

Sensibilidad en la elección de la disciplina de servicio sobre el sistema GPRS

J. Bada, F. Casadevall

Grupo de Comunicaciones Radio

Departamento de Teoría de la Señal y Comunicaciones

Universidad Politécnica de Cataluña (UPC)

ferranc@tsc.upc.es

Resumen. - El Servicio General de Paquetes vía Radio (GPRS) es el nuevo sistema de conmutación de paquetes, sobre la infraestructura de GSM, para la transmisión de datos con calidad de servicio. En un sistema como GPRS, garantizar la calidad de servicio al máximo número de usuarios requiere de una eficiente gestión de los escasos recursos del interfaz radio, o lo que es lo mismo, de la selección de una disciplina de servicio óptima. Considerando tráfico heterogéneo, en este artículo se analizan cuatro disciplinas de servicio que basan la gestión de los recursos en la prioridad. Dichas disciplinas son: SPS (Static Priority Scheduling), ARR (Adapted Round Robin), WPQ (Windowed Priority Queuing) y SJN (Shortest Job Next).

Los resultados muestran que las disciplinas que clasifican las tramas LLC en función de su longitud (SJN) o bien que reparten equitativamente los recursos entre las diferentes prioridades (ARR) presentan un mejor comportamiento en el cumplimiento de los atributos de calidad de servicio frente a las que simplemente extraen los paquetes según el orden de llegada (SPS, WPQ).

1. Introducción

En los últimos años, en el mundo de las telecomunicaciones, hay dos sectores que han despuntado respecto al resto. Estos son las comunicaciones móviles y las redes de conmutación de paquetes para datos, especialmente Internet. El primero de ellos ha explotado en la última década convirtiéndose en uno de los servicios estrella en todo el mundo. La culpa de esta revolución se debe a la aparición de los sistemas móviles de segunda generación, como por ejemplo el **GSM** (*Global System For Mobile Communications*) que presenta mejores prestaciones que sus antecesores al utilizar una técnica digital en la codificación de la voz. En segundo lugar, las aplicaciones de datos han experimentado un crecimiento espectacular y cada vez se están convirtiendo en más imprescindibles en todos los ámbitos de la sociedad mundial. Ejemplos de estas aplicaciones son, la navegación por Internet a través de páginas Web o el correo electrónico, entre otras. Para la transferencia de los datos, todas ellas se basan en el protocolo de nivel de red **IP** (*Internet Protocol*). Por todo esto parece lógico que en un futuro cercano estas dos áreas tiendan a converger, de manera que se incremente la demanda de acceso a aplicaciones de Internet a través de sistemas móviles. En consecuencia, y como primer paso imprescindible de la convergencia, será necesario el dotar a las redes móviles actuales, diseñadas principalmente para dar soporte al tráfico de voz, de los mecanismos necesarios para adaptarlas a los servicios o aplicaciones de datos. Por todo ello, y basado en el estándar GSM, se ha desarrollado un nuevo sistema que incorpora servicios en modo paquete y que se conoce como **GPRS** (*General Packet Radio Service*).

La principal característica de este nuevo servicio, GPRS, es la introducción de la conmutación de paquetes dentro de la propia red GSM, proporcionando los mecanismos necesarios para la interconexión con redes externas de datos¹, y el disfrute de velocidades teóricas de transferencia de hasta aproximadamente 171 Kbps, debido a la utilización de múltiples ranuras temporales ó timeslots por usuario sobre una misma portadora y a la introducción de tres nuevos esquemas de codificación (CS-2 a CS-4) además del propio utilizado por GSM (CS-1). Con la inserción de la transmisión en modo paquete extremo a extremo, GPRS es un sistema de transmisión apropiado para aplicaciones que generan tráfico a ráfagas como pueden ser; WWW (WAP), E-mail, Telnet, FTP entre otras. Además, con vistas al futuro usuario móvil, el sistema introduce dos nuevos aspectos muy destacables a tomar en consideración; la tarificación, basada en el volumen de datos intercambiados, y la posibilidad de estar conectado permanentemente, funcionalidad conocida como "Always on". Esto último significa que el terminal móvil (**MS**) permanece a la escucha permanente del sistema por si le llegan nuevas

¹ Estas redes de datos serán, básicamente, Internet o Intranets aunque también se especifica la interconexión con redes X.25.

transferencias, sin tener que realizar una nueva conexión. Además de las anteriores, una de las características que merecen una atención especial dentro de las redes GPRS es la introducción del concepto de calidad de servicio (**QoS**) con atributos tales como la prioridad, fiabilidad, retardo y velocidad de transferencia media y de pico. Como muestra de estos parámetros, en la siguiente tabla se muestra los valores estandarizados para el retardo medio de acceso y el percentil al 95% dentro del segmento GPRS de una conexión móvil de datos. Una descripción más detallada del servicio GPRS se puede encontrar en [1], [2] y [3].

Clases de retardo	Retardo			
	Tamaño SDU: 128 octetos		Tamaño SDU: 1024 octetos	
	Retardo medio transferencia (s)	Retardo 95% (s)	Retardo medio transferencia (s)	Retardo 95% (s)
1 (predictive)	<0.5	<1.5	<2	<7
2(predictive)	<5	<25	<15	<75
3(predictive)	<50	<250	<75	<375
4(best effort)	Sin especificar			

Tabla 1. Retardo para la red GPRS

En la tecnología GPRS, al tratarse de una red de conmutación de paquetes, es imprescindible la elección de una estrategia óptima de gestión de los recursos radio², **RRM** (*Radio Resource Management*) con el objetivo de :

- Gestionar de la forma más eficiente posible los escasos recursos disponibles en la interfaz radio entre los diferentes usuarios y servicios.
- GPRS es una red que introduce el concepto de calidad de servicio que debe garantizarse, siendo el segmento más restrictivo la interfaz radio.

Debido, entre otras, a estas dos razones en esta ponencia se presenta un estudio comparativo de diversas disciplinas de servicio, considerando varios escenarios y un tráfico heterogéneo, que es el que se espera en el nuevo sistema. Las tipologías de tráfico consideradas en el estudio son; correo electrónico, navegación por Internet (WWW), aplicaciones para gestión de flotas (Mobitex) y el servicio de mensajes cortos (SMS).

La organización del presente trabajo es la siguiente. En primer lugar, se presenta una descripción específica de las características más relevantes de las cuatro disciplinas de servicio analizadas. En el apartado tercero, se presentan los resultados obtenidos considerando diferentes escenarios para las cuatro disciplinas de servicio presentadas anteriormente. Finalmente, en el último punto se muestran las conclusiones alcanzadas.

2. Disciplinas de servicio

En los sistema vía radio está claro que existe una limitación en la disponibilidad de recursos radio. Para la tecnología GPRS que trabaja en un entorno multiservicio si se quiere garantizar los diferentes perfiles de calidad de servicio y optimizar la capacidad del sistema, es imprescindible la utilización de estrategias de gestión de recursos radio eficientes. Estas estrategias no se encuentran definidas dentro de las especificaciones del sistema, siendo por tanto decisión de los fabricantes la implementación de la política de gestión más óptima. El elemento clave en los algoritmos de RRM es la disciplina de servicio encargada de determinar el orden en el cual los usuarios de un mismo canal tienen permiso para acceder al mismo.

En la literatura se describen una infinidad de disciplinas de servicio (scheduling) como por ejemplo; First-In-First-Out (FIFO), Static Priority Scheduling (SPS), Virtual Clok (VC), Weighted Fair Queuing (WFQ), Start-Time Fair Queuing (STFQ), Worst-case Fair Weighted Fair Queuing (WF²Q), Earliest Deadline First (EDF), Stop-and-Go, Weighted Round Robin (WRR), Deficit Round Robin (DRR), etc., que han sido evaluados en redes fijas [4]. No obstante, sólo algunas de ellas cumplen con los requisitos básicos para que, con las consiguientes adaptaciones, puedan reutilizarse en entornos GPRS.

Las disciplinas presentadas en este artículo son las siguientes; Static Priority Scheduling (**SPS**) [5], Adapted Round Robin (**ARR**), Windowed Priority Queuing (**WPQ**) y Shortest Job Next (**SJN**) [6]. En

² Debe recordarse que el servicio GPRS comparte los recursos de interfaz radio con el sistema GSM

todas ellas a cada uno de los usuarios activos se les asocia un nivel de prioridad. Las disciplinas sirven los paquetes (tramas LLC) en concordancia con la prioridad asociada siguiendo unas reglas determinadas que diferencian y caracterizan a cada una de ellas. Debe recordarse que en el sistema GPRS se ha definido tres tipos de prioridades: *alta*, *media* y *baja*.

El trato con que se gestionan los mensajes de señalización y retransmisiones es idéntico para todas las disciplinas dándoles la máxima prioridad. Es decir, todos estos mensajes se sitúan en la cola de máxima prioridad a la cabeza de la misma (señalización) o por delante de los radio bloques a transmitir (retransmisiones) y son los primeros en recibir servicio. En los siguientes sub-apartados se presenta una descripción de cada una de ellas.

A. Static Priority Scheduling (SPS)

La disciplina SPS también conocida como Priority Queuing (**PR**) trabaja en colas separadas por prioridad. Esta estrategia se ha estructurado en tres colas; cola de prioridad alta, de prioridad media y prioridad baja. Los datos de usuario, y en función de la prioridad asociada en el momento de negociar el perfil de calidad de servicio, se sitúan en una de ellas. Dentro de cada una de éstas colas el orden de servicio vendrá marcado en función del orden de llegada. Como puede verse en este caso se produce una agregación de los usuarios en función de la prioridad. Esto en principio es óptimo, en el sentido que reduce la complejidad en la implementación de la disciplina. Ahora bien, como se verá posteriormente esta disciplina es poco adecuada en un entorno móvil como es GPRS. El criterio principal escogido para servir los paquetes es empezar siempre por las colas de más alta prioridad y proseguir sucesivamente con el resto de colas de prioridad inferior, siempre y cuando las inmediatamente superiores se encuentren vacías, es decir se hayan servido todas las tramas LLC. El orden de servicio dentro de una misma cola como se ha comentado es FCFS (First Come First Served).

Para ilustrar como trabaja esta disciplina en la Figura 1 se muestra un ejemplo en donde la cola A representa la cola de mayor prioridad mientras que la cola C es la de menor prioridad. Supóngase que en un instante determinado de tiempo en la cola A hay un total de 9 tramas pendientes de ser servidas, 4 en la cola B y 3 en la cola C. Atendiendo a las características de esta disciplina el sistema sirve en primer lugar las 9 tramas de la cola de alta prioridad, posteriormente, siempre y cuando no entren nuevas tramas en la cola de alta prioridad, sirve las de la cola B y finalmente si en las dos anteriores no hay ninguna trama pendiente de servir procesará las de la cola de menor prioridad.

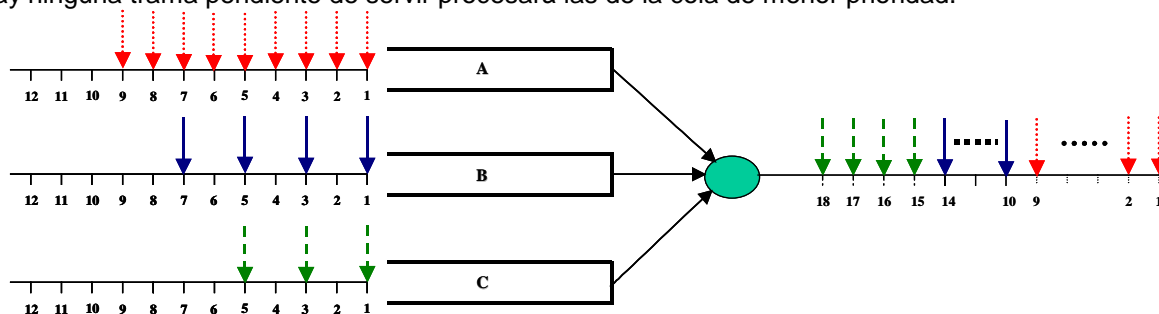


Figura 1. Ejemplo de servicio SPS

Como puede verse en el ejemplo presentado, esta disciplina tiene el problema de la alta dependencia de sus prestaciones con el nivel de tráfico generado por los usuarios de prioridad alta, nivel de tráfico que repercute directamente al resto de prioridades. Esto en principio podría no ser un problema grave, siempre y cuando los paquetes de una cola presenten la misma tipología de tráfico, pero como se ha visto, el tráfico en GPRS no presenta esta propiedad y por tanto, como se verificará en el apartado de resultados, esta estrategia presenta un funcionamiento no demasiado adecuado para este tipo de entornos.

B. Windowed Priority Queuing (WPQ)

Como se ha visto anteriormente con la estrategia SPS existe una fuerte dependencia de sus prestaciones con el tráfico generado por los usuarios de la clase de servicio más restrictiva, lo que puede conllevar que las otras clases lleguen a estar retardadas de forma significativa o incluso a no ser nunca atendidas, fenómeno conocido como "*starvation*". Una posible solución a este grave problema es el algoritmo que se presenta a continuación llamado *Windowed Priority Queuing* (WPQ). En este algoritmo para cada una de las diferentes colas presentes en nuestro sistema se limita el número

máximo de tramas (radio bloques) que pueden ser servidos en cada vuelta. Con ello se consigue que las colas de baja prioridad sean atendidas aunque en las colas de alta prioridad existan tramas esperando el servicio. La forma de elegir el número máximo de radio bloques a servir para cada una de las colas dentro de una vuelta debe estar estrechamente relacionado con los atributos de calidad de servicio (prioridad). Además, debe considerarse que este máximo no tiene por que ser el mismo para cada una de las diferentes prioridades definidas en nuestro sistema. En el caso del sistema GPRS se entiende por una vuelta al conjunto de radio bloques elegidos de las distintas colas que deben servirse en una multitrama. Considerándose que la multitrama en GPRS esta formada por 12 radio bloques a cada una de las prioridades se les asignara un subconjunto de los mismos de manera que siempre se cumpla $\sum RB_{ALTA} + \sum RB_{MEDIA} + \sum RB_{BAJA} \leq 12RB$. El número de radio bloques se mantendrá fijo durante toda la simulación para cada una de las prioridades. Como el número de radio bloques es función de la prioridad, parece lógico que la prioridad alta es la que tenga más RB's asignados por multitrama. Si para alguna prioridad hay en un instante determinado un número inferior de radio bloques al asignado el sistema asignará recursos sobrantes a la siguiente prioridad. En la Figura 2 se muestra un ejemplo de esta estrategia en donde se aprecia que aunque hayan radio bloques pendientes por transmitir en la cola de mayor prioridad (A) el sistema da servicio a las otras colas. Con objeto de facilitar la comparación con la estrategia anterior, la figura 2 presenta la misma situación que en el caso del algoritmo SPS. En este caso la asignación de radio bloques por prioridad es la siguiente:

PRIORIDAD	Nº Radio Bloques
ALTA	5
MEDIA	2
BAJA	1

Aunque la estrategia no es del todo equitativa, si que es cierto que en conjunto se comporta de una manera más optima que el algoritmo SPS, ya que en ningún momento espera la finalización de los paquetes de una cola para servir otra de menor prioridad.

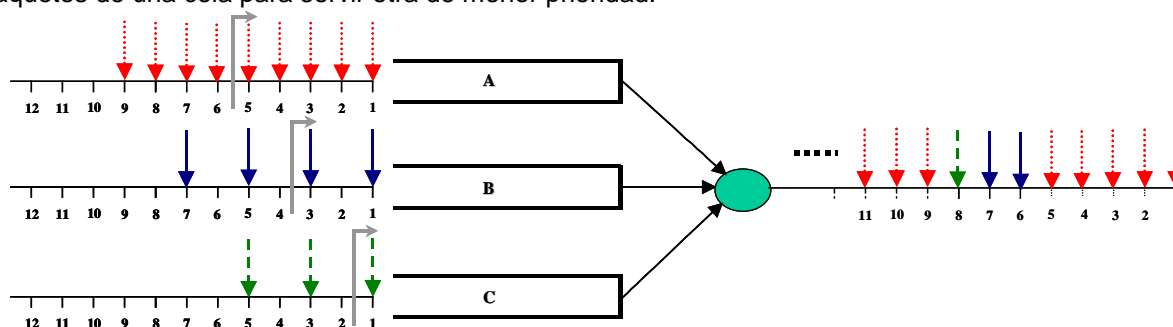


Figura 2. Ejemplo de servicio WPQ

Se puede observar que si, para las diferentes colas, se escoge un numero de radio bloques a servir por vuelta elevado el sistema tiende a comportarse como el algoritmo SPS mientras que si dicho valor se reduce simula una estrategia del tipo Round Robin (RR). En consecuencia, es primordial realizar una elección muy depurada de la cantidad de radio bloques que pueden servirse para cada una de las colas.

C. Shortest Job Next (SJN)

A diferencia de las vistas anteriormente, la estrategia Shortest Job Next (SJN) se basa en la ordenación de los paquetes según la longitud de los mismos o en su defecto en GPRS en las tramas LLC. Por consiguiente, esta estrategia al igual que las anteriores está formada por tres colas para las diferentes prioridades. Las tramas LLC que llegan al sistema se depositan en función de la prioridad en una de ellas. Posteriormente, en función del tamaño de la trama LLC se produce la reorganización de las mismas dentro de cada una de las colas, situándose las de menor longitud a la cabeza para su transmisión. La manera de servir las tramas LLC, una vez clasificadas por su longitud, vendrá dado en función de la prioridad. Es decir, en primer lugar se prestará servicio a los usuarios de prioridad alta; en la situación que dicha cola este vacía se pasa a la siguiente y así sucesivamente. En esta disciplina se puede observar que los instantes de la toma de la decisión no son equidistante sino que dependen de la longitud de las tramas LLC. Si se presentan dos tramas LLC que tienen la misma longitud se seleccionará para servir aquella que haya llegado con anterioridad.

En la siguiente figura se presenta un ejemplo en donde se puede apreciar el funcionamiento de esta disciplina. En ella se puede ver que hay tres colas en función de la prioridad del servicio. En un instante determinado en el sistema se realiza una fotografía instantánea con las tramas LLC que están pendientes de transferir³. Antes de proceder a su transferencia tiene lugar la ordenación de las tramas en función de su longitud. Una vez terminada la ordenación en las diferentes colas se procede a la transferencia empezando en primer lugar por las colas con mayor prioridad.

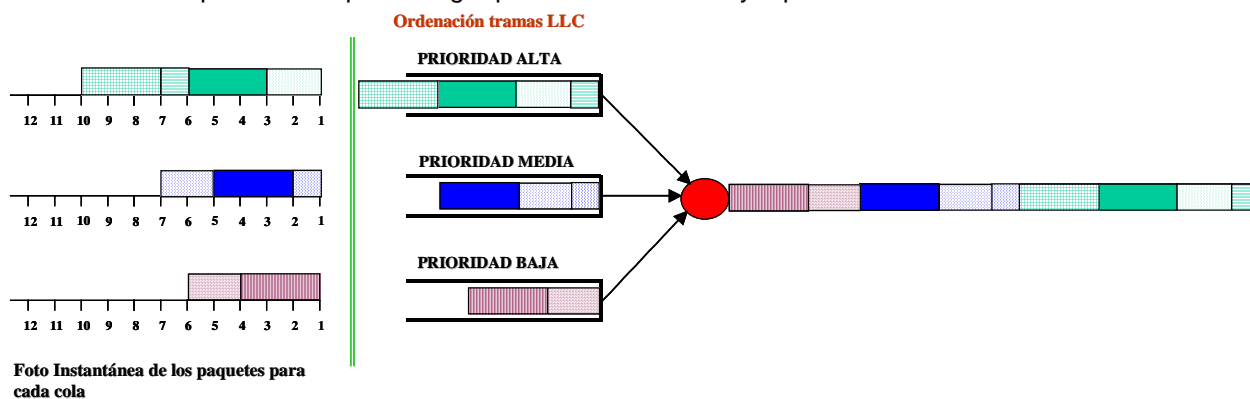


Figura 3. Ejemplo de servicio SJN

Uno de los problemas que se pueden presentar con esta estrategia es que no se sabe cual será la próxima trama LLC a servir ya que depende del siguiente instante en la toma de la decisión y de los instantes de llegada de estas al sistema. Si todos los flujos tienen la misma longitud el sistema se comporta como el algoritmo First Come First Served (FCFS) presentando los mismos problemas que éste.

D. Adapted Round Robin (ARR)

En este apartado se presenta la disciplina *Adapted Round Robin* (ARR) que se basa en una estrategia RR (Round Robin) pero con la clasificación de los usuarios de una misma prioridad. Esta disciplina tiene la capacidad de trabajar con usuarios con tipologías de tráfico diferentes. A diferencia de las disciplinas expuestas anteriormente en este caso no se realiza la agregación de usuarios por prioridad en una sola cola sino que se ha optado por colocar tantas colas como usuarios y además al igual que las anteriores se etiquetan en función de la prioridad. En consecuencia, en esta disciplina habrán dos variables a considerar en la toma de la decisión del siguiente flujo a servir como son la prioridad y el número de usuarios asociados a cada una de ellas. Entonces, las tramas LLC llegan al sistema y en función del usuario y prioridad se emplazan en las mismas. Al utilizar la prioridad para seleccionar a los usuarios que transmitirán por multitrama se debe realizar una pre-asignación de los radio bloques. Esta asignación se realizará de la misma forma que la descrita para la disciplina WPQ. A diferencia de la disciplina WPQ, en este caso los recursos de una prioridad por multitrama se reparten por igual entre todos los usuarios activos. Con ello se consigue realizar una repartición de la forma más justa posible de los recursos entre los diferentes usuarios y las prioridades.

En la Figura 4 se presenta un ejemplo de como se realiza el reparto de recursos entre los diferentes usuarios presentes en un instante determinado para esta disciplina. Para ello cada usuario dispone de su propia cola presentándose 3 usuarios de prioridad alta, 2 de prioridad media y por último 2 de prioridad baja. En la siguiente tabla se refleja por multitrama la asignación de radio bloques por prioridad:

PRIORIDAD	Nº Radio Bloques
ALTA	7
MEDIA	4
BAJA	1

³ Se supone para no complicar el dibujo que en el sistema no entran nuevas tramas LLC hasta terminar con las existentes.

Con estos valores prefijados se observa en dicha figura como se servirían los diferentes radio bloques de las tramas LLC para las multitramas M y M+1 para los usuarios activos en un determinado instante y en función de las prioridades asignadas. Se puede ver, por multitrama, como en primer lugar se ocupan los radio bloques para prioridad alta, seguidos de los de media y por último los de baja.

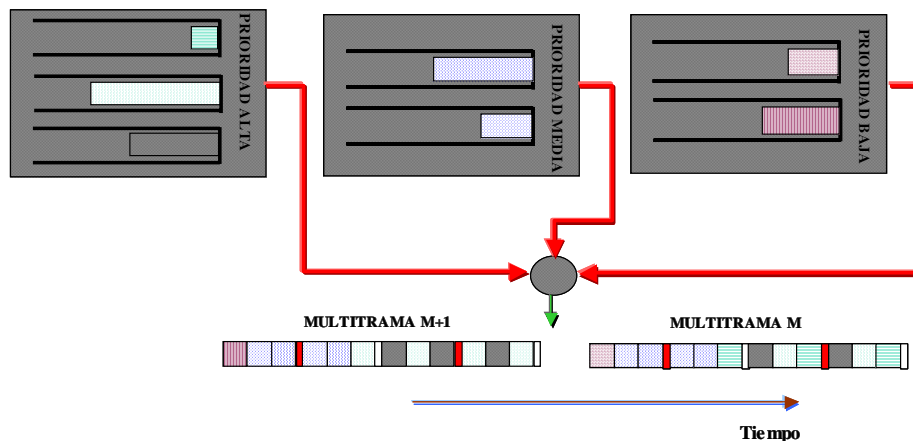


Figura 4. Ejemplo de servicio ARR

3. Análisis de resultados

En esta sección se presentan los resultados de las simulaciones sobre el comportamiento de las diferentes disciplinas de asignación de recursos analizadas en el punto anterior. Para el análisis de los resultados se ha construido un simulador basado en la herramienta **MATLAB** (*Matriz Laboratory*). El simulador implementado soporta una jerarquía de bloques en el modelado del sistema, y en él se han desarrollado todas las capas de la pila de protocolos GPRS relacionadas con la interfaz radio, tanto desde el punto de vista del terminal móvil como del subsistema de estaciones base (BSS), y se ha considerado la posibilidad de incluir varios terminales móviles en ambos sentidos de la transmisión.

En nuestro análisis se considera una única portadora, de la que sólo se utilizan 2 de los Time Slots (TSs) de forma fija⁴ para GPRS mientras que el resto están en disposición del sistema GSM. Todos los móviles se mueven por la celda durante la transferencia según el modelo descrito en [7] y su estado de movilidad inicial es READY, por tanto no es necesario realizar el proceso de búsqueda. Se trabaja a nivel RLC con reconocimientos, con un máximo de 3 retransmisiones para radio bloque erróneo, y a nivel LLC sin reconocimientos. La longitud máxima de las tramas LLC es de 1520 bytes. Se presupone que la disciplina de servicio se encuentra implementada en el nivel RLC/MAC por tanto se trabaja con radio bloques. La transferencia de datos se considera en ambos enlaces aunque principalmente se mostraran resultados para el enlace descendente por ser el que maneja un volumen tanto de tráfico como de usuarios más elevado. En el modelo de canal radio se ha tenido en cuenta el efecto producido por los desvanecimientos lentos, con un entorno celular con celdas trisector. En particular, se ha considerado un canal Typical Urban a 50 Km/h y sin Frequency Hopping (TU50-NoFH).

Los parámetros principales de estudio para poder caracterizar las disciplinas de servicio son los que se enuncian a continuación:

- *Retardo medio*: El retardo es uno de los parámetros más críticos dentro de los atributos que el sistema GPRS utiliza para la descripción de la QoS y por tanto las diferentes disciplinas de servicio se evalúan en función del grado de cumplimiento del mismo. Este se presenta normalizado según se expone en {[8], [9]} debido a la gran variabilidad en el tamaño de los paquetes de las aplicaciones analizadas y para las diferentes calidades de servicio. Uno de los variables interesantes de medir en cuanto al retardo medio es la desviación que presenta con respecto a la media.

⁴ Se supone que estos 2 TSs son canales esclavos y se utiliza uno más para la señalización (TS-0)

- Velocidad media de transferencia por paquete: Para cada aplicación, indica la velocidad con la cual el sistema realiza las descargas de información solicitadas o transferencias de mensajes hacia o desde un servidor para cada uno de los paquetes de las diferentes aplicaciones que forman parte del tiempo de simulación. Para su calculo se determina, para cada uno de los paquetes generados durante el transcurso de la simulación, la velocidad de transmisión. Posteriormente se realiza el promediado de las muestras obtenidas. Este parámetro se expresa en bits por segundo (bps). El tiempo de transmisión toma en consideración el tiempo transcurrido desde el inicio de la generación del paquete hasta la recepción de los reconocimientos ACK por parte del sistema receptor a nivel RLC/MAC. Por tanto, se toma en consideración tanto el tiempo de permanencia en cola como el tiempo utilizado para la transmisión de la información en la interfaz radio.

$$\text{Velocidad media de transferencia} = \frac{\sum_{i=1}^N P_i}{N} / \text{Tiempo transmisión}_i \text{ (bits / seg)}$$

- Tiempo de espera transmisión del primer RB: Es el tiempo que transcurre desde que el paquete llega a la unidad controladora de paquetes (**PCU**, Packet Control Unit), con lo cual ya dispone de recursos asignados para la transferencia, hasta que se empieza a enviar el primer radio bloque de dicho paquete. Este es un parámetro que será de utilidad en la comparación de las disciplinas de servicio. Se expresa en mseg.
- Región de planificación (scheduling): La región de planificación es al mismo tiempo una técnica que permite un control de admisión eficiente y una manera de medir el comportamiento del sistema. Es conocido que una disciplina de servicio debe reservar recursos para garantizar los requisitos de las conexiones con calidad de servicio. Pero como los recursos son limitados, esto implica que la disciplina que quiere cumplir los requisitos del sistema sólo puede servir un número limitado de conexiones. Al conjunto de posibles combinaciones de las conexiones que garantizan el comportamiento adecuado de la red de manera simultanea sin que degrade la calidad de ninguna (o sea todos los parámetros estén dentro de las limitaciones impuestas por el sistema) se le llama región de planificación. Dada una región de planificación, realizar el control de admisión es muy simple ya que únicamente requiere mirar si la combinación resultante de los parámetros de las conexiones se encuentra dentro o no de la región de planificación. En conclusión en nuestro trabajo la región de planificación nos será de utilidad para comparar el número de usuarios simultáneos que puede aceptar el sistema con condiciones optimas de cumplimiento de la QoS. Este valor siempre esta referenciado a la calidad de servicio más restrictiva.
- Factor de utilización: Representa el porcentaje de uso que se esta realizando de los recursos disponibles por la red. Para su determinación únicamente se han considerado como ocupados aquellos recursos que transportaban datos de usuario. Se calcula como el cociente entre el número de radio bloques (RB) ocupados en la transmisión de datos (se incluyen también los RB utilizados para las retransmisiones y señalización para reconocimientos) y el número total de radio bloques disponibles durante el tiempo de la simulación. Se expresa en porcentaje

$$\text{Factor de utilización (FA)} = \frac{\text{N}^\circ \text{ RB datos}}{\text{N}^\circ \text{ RB duración simulación}}$$

Por lo que respecta a tráfico generado por los diferentes usuarios, en este trabajo se presentan dos escenarios claramente diferentes que permiten realizar un estudio más detallado por lo que respecta a la caracterización de las diferentes disciplinas de servicio analizadas. En los siguientes sub-apartados se presentan los resultados más relevantes para ambos escenarios.

3.1 Escenario A: Equilibrio entre las diferentes prioridades

En este escenario, se analiza el comportamiento de las disciplinas de servicio con prioridad presentadas en el punto anterior en un entorno donde existe un equilibrio de usuarios entre las diferentes prioridades. En este estudio se puede observar como con las disciplinas de servicio por prioridad consideradas en la

evaluación y bajo condición de carga idéntica entre prioridades, el sistema puede garantizar las cotas de retardo negociadas entre el terminal móvil y la red para las diferentes calidades de servicio. Los parámetros que describen este escenario son los siguientes:

Parámetros generales						
Esquema de codificación	CS-1					
Capacidad multislot	2(DL): 1 (UL)					
Retransmisiones RLC/MAC	3					
Duración de la simulación	1 hora					
Enlace	Ambos sentidos transmisión					
Parámetros de los modelos de tráfico	DOWNLINK			UPLINK		
Servicio	ALTA	MEDIA	BAJA	ALTA	MEDIA	BAJA
Topología www (%)	25	25	25			
Topología e-mail (%)	50	50	50	50	50	50
Topología SMS (%)	15	15	15	30	30	30
Topología MOBITEK (%)	10	10	10	20	20	20
Duración media sesión WWW (min)	7,5	7,5	7,5			
Duración media sesión e-mail (min)	5	5	5	5	5	5
Duración media sesión SMS (seg)	30	30	30	30	30	30
Duración media sesión MOBITEK (seg)	30	30	30	30	30	30
Parámetros de las disciplinas de servicio						
ARR (RB por Multitrama y prioridad)	7	4	1	7	4	1
WPQ (RB por Multitrama y prioridad)	7	4	1	7	4	1
Parámetros de QoS GPRS						
Prioridad	ALTA (40%)		MEDIA (30%)		BAJA (30%)	
Retardo	1		2		3	

Tabla 2. Parámetros del escenario A

Al existir diferentes prioridades, el número de usuarios simultáneos de los escenarios se ha estimado oportuno que sea función del peso de cada una de las prioridades así como del máximo número de usuarios simultáneos por enlace. En la Tabla 3 se detalla el conjunto de usuarios para cada enlace y prioridad.

Enlace	DESCENDENTE			ASCENDENTE		
	ALTA	MEDIA	BAJA	ALTA	MEDIA	BAJA
10	1	1	1	0	0	0
30	4	3	3	2	1	1
50	6	4	4	3	2	2
70	9	6	6	4	3	3
90	11	8	8	5	4	4

Tabla 3. Usuarios simultáneos escenario A

Para este escenario, en la Figura 5 se aprecia el comportamiento, en términos del retardo medio normalizado para el conjunto de aplicaciones, de las cuatro disciplinas evaluadas. Se considera que los usuarios cumplen con las cotas de retardo especificado para las diferentes tramas LLC siempre y cuando el retardo medio normalizado se encuentre en un valor inferior a la unidad [9]. De la figura se puede observar que las disciplinas que no consideran en su elección la longitud de las tramas, es decir SPS y WPQ, presentan el peor comportamiento. Esto es debido a que las tramas LLC de menor

longitud tienen un tiempo de servicio muy restrictivo para cumplir con el retardo medio (aplicaciones Mobitex y SMS) que es el parámetro que se desea evaluar.

La estrategia WPQ es la que presenta el peor comportamiento puesto que reparte su ancho de banda entre todas las prioridades y no toma en consideración la longitud de las tramas LLC en su elección del servicio. Conclusiones similares se pueden extraer de la disciplina SPS aunque en menor medida ya que esta da servicio en primer lugar a los usuarios de prioridad alta. En contraposición, las disciplinas que consideran en su elección la longitud de las tramas (SJM) o reparten el ancho de banda de forma equitativa entre los diferentes usuarios (ARR) son las que ofrecen los mejores resultados. Así, para la disciplina SJM se puede ver que aún incluso en el caso de presentar un nivel de carga equivalente al **90%** de usuarios simultáneos en el sistema este cumple con la calidad de servicio. Por su parte, para la disciplina ARR este valor se reduce a un **70%**.

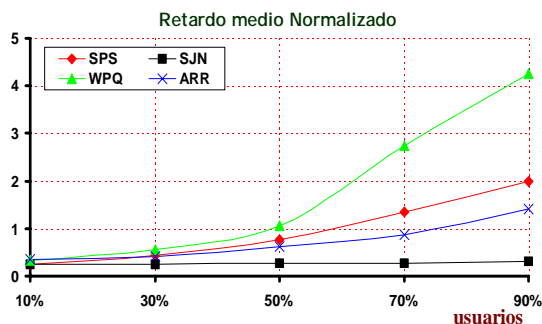


Figura 5. Retardo medio normalizado (ALTA)

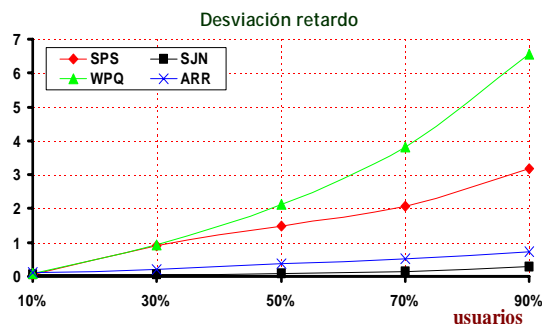


Figura 6. Desviación del retardo medio

En la Figura 6 se observa la desviación del retardo medio. Como era previsible las disciplinas SPS y WPQ presentan una gran dispersión en sus valores, debido, entre otros factores, a la aleatoriedad en la selección de los flujos a servir. Por el contrario las disciplinas ARR y más específicamente SJM escogen los paquetes a servir en función de su dimensión, por lo que sirven primero las tramas LLC más pequeñas lo que optimiza el uso de los recursos radio.

Cuando se analiza el comportamiento de los usuarios de prioridad media (Figura 7) se observa que todas las disciplinas cumplen con los requisitos definidos en el estándar en términos de retardo medio normalizado. Esto se debe principalmente a que para esta prioridad la cota del retardo no es tan restrictiva como la prioridad alta y, por tanto, hay una mayor permisividad para los usuarios de esta categoría tal y como se desprende de la Tabla 1. Se puede ver que hasta aproximadamente un nivel de carga equivalente al **50%** de usuarios simultáneos prácticamente no existe diferencia en el trato dispensado por las distintas disciplinas de servicio. Además se constata que la disciplina más óptima para usuarios de prioridad alta, como era SJM, pierde en este caso todas sus ventajas al tratarse de una disciplina que depende del estado de la cola de prioridad alta. Esto se analizará con mayor profundidad en el escenario siguiente.

Otro de los parámetros interesantes de comparar entre disciplinas es el retardo en transmitir el primer radio bloque. En la Figura 8 se presenta la diferencia del retardo en la transmisión del primer radio entre la prioridad media y la prioridad alta. Para las disciplinas que comparten el ancho de banda disponible entre los distintos usuarios se aprecia como la diferencia es muy pequeña, casos WPQ y ARR. En cambio, cuando la disciplina es dependiente del comportamiento de la cola de prioridad alta, la diferencia es mucho mayor y más destacada con el incremento de usuarios simultáneos.

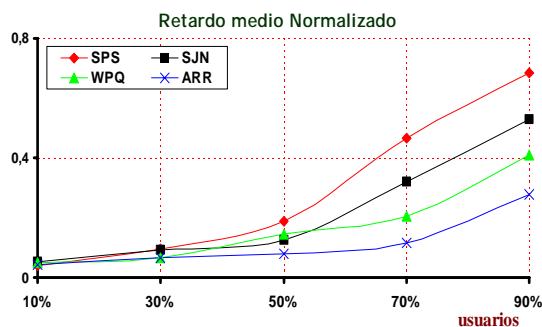


Figura 7. Retardo medio normalizado (MEDIA)

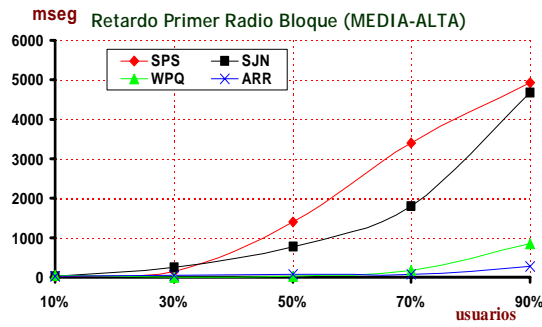


Figura 8. Retardo transmisión primer RB

En la Figura 9 se muestra para la disciplina SJN la velocidad media de transferencia para el servicio E-mail para las diferentes prioridades. De ella, se concluye que cuando hay poco usuarios simultáneos el sistema prácticamente no distingue entre la prioridad alta y la media, en cambio si que hay diferencia con la prioridad baja (aproximadamente 2-3 Kbps). Con el aumento del número de usuarios se puede observar como las prioridades medias y bajas presentan un descenso apreciable en la velocidad media de transferencia mientras que para la prioridad alta esta se mantiene prácticamente constante entorno a los 14 Kbps.

En la Figura 10 se analiza la región de planificación para este escenario de estudio. De ella se puede concluir que la disciplina que presenta mejores resultados es la SJN debido al tratamiento que se realiza con las tramas de menor longitud. En cambio, disciplinas como SPS y WPQ son las que en general presentan peores comportamientos, aunque tienen como ventaja su fácil implementación. Como se puede ver en la región de planificación únicamente se ha presentado las prioridades altas y medias. Este hecho se debe a que la prioridad baja es muy poco restrictiva y en consecuencia el número de usuarios que en ella se pueden emplazar simultáneamente en este escenario es poco representativo.

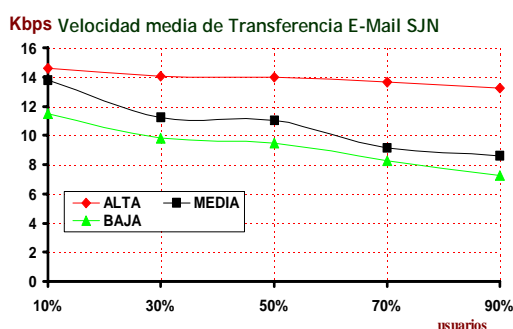


Figura 9. Velocidad media transferencia SJN

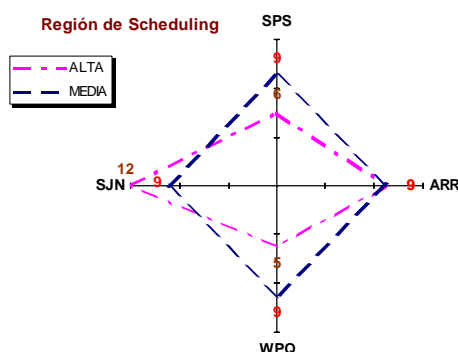


Figura 10. Región de planificación

En conclusión se ha presentado un escenario en donde había un equilibrio entre las diferentes prioridades posibles que admite el servicio GPRS. En él, se ha puesto en evidencia que disciplinas que toman en consideración la longitud media de las tramas LLC o comparten el ancho de banda entre los diferentes usuarios simultáneos se erigen como disciplinas óptimas para un entorno GPRS con una tipología de tráfico heterogénea.

3.2 Escenario B: Predominio usuarios de alta prioridad

En este escenario, se estudia el comportamiento de las disciplinas de servicio con prioridad analizadas en este trabajo pero en un entorno donde predomina los usuarios de prioridad alta.

Los parámetros que describen este escenario son los mismos que los presentados en el escenario A pero con las siguientes modificaciones en las prioridades por usuario:

Parámetros de QoS GPRS			
Prioridad	ALTA (70%)	MEDIA (20%)	BAJA (10%)
Retardo	1	2	3

Tabla 4. Parámetros del escenario B

Al igual que en el caso anterior, con esta caracterización de usuarios por prioridad, el número de usuarios simultáneos para cada prioridad se detalla en la siguiente tabla:

Enlace	DESCENDENTE			ASCENDENTE		
	ALTA	MEDIA	BAJA	ALTA	MEDIA	BAJA
10	2	0	0	1	0	0
30	7	2	1	3	1	0
50	11	3	1	5	1	0
70	16	4	2	7	2	1
90	20	5	2	9	2	1

Tabla 5. Usuarios simultáneos escenario B

En la Figura 11 se muestra el retardo medio normalizado para las diferentes disciplinas de servicio cuando el tráfico predominante se debe a los usuarios de alta prioridad. De la figura se deduce que, en una situación con distintas prioridades y alta carga de usuarios de prioridad alta, la mejor disciplina es aquella que sirve en primer lugar las tramas de longitud más pequeña puesto que son las más críticas en el cumplimiento de la calidad de servicio (parámetro retardo medio). De entre todas las estudiadas, la disciplina de servicio que obtiene mejores resultados, al igual que sucedía en el escenario anterior, es la denominada SJN, en donde la clasificación se realizaba en función de la longitud de las tramas. En segundo lugar ARR, al distribuir entre todos los usuarios presentes los recursos disponibles, consigue presentar un retardo acotado. Por último, SPS y WPQ como era de esperar, presentan peor comportamiento. Si se compara dicha figura con la mostrada en el escenario A (Figura 5) se comprueba que con el aumento del número de usuarios simultáneos para la prioridad alta el retardo de todas las disciplinas aumenta de forma considerable siendo más evidente para las disciplinas SPS y WPQ que triplican el retardo. Se ha comprobado que en el retardo medio normalizado para los algoritmos SPS y WPQ viene muy marcado por las aplicaciones que generan poco volumen de tráfico mientras que las disciplinas ARR y SJN el efecto es totalmente el contrario, siendo las aplicaciones que generan un volumen elevado de tráfico quien modela el comportamiento.

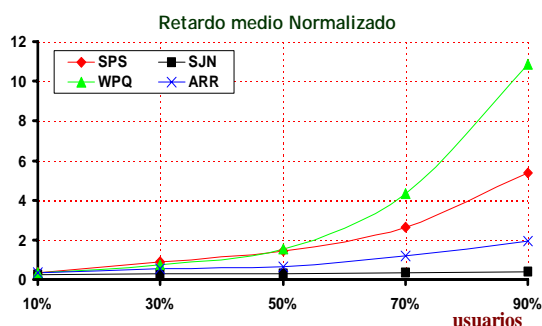


Figura 11. Retardo medio normalizado (ALTA)

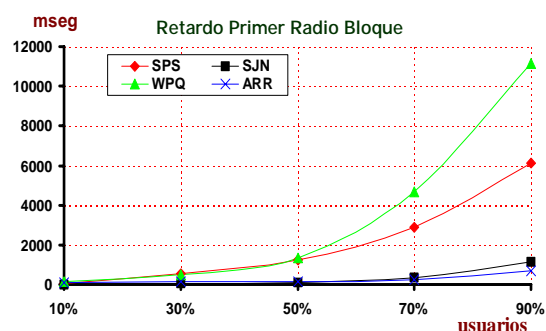


Figura 12. Retardo primer radio bloque (ALTA)

Si se analiza el retardo en la transmisión del primer radio bloque se observa que la disciplina que presenta un mejor comportamiento es la estrategia ARR ya que comparte los recursos disponibles entre todos los usuarios simultáneos presentes en el sistema (Figura 12). Por su parte, disciplinas como SPS o WPQ presentan el peor comportamiento debido a que no realizan ninguna ordenación de los flujos. En algunas disciplinas, y como consecuencia de mantener prácticamente constante el retardo medio normalizado, se consigue obtener una velocidad media de transferencia también constante. En la Figura 13 se presenta para las cuatro disciplinas analizadas un ejemplo del comportamiento de la aplicación WWW para usuarios de alta prioridad. Mientras la disciplina SJN mantiene prácticamente constante la velocidad de transmisión con una fluctuación máxima de 2 Kbps entre un 10% hasta el 90% de usuarios simultáneos, para la disciplina WPQ este valor varía desde los 14 Kbps para un nivel de carga equivalente al 10% de usuarios hasta situarse aproximadamente en los 4 Kbps para un 90% de usuarios.

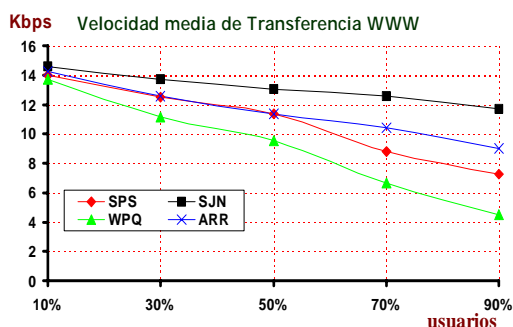


Figura 13. Velocidad media de transferencia WWW (ALTA)

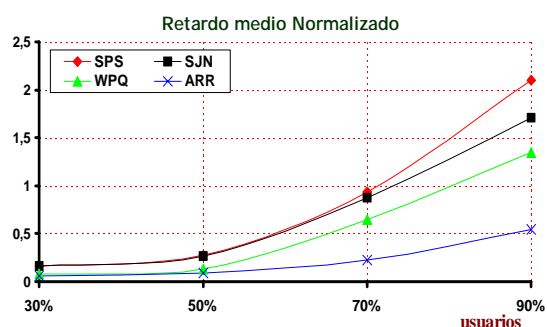


Figura 14. Retardo medio normalizado (MEDIA)

Cuando se evalúan los mismos esquemas de asignación de recursos, pero para usuarios de prioridad media, se ve como para la estrategia SJN desaparece una de sus mayores ventajas por lo que respecta al tratamiento de los paquetes de longitud menor. Básicamente, esto se debe a que, en el escenario analizado, el porcentaje de usuarios de alta prioridad es muy elevado y por tanto ello repercute en un consumo muy elevado de recursos, lo que afecta al resto de usuarios de prioridades inferiores. Este hecho no se había observado con demasiada claridad en el escenario anterior debido al equilibrio entre usuarios. Es decir, aunque las tramas generadas por estos usuarios lleguen en primer lugar al gestor de recursos radio, permanecen en la cola un tiempo más elevado que el permitido para cumplir con la calidad de servicio (Figura 14) al estar el sistema sirviendo a usuarios de prioridad más alta. En general, tanto el sistema SPS como SJN tienen prácticamente el mismo comportamiento mientras que el algoritmo WPQ mejora su comportamiento. Esto se debe a que el algoritmo WPQ asigna un mínimo ancho de banda para la transmisión de paquetes de prioridad media (4 RB por multitrama). Por último, se observa que el algoritmo ARR es el que mejor se comporta puesto que como ya se ha comentado comparte los recursos disponibles entre todos los usuarios.

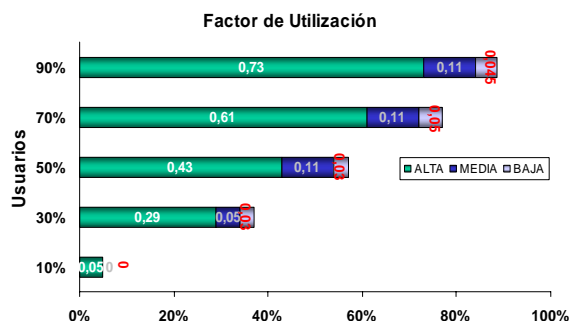


Figura 15. Factor de utilización

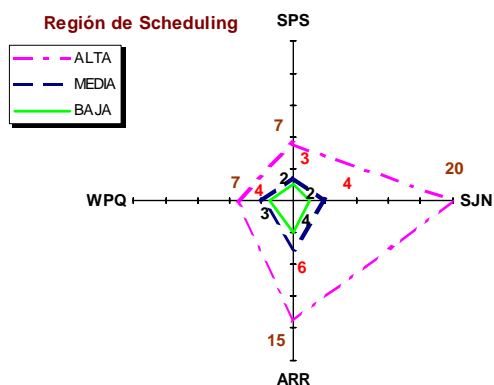


Figura 16. Región de planificación

En la Figura 15 se presenta el factor de utilización para este escenario. Como era de esperar el factor de utilización no depende de la disciplina de servicio, por lo cual sólo se han presentado resultados para una de ellas (SJN). En la misma puede verse como para un 90% de usuarios simultáneos el sistema alcanza un factor de utilización cercano al 85% para el conjunto de aplicaciones, siendo la prioridad alta con casi un 75% la predominante en la utilización de los recursos disponibles.

Finalmente, en la Figura 16 se constata como las disciplinas SJN y ARR son las que presentan una región de planificación más adecuada para entornos GPRS en donde existe un número de usuarios dominantes de prioridad elevada. Los resultados para este escenario son más concluyentes con relación al comportamiento de las diferentes estrategias analizadas que los mostrados en el escenario anterior. En efecto, con las disciplinas de servicio por prioridad consideradas en la evaluación y bajo condiciones de carga de la red elevada (prioridad alta) se obtiene mayores diferencias entre las disciplinas en cuestión, llegándose a la conclusión que disciplinas como SPS y WPQ no son óptimas para un entorno GPRS con tráfico heterogéneo.

Esto se debe a que en este caso se ha asignado a la prioridad alta un mayor número de usuarios y además cabe recordar que esta es la más restrictiva en cuanto a retardo medio.

4. Conclusiones

En este estudio se ha realizado una evaluación del comportamiento de las estrategias de gestión de recursos radio para el sistema GPRS, considerando la calidad de servicio en la red. Para ello se han tomado en consideración diferentes modelos de tráfico representativos de algunos de los servicios de Internet, un modelo de movilidad del terminal móvil y se han analizado sobre cuatro disciplinas de servicio basadas por prioridad.

Los resultados muestran la importancia que tiene la selección de una disciplina de servicio en un entorno con tráfico heterogéneo para poder cumplir con los atributos de calidad de servicio negociados entre el terminal móvil y la red para el mayor número posible de terminales móviles. De ellos se constata que las disciplinas que consideran la longitud de las tramas LLC o la tipología de tráfico en su defecto son las que mejor comportamiento presentan, por ejemplo SJN. Sucede efectos similares si se reparte el ancho de banda simultáneamente entre los diferentes usuarios presentes en el sistema, ARR.

Por contra, disciplinas que sirven los usuarios de forma aleatoria, SPS o WPQ presentan comportamientos muy dispares y no se consideran óptimas de instalar en el sistema aunque presenten una fácil implementación.

Referencias

- [1] European Telecommunications Standards Institute, Digital cellular telecommunications system (Phase 2+) General Packet Radio Service- Overall description of the GPRS radio interface; Stage 2, GSM 03.64 version 7.0.0 Release 1998
- [2] European Telecommunications Standards Institute, Digital cellular telecommunications system (Phase 2+) General Packet Radio Service- Mobile Station (MS) – Base Station System (BSS) interface; Radio Link Control/Medium Access Control (RLC/MAC), GSM 04.60 version 7.0.1 Release 1998
- [3] Digital cellular telecommunications system (Phase 2+) (GSM) Universal Mobile Telecommunications System (UMTS); General Packet Radio Service- Service description; Stage 1, 3GPP TS 22.060 version 3.5.0 Release 1999.
- [4] Zhang H. "Service Disciplines for Guaranteed Performance Service in Packet_Switching Network" Proceedings of the IEEE, vol. 83, no 10, pp.1374-96, Oct 1995
- [5] Hiroshi Saito, "Optimal Queuing Discipline for Real-Time Traffic at ATM Switching Nodes", IEEE Transactions on Communications, December 1990, pp. 2131-2136.
- [6] <http://www.cs.wpi.edu/~cs3013/c02/week3-sched/week3-sched.html>
- [7] ETSI Technical Specification UMTS 30.03, "Universal Mobile Telecommunications System (UMTS); Selection procedures for de choice of radio transmission technologies of the UMTS", version 3.0.1.
- [8] Bada J., Casadevall F., "Análisis comparativo de estrategias de asignación de recursos radio en el sistema GPRS. XI Jornadas de I+D en Telecomunicaciones", Telecom. I+D, Septiembre 2001
- [9] Sau J., Sccholefield C. "Scheduling and Quality of Service in the General Packet Radio Service", ICUP 1998.