribuida los que deciden cuándo y cómo realizar la transmisión de los datos, eliminando la necesidad de enviar mensajes de señalización para este cometido. En cierto modo, ambos gestores mostrados en la Figura 1 están lógicamente distribuidos entre todos los terminales de usuario presentes en el sistema.

3.3.- Descripción del mecanismo de gestión

Podemos encontrar en la literatura algunas propuestas de gestión que se apoyan en un concepto conocido como Crédito de Servicio o Servicio Pendiente. Todas ellas están basadas en la idea usada en los sistemas ATM conocida como 'leaky bucket' [6]. Este mecanismo consiste en asignarle a cada conexión, de forma regular en el tiempo, unos testigos o *tokens*, y quitárselos después a medida que va recibiendo servicio del sistema. La tasa de generación de testigos para cada conexión depende de su tasa de transmisión. De este modo, una conexión sólo podrá recibir servicio cuando tenga testigos disponibles, es decir, cuando el sistema le 'deba' servicio.

Nuestro estudio propone una extensión de esta idea a un entorno de comunicaciones móviles con canal radio, donde podemos redefinir los testigos (unidades discretas) y utilizar en su lugar el concepto de Crédito de Servicio (CrS), o Crédito Disponible de Transmisión. Este valor es un número real, no necesariamente entero, que mide la diferencia entre el servicio que un usuario ha 'contratado' con el sistema, y el servicio real que éste le ha ofrecido. En cada unidad de tiempo transcurrido, el valor de CrS de cada conexión es incrementado en la cantidad proporcional de servicio que debería haber recibido la conexión, normalizada a la unidad. Cada vez que el sistema le asigna servicio de forma exitosa a dicha conexión, su valor de CrS se decrementa en una unidad, indicando que el sistema ha saldado parte de su 'deuda'. De este modo, valores positivos de CrS suponen que una conexión ha recibido hasta ese momento menos servicio del sistema del que pactó inicialmente. Cuanto mayor sea CrS, tanto más servicio debe darle el sistema a la conexión para ponerse 'al día'. Por otro lado, un valor negativo de CrS supondría 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, y darles prioridad en la asignación a aquellas que tengan un CrS mayor. También es posible establecer un criterio de prioridad entre las diferentes CS, de modo que el reparto de recursos no sigan únicamente un orden creciente de CrS.

Existen, sin embargo, situaciones que deben tratarse con cuidado cuando se hace uso de una estrategia de gestión 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 vuelve 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'. Existen al menos dos estrategias diferentes para abordar este problema, evitando los perjuicios que de él pueden derivarse:

1. 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.

2. Establecer un máximo de recursos asignables por el sistema de forma simultánea para una conexión.

Es importante reseñar que para sistemas de transmisión por radio, especialmente en entornos móviles, sólo se deberá reducir el valor del CrS cuando se haya recibido con éxito la notificación de que se ha servido correctamente a una conexión, no sencillamente cuando se realiza la transmisión de un paquete de datos de la misma.

3.3.- Algoritmo de asignación de recursos

Una vez definido el concepto de CrS y su método de cálculo, debemos detallar de qué manera el sistema asigna recursos a las diferentes CS en función del CrS. En principio, cada CS tiene asignada una cierta tasa de transmisión que llamaremos básica. Si los recursos disponibles lo permiten, es posible que resulte conveniente transmitir la información de una cierta conexión a una tasa mayor que la básica que tiene asignada. Ello será posible siempre que la calidad de las conexiones en curso no se vea degradada. En el sistema estudiado, para realizar una variación de la tasa de transmisión de las conexiones, se ha hecho uso de la técnica de transmisión multicódigo, que es totalmente equivalente a utilizar una variación de la ganancia de procesado.

El algoritmo de asignación de recursos sigue los siguientes pasos:

1. Se ordenan las conexiones activas según su CrS, de mayor a menor, teniendo en cuenta la prioridad que puedan tener las CS a las que pertenecen.
2. Se realiza una asignación temporal de recursos a la conexión con mayor CrS.
3. Se verifica que no se degrada por debajo de lo admisible la calidad de las conexiones con recursos ya asignados. En caso afirmativo se fija la asignación de recursos para la conexión. En caso negativo se finaliza el proceso de asignación de recursos.
4. Se calcula el valor de CrS que tendría la conexión en el caso de que la transmisión futura tuviera éxito.
5. Se reordenan las conexiones según su CrS haciendo uso del valor virtual calculado en el punto 4 para las conexiones que ya tienen recursos asignados.
6. Se vuelve al punto 1.

De este modo, si una conexión acumula suficiente prioridad como para que se le puedan asignar más recursos que los que le corresponderían según su tasa básica, el algoritmo presentado le dará todos aquellos recursos que sea posible asignarle. De cara a la evaluación de la interferencia multiusuario del acceso CDMA, cuando a un usuario se le asignen n veces los recursos básicos de su conexión, se considerará el caso equivalente de tener n usuarios activos de dicha conexión.

4.- Evaluación de las prestaciones

Se han obtenido algunos resultados del rendimiento del sistema mediante simulaciones por ordenador, de cara a verificar la bondad de las propuestas realizadas. El sistema simulado consiste en una célula con una estación base que da servicio a un conjunto de usuarios que utilizan un control de potencia en lazo cerrado ideal. En el acceso radio CDMA, la interferencia multiusuario para el cálculo del BER se ha modelado haciendo uso de la hipótesis gaussiana [4]. El canal está dividido en tramas de 10 ms en las que se transmite cada uno de los paquete de datos. Se han utilizado 3 minislots de control por trama para el acceso DQRAP/CDMA y las secuencias de petición de acceso son de 256 chips. El receptor de dichas secuencias es el descrito en [4].

Las Figuras 2 y 3 muestran una comparativa entre las prestaciones del sistema sin ningún tipo de gestión y las obtenidas haciendo uso de la propuesta del estudio. En estas gráficas se presentan las funciones de distribución (PDF) del retardo de los paquetes para los diferentes servicios. Se han establecido cuatro clases de servicio diferentes, cada una con sus características de transmisión y requerimientos de calidad. Para el caso con *scheduling*, estas clases de servicio han sido ordenadas según lo restrictivo de dichos requerimientos, en orden descendente. De este modo, existe un cierto orden implícito de prioridad a la hora de asignar los recursos del sistema. Básicamente, las cuatro clases de servicio usadas tratan de representar,

respectivamente, usuarios de voz, datos interactivos, datos no interactivos (IP) con retardo máximo, y conexiones *best-effort*. Las características del tráfico de cada tipo de usuario se describen en la Tabla 1. El número de usuarios activos de cada clase de servicio se especifica en las gráficas citadas.

Tabla 1

	Servicio 0 (voz)	Servicio 1 (datos interactivos)	Servicio 2 (datos no interactivos, IP)	Servicio 3 (<i>best-effort</i>)
Modelo de generación tráfico	ON-OFF (actividad 44% por usuario)	MMPP (act. 50%, $\lambda=0,02$ y $1/\mu=1000$ bits por usuario)	IEEE 802.14	Poisson ($\lambda=0,01$ y $1/\mu=6000$ bits por usuario)
Tasa media para cálculo CrS (Kbps)	6,6	1	0,5	NA
Tasa de pico (Kbps)	15	30	60	120
Ganancia de procesado	256	128	64	32
Retardo máximo paquetes	50 ms	500 ms	1 s	No
Porcentaje máximo paquetes perdidos	1%	0,1%	0,01%	0%
VER máximo	10^{-3}	10^{-6}	10^{-6}	10^{-6}

En estas gráficas se observa explícitamente como el sistema da prioridad a los servicios según lo restrictivo de sus parámetros de calidad. Así, para el primer ejemplo, se puede comprobar cómo el 90% de los paquetes de información del servicio 1 se entregan en menos de 185 ms, mientras que el mismo porcentaje de paquetes del servicio 2 se entregan en menos de 270 ms. Comparando ambas gráficas podemos observar una muestra de la ganancia que introduce el sistema de gestión de cara a mejorar las prestaciones y eficiencia del sistema, sin cambiar ninguno de los parámetros físicos del mismo.

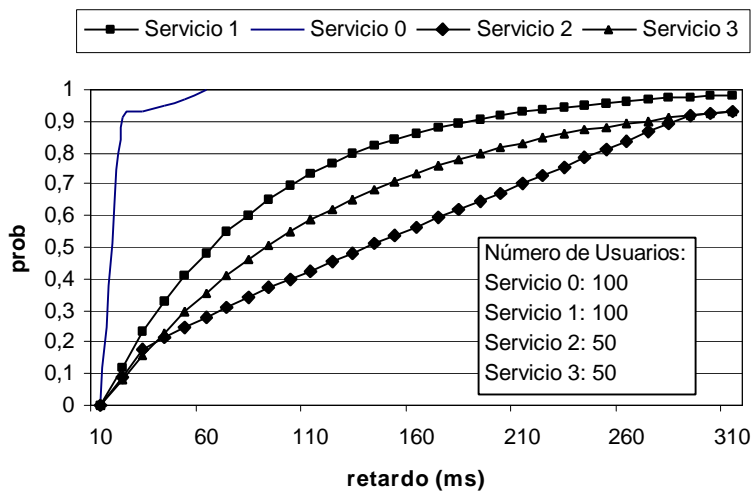


Figura 2. PDF Retardo paquetes (CON scheduling)

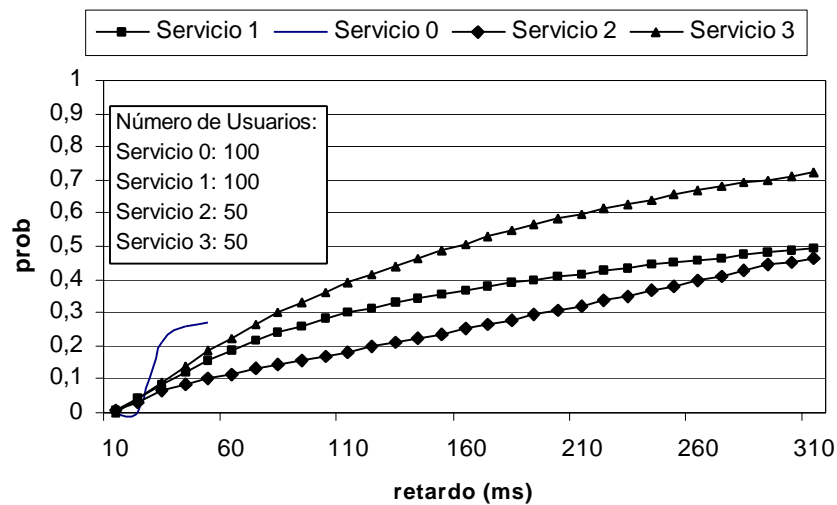


Figura 3. PDF Retardo paquetes (SIN scheduling)

5.- Conclusiones

Se ha propuesto y analizado un sistema de gestión de transmisión por paquetes con calidad garantizada, basado en el protocolo de acceso DQRAP/CDMA y en el concepto de Crédito de Servicio. Se han realizado simulaciones por ordenador y los resultados obtenidos resultan prometedores de cara a considerar el esquema propuesto como un candidato a ser utilizado en los futuros sistemas de comunicaciones móviles de tercera generación, gracias a su combinación de simplicidad y eficiencia.

Referencias

- [1] <http://www.3gpp.org>
- [2] Erik Dahlman, Per Beming, Jens Knutsson, Fredrik Ovesjö, Magnus Persson, Christiaan Roobol, "WCDMA- The Radio Interface for Future Mobile Multimedia Communications", IEEE Transactions On Vehicular Technology, Vol 47, nº 4, November 1998, pp 1105-1118
- [3] S. Lu, V. Bharghavan and R. Srikant, "Fair Scheduling in wireless packet networks", Proc. ACM SIGCOMM'97, Palais de Festivals, Cannes, France, Sept. 1997.
- [4] Luis G. Alonso, Ramon Agustí, Oriol Sallent, "A Near-Optimum MAC Protocol Based on the Distributed Queueing Random Access Protocol (DQRAP) for a CDMA Mobile Communication System", aceptado para publicación en IEEE Journal on Selected Areas in Communications. Será publicado en el tercer cuarto de 2000.
- [5] Luis G. Alonso, Ramón Agustí, Oriol Sallent, "A Near-Optimum Medium Access Protocol Based on the Distributed Queueing Random Access Protocol (DQRAP) for a CDMA Third Generation Mobile Communication System", PIMRC'99, pp. 1300-1304, Osaka, septiembre 1999.
- [6] Dimitri Bertsekas, Robert Gallager, *Data Networks*. Second Edition. Prentice Hall International Editions. 1992.
- [7] R. Andel, M.N. Huber, S. Schroder. *ATM Networks: Concepts, Protocols, Applications*. Second Edition. Addison-Wesley, 1994.