

Programa de doctorado 'Tecnología de las Telecomunicaciones'
Departamento de Señales y Comunicaciones
UNIVERSIDAD DE LAS PALMAS DE GRAN CANARIA



TESIS DOCTORAL

Contribuciones a la gestión de los recursos radio en
redes móviles celulares de radio paquetes con soporte de
calidad de servicio

Octubre 2011
Las Palmas de Gran Canaria

Autor: Carlos M. Ramírez Casañas
Director: Josep Paradells Aspas
Codirector: Rafael Pérez Jiménez

A Irene, Sofia y Victoria
A mis padres y a mi hermana
A mi abuelo materno Manuel

Agradecimientos

Desde estas líneas quiero agradecer, de forma muy sincera y especial por su guía y su dedicación, a mi director de tesis Josep Paradells. También quiero agradecer al codirector de la tesis Rafael Pérez por su total apoyo.

Quiero destacar y agradecer a Sonia P. Mansilla por su tiempo y su colaboración, esenciales para *ver* de lo particular lo general.

Por último, desde aquí agradecer a todos los amigos y compañeros del Departamento de Ingeniería Telemática de la Universidad Politécnica de Cataluña, del Departamento de Ingeniería Telemática y del Departamento de Señales y Comunicaciones de la Universidad de Las Palmas de Gran Canaria que he conocido a lo largo de este tiempo y que han hecho que esto haya valido aún más la pena.

Índice general

1. Introducción	15
1.1. Motivación	15
1.2. Evolución de las redes móviles celulares de paquetes de datos	16
1.2.1. Sistemas celulares de segunda y tercera generación	17
1.2.2. Más allá de la tercera generación de redes móviles, 3.5G y 4G	21
1.2.3. Interfaz radio	23
1.2.4. Introducción a la calidad de servicio en redes celulares	29
1.2.5. Mecanismos de soporte a la calidad de servicio en redes celulares	34
1.3. Objetivos de la presente tesis doctoral	43
2. Modelado de los recursos radio en redes celulares de paquetes de datos	45
2.1. Introducción	45
2.2. La asignación de los recursos radio en las redes móviles celulares	46
2.2.1. Fundamentos de la asignación de los recursos radio	52
2.2.2. Modelos de asignación de canales	53
2.3. Trabajos anteriores acerca de los métodos de asignación de canales	56
2.4. Criterios de calidad de servicio y reglas de dimensionado	59
2.4.1. Conceptos de ingeniería de tráfico	59
2.4.2. Modelos de colas y caracterización del sistema móvil celular	60
2.5. Conclusiones	65
3. Asignación compartida de recursos en la red móvil EDGE	67
3.1. Introducción	67
3.2. Modelo de asignación compartida por ráfagas	68

3.2.1.	Análisis del modelo	70
3.2.2.	Concepto de ráfagas: programación y análisis	73
3.3.	Esquema de asignación repacking	83
3.3.1.	Modelo repacking	83
3.3.2.	Resultados	86
3.4.	Conclusiones	88
4.	Asignación con reserva de recursos para el sistema EDGE	91
4.1.	Introducción	91
4.2.	Reserva de recursos radio en EDGE	92
4.2.1.	Esquema de asignación de recursos compartida con reserva para tráfico de voz	92
4.2.2.	Modelo analítico	93
4.2.3.	Resolución del modelo	96
4.2.4.	Medidas de rendimiento	100
4.2.5.	Esquema de asignación de recursos compartida con reserva para tráfico de datos	103
4.3.	Conclusiones	106
5.	Asignación de recursos con reserva y buffering para EDGE	109
5.1.	Introducción	109
5.2.	Esquema de asignación de recursos con reserva y buffering	110
5.2.1.	Modelo analítico	111
5.2.2.	Resolución del modelo	114
5.2.3.	Medidas de rendimiento	120
5.3.	Conclusiones	126
6.	Conclusiones y líneas futuras	129
6.1.	Conclusiones	129
6.2.	Líneas futuras de investigación	131

Índice de figuras

1.1. Capas y subcapas de la interfaz radio	24
1.2. Elementos de QoS en el equipo terminal	39
2.1. Funciones del Controlador de Red Radio	49
2.2. Interacción para la gestión de recursos radio común	51
2.3. Asignación de canales mediante repositorios	53
2.4. Método de <i>repacking</i>	54
2.5. Asignación de canales mediante un único repositorio	55
2.6. Asignación recursos de voz GSM y datos GPRS (Ni and Häggman)	57
2.7. Cadena de Markov binomial para dos clases de tráfico	62
2.8. Cadena de Markov para un proceso IRP	63
2.9. Cadena de Markov unidimensional para un proceso NIRP	64
3.1. Situación momentánea de imposibilidad de capacidad multi-slot para tráfico de datos	68
3.2. Ejemplo de asignación de time-slots	70
3.3. Diagrama de transición para el estado 5 con 8 time-slots	72
3.4. Probabilidades de encontrar estados con una probabilidad de bloqueo del 1%	73
3.5. Probabilidades de encontrar estados con probabilidad de bloqueo del 2%	74
3.6. Estado 2 con ráfagas de 6 y 1 time-slots libres consecutivos	75
3.7. Probabilidad de encontrar ráfagas con probabilidad de bloqueo del 1%	76
3.8. Probabilidad de encontrar ráfagas con probabilidad de bloqueo del 2%	77
3.9. Diferencia entre tasas de pico con probabilidad de bloqueo del 1% frente al 2%	78

3.10. Probabilidad acumulativa de encontrar ráfagas con probabilidad de bloqueo del 1%	79
3.11. Representación de los estados con ráfagas de $m=2$ (gris claro) en posiciones central-izquierda	82
3.12. Representación de los estados en el método repacking	85
3.13. Proceso de intra-cell handover	86
3.14. Distribución de probabilidad con una probabilidad de bloqueo del 1% usando el método repacking	87
3.15. Distribución de probabilidad con una probabilidad de bloqueo del 2% usando el método repacking	87
4.1. Asignación de time-slots con reserva para tráfico de voz	93
4.2. Diagrama de transición para el estado 9	98
4.3. Diagrama de transición para el estado 20	100
4.4. Escenarios con $n-m$ desde 0 hasta 4 time-slots reservados y probabilidad de bloqueo del 1%	101
4.5. Escenarios con $n-m$ desde 0 hasta 4 time-slots reservados y probabilidad de bloqueo del 5%	102
4.6. Asignación de time-slots con reserva para tráfico de datos	103
4.7. 8 time slots con 1 time-slot reservado y 7 time-slots compartidos	104
4.8. 8 time slots con 2 time-slots reservados y 6 time-slots compartidos	105
5.1. Asignación de time-slots con reserva y buffering	110
5.2. Diagrama de transición de estados para $n=3$ time-slots, $m=1$ time-slot compartido, $q=1$ elemento en cola)	115
5.3. Diagrama de transición para el estado 20 ($n=4, m=2, q=1$)	116
5.4. Diagrama de transición para el estado 24 ($n=4, m=2, q=1$)	117
5.5. Probabilidad de Pérdida y Probabilidad de Bloqueo para un sistema en el que el tráfico de voz es mayor que el tráfico de datos, $n=8, m=1$	121
5.6. Probabilidad de Pérdida y Probabilidad de Bloqueo para un sistema en el que el tráfico de voz es mayor que el tráfico de datos, $n=8, m=2$	122
5.7. Probabilidad de Pérdida y Probabilidad de Bloqueo para un sistema en el que el tráfico de voz se aproxima al tráfico de datos, $n=8, m=1$	123

5.8. Probabilidad de Pérdida y Probabilidad de Bloqueo para un sistema en el que el tráfico de voz es menor que el tráfico de datos, $n=8$, $m=1$ 124

5.9. Probabilidad de Pérdida y Probabilidad de Bloqueo para un sistema en el que el tráfico de voz es menor que el tráfico de datos, $n=8$, $m=4$ 125

Índice de cuadros

1.1. Parámetros de QoS en 3GPP Release'97/98	31
1.2. Parámetros de QoS definidos para GERAN y UTRAN en 3GPP Release'99	33
4.1. Comparativa de probabilidades de encontrar ráfagas sin reser- va y con reserva de 1 time-slot con Probabilidad de Bloqueo del 1%	105
4.2. Comparativa de probabilidades de encontrar ráfagas sin reser- va y con reserva de 1 time-slot con Probabilidad de Bloqueo del 2%	106
4.3. Comparativa de probabilidades de encontrar ráfagas sin reser- va y con reserva de 2 time-slots con Probabilidad de Bloqueo del 1%	106
4.4. Comparativa de probabilidades de encontrar ráfagas sin reser- va y con reserva de 2 time-slots con Probabilidad de Bloqueo del 2%	107

Capítulo 1

Introducción

1.1. Motivación

Las redes móviles celulares de radio paquetes han supuesto la introducción de nuevos servicios y el soporte a nuevas aplicaciones en un entorno móvil. Este hecho ha permitido la posibilidad de disponer de un acceso a contenidos de información multimedia a través de la red de datos global Internet. En concreto GPRS (General Packet Radio Service) [ETSI GPRS (2000)] y EGPRS (Enhanced GPRS)[Molkdar et al. (2002)] han sido los servicios de paquetes de datos dentro de la red GSM (Global System for Mobile Communications) para el soporte del transporte de datos en los sistemas móviles celulares de segunda generación. Con el desarrollo del estándar GSM para las redes móviles celulares se ha incrementado el número de usuarios de forma global en las comunicaciones interpersonales. En la evolución de esta popular red, cada vez son más los usuarios que requieren la disponibilidad de nuevos servicios que actualmente disfrutan a través de las redes fijas de datos para ser utilizadas en este nuevo entorno móvil. Este hecho ha producido la necesidad de incrementar la capacidad de los servicios de datos con tasas mayores a las proporcionadas por GSM y mejoras en los servicios de voz. Los servicios de conmutación de paquetes basadas en GPRS han estado disponibles en Europa desde el año 2001. Debido a este nuevo servicio se han ofrecido tasas de pico por encima de los 64Kbits/s a las aplicaciones de datos en entornos móviles.

Con objeto de obtener tasas de datos altas el Instituto de Estándares de Telecomunicación (ETSI) ha desarrollado el estándar EDGE (Enhanced

Data rates for GSM Evolution)[3GPP GERAN (2004)] que formando parte de la tercera generación de los sistemas móviles celulares y, en particular, dentro del estándar UMTS (Universal Mobile Telecommunication System)[UMTS ETSI (1997)] consigue tasas de pico por encima de 144 Kbits/s en un amplio escenario de cobertura y hasta tasas máximas de 384Kbits/s en ambientes móviles cercanos. En el caso del propio sistema UMTS se logran tasas de 2Mbits/s en escenarios interiores. Estos sistemas han significado un proceso de extensión y de desarrollo de los sistemas celulares existentes. La ventaja de este proceso de evolución es la rápida disponibilidad de los servicios que ya se utilizaban en los sistemas de segunda generación, además de proporcionar la oportunidad de preparar a los usuarios para los nuevos servicios de la tercera y cuarta generación.

1.2. Evolución de las redes móviles celulares de paquetes de datos

Los servicios de comunicaciones personales que incluyen las redes móviles celulares de paquetes de datos, desde una perspectiva actual, se pueden abordar por su significado desde dos puntos de vista[Cheng et al. (2008)] bien diferenciados. La primera visión es desde una perspectiva puramente tecnológica diferenciando las características más relevantes que distinguen una generación de redes frente a otras así como su desarrollo a lo largo de los últimos años. La segunda perspectiva se utiliza bajo el paraguas que incluye varios servicios de acceso inalámbrico y de movilidad personal con el objetivo último de permitir a los usuarios comunicarse entre ellos libremente en cualquier lugar y a cualquier hora de acuerdo con sus necesidades (voz, datos, imágenes y vídeo).

Atendiendo a la primera visión, la evolución tecnológica de las redes móviles, en concreto las redes celulares de paquetes de datos tienen su origen y desarrollo a partir de los años 90, por medio de la digitalización de la transmisión de las señales vocales a través de una red de transmisión digital que solo empleaba la conmutación de circuitos (2G, Segunda Generación de redes celulares). A partir de este tipo de redes móviles celulares se ha empezado a introducir la transmisión de servicios de datos básicos a bajas velocidades. La traslación de las nuevas aplicaciones utilizadas por los usuarios desde las redes fijas de comunicaciones al entorno móvil, ha supuesto la necesidad del

transporte de datos con mayor intensidad que, con la incorporación de un servicio propio para la transmisión de datos y utilizando la conmutación de paquetes junto a una red troncal basada en el protocolo IP, hacen posible el transporte de datos de forma más eficiente.

La evolución desde la segunda generación de redes móviles celulares hacia la tercera generación (3G) ha ido de forma gradual añadiendo nuevas funciones y posibilidades a las redes existentes. Esta evolución comienza con la actualización de las redes de 2G a las denominadas redes 2.5G y a partir de éstas a las actuales redes móviles.

1.2.1. Sistemas celulares de segunda y tercera generación

1.2.1.1. Segunda generación de sistemas móviles celulares

La segunda generación de redes móviles celulares (2G) se caracteriza fundamentalmente por la transmisión digital de la información por medio de la conmutación de circuitos. De forma casi exclusiva está implementada para dar servicio de voz aún pudiendo además proporcionar servicio de transmisión de datos con tasas de hasta 9,6Kbps. Las redes de segunda generación introducen nuevas bandas de frecuencia y multiplica el número de usuarios que podía atenderse con respecto a sistemas de la primera generación de redes móviles. Supuso la estandarización de forma global entre países que garantizaba la compatibilidad entre los distintos operadores y regiones a través de la itinerancia o *roaming* siempre que utilizaran la misma tecnología estándar. La incorporación de las redes 2G, como por ejemplo GSM (Global System for Mobile Communications), significó hacer rentable los sistemas de comunicaciones móviles para los operadores, además de utilizar protocolos de codificación más sofisticados e incluir técnicas de autenticación para los mensajes, tanto de señalización como de transmisión de la información de los usuarios, lo cual implicaba comunicaciones seguras frente a la primera generación de redes móviles (1G).

Las redes móviles celulares de segunda generación, en función de dónde han sido implementadas y estandarizadas [Vriendt et al. (2002)], se denominaron como:

- GSM (Global System for Mobile Communications) en Europa (conocida en Estados Unidos como PCS, Personal Communication Services), basada en la tecnología de acceso TDMA (Time Division Multiple Ac-

cess) operando en 900 y 1800 MHz en Europa y en 1900 MHz en América.

- TDMA IS-54 también denominada DAMPS (Digital Advanced Mobile Phone Service) ó ANSI-13 que posteriormente y con un desarrollo más amplio se denominó IS-136 ó Digital PCS similar a GSM, se implantó en Norteamérica; por la propia denominación utiliza la tecnología de acceso TDMA y como principal característica se adapta al sistema AMPS conservando el mismo ancho de banda de los canales AMPS. Trabaja en frecuencias de 800 y 1900MHz.
- E-TDMA (Extended-TDMA) representa una extensión del TDMA IS-136, usando como característica destacada con respecto al IS-54 que los time slots se asignan de forma dinámica de acuerdo con los requerimientos de servicio que se presenten en cada momento.
- CDMA IS-95 (también denominado cdmaOne ó Narrowband-CDMA (N-CDMA)) lanzado comercialmente en Hong Kong y que opera en distintos países de Sudamérica, África y Asia, incluye un sistema de 1,9GHz que opera en Estados Unidos. Caracterizado principalmente por utilización de celdas de radio pequeño y la utilización de espectro ensanchado para proporcionar alta seguridad y calidad en la transmisión de la señal.
- PDC (Personal Digital Communications), este sistema desarrollado en Japón utiliza una tecnología de acceso TDMA que emplea pocos time slots para las comunicaciones de voz y también utilizado para la transmisión de datos con tasas de 9,6Kbps. La arquitectura de la red es muy similar a la de GSM. Opera con frecuencias de 900 y 1500MHz.

1.2.1.2. Generación 2.5G, transición de la 2G a la 3G

La ITU (International Telecommunication Union) definió un conjunto de requisitos para el desarrollo de las redes móviles hacia los sistemas de tercera generación. El principal de estos requisitos fue el soporte para los servicios de datos con tasas por encima de los 144Kbps en entornos abiertos y de 2Mbps en entornos cerrados, así mismo otros requisitos fueron el soporte eficiente del tráfico asimétrico y la transmisión por medio de la conmutación de paquetes y la alta eficiencia del espectro. Atendiendo a estos requisitos, a partir de 1998

aparece y se desarrolla la generación de sistemas móviles 2.5G ó también denominada 2G+ de camino hacia la 3G. Estos sistemas se caracterizan por ofrecer capacidades adicionales a los sistemas de segunda generación, tal es la utilización de la conmutación de paquetes para el transporte de datos, además de otras características adicionales como la de optimizar el uso eficiente de los recursos radio. Asimismo la incorporación de estos sistemas ha facilitado la utilización de Internet móvil. Estas redes móviles celulares han representado y representan la forma más adecuada de transitar entre la segunda y tercera generación de redes móviles celulares. La incorporación de los sistemas 2.5G en las operadoras europeas y de Estados Unidos comenzaron inicialmente a partir del año 2001, en Japón el paso de la segunda generación a la tercera fue directa.

Estos sistemas 2.5G prestan servicios de voz y de datos así como servicios de mensajería y tarificación en función del volumen de datos transmitidos. Como sistemas de la 2.5G se incluyen:

- HSCSD (High Speed Circuit Switched Data)[HSCSD ETSI (1997)] aparece como una mejora de los servicios de datos de la red GSM. Permite tasas de hasta 57,6Kbps pero al estar basada en la conmutación de circuitos, de forma inherente, es ineficiente para el tráfico a ráfagas. Este hecho unido a la pronta incorporación de otras arquitecturas de red que utilizan la conmutación de paquetes de datos, ha provocado que la mayor parte de los operadores hayan desistido de la implantación de esta tecnología en la práctica.
- GPRS (General Packet Radio Service) incorpora a la red GSM la conmutación de paquetes para la transmisión de datos de forma eficiente. Manteniendo la estructura de trama, la modulación de la interfaz radio y las bandas de frecuencia; se caracteriza por permitir el envío y recepción de datos en cualquier momento, la mejora del uso de los recursos radio y la facturación basada en el volumen de datos transmitidos.
- CDMA2000 que representa la evolución de cdmaOne, se trata de una tecnología de acceso CDMA multiportadora. La designación 1xRTT (1 time Radio Transmission Technology) se utiliza para identificar la versión de la tecnología CDMA2000 que opera con un par de canales de 1,25 MHz y casi duplica la capacidad de voz sobre las redes IS-95. Aunque capaz de soportar altas velocidades de datos, la mayoría de desarrollos están limitados a una velocidad de pico de 144 kbits/s.

1.2.1.3. Tercera generación de sistemas móviles celulares

La Unión Internacional de las Telecomunicaciones (ITU) definió por medio de la recomendación ITU-R-M687-2 los sistemas de la tercera generación de redes móviles (3G) o también denominados sistemas IMT-2000 (International Mobile Telecommunications-2000). Dicha recomendación incluye aspectos que deben cubrir estos sistemas 3G como son: calidad de servicio, QoS, equiparable a la de las redes fijas, compatibilidad y adecuación con los anteriores estándares de generaciones anteriores, una arquitectura flexible, acceso a Internet por medio de los dispositivos móviles, una gestión eficiente de la red que incluya la reducción de costes, así como una mejor oferta de servicios con respecto a las redes implantadas anteriormente. Además de estas características, se insiste en el hecho de que el acceso ha de ser posible en cualquier momento y lugar desde cualquier dispositivo.

Los sistemas de tercera generación IMT-2000 se pueden agrupar en cuatro categorías en función del tipo de tecnología de acceso:

- TDMA. Dentro de esta tecnología se incluye el sistema EDGE (Enhanced Data rates for GSM Evolution) que representa la evolución y desarrollo de GSM e IS-136, también denominado como UWC-136 (Universal Wireless Communications - 136). Este sistema complementa, en lugares con densidad de abonados[Hälönen et al. (2002), Ramirez et al. (2007)] pequeña a UMTS utilizando la red de acceso radio GSM/EDGE Radio Access Network (GERAN).
- WCDMA (Wideband Code Division Multiple Access). Dentro de esta tecnología de acceso se incluye el sistema UMTS (CDMA2000, 3xRTT, en Estados Unidos) que proporciona mejoras considerables de QoS y de gestión eficiente de los recursos.
- TD-CDMA (Time Division - Code Division Multiple Access). Implica a aquellos sistemas que contienen el componente TDD (Time Division Duplex) de UMTS y el TD-SCDMA (Chinese Time Division-Synchronous Code Division Multiple Access).
- FD-TDMA (Frequency Division - Time Division Multiple Access). Utilizado para el desarrollo de la telefonía digital sin cables DECT (Digital Enhanced Cordless Telecommunications).

El 3GPP (The 3rd Generation Partnership Project)¹ se organiza en grupos específicos TSGs (Technical Specification Groups) que definen: el núcleo de la red, los aspectos relacionados con los servicios y los sistemas, la red de acceso radio (RAN, Radio Access Network) y los aspectos relacionados con los terminales móviles.

1.2.2. Más allá de la tercera generación de redes móviles, 3.5G y 4G

Actualmente las redes móviles correspondientes a la tercera generación están en plena transición hacia redes cuyos principales objetivos se centran en tecnologías capaces de incrementar sus tasas de datos y la capacidad de transferencia de datos tanto para el enlace descendente como para el ascendente. En concreto el 3GPP introduce el estándar High Speed Packet Access (HSPA)[3GPP HSPA (2007)] con dichos objetivos. La evolución de este estándar, denominado HSPA+, mejora el rendimiento de la red radio en términos de eficiencia del espectro, tasas de pico y el retardo, así mismo explota el potencial del modo de acceso WCDMA. Las modificaciones en este sentido se concretan en la utilización de antenas MIMO (Multiple Input Multiple Output), un orden superior de modulación para el enlace de subida (16QAM) y el de bajada (64QAM), mejoras de los protocolos de nivel 2 y la conectividad continua para la transferencia de los paquetes de datos. De esta forma se logra conseguir tasas desde 14 hasta 42 Mbits/s en el enlace de bajada y desde 5,8 hasta 11 Mbits/s en el enlace de subida.

Con objeto de asegurar la evolución de este tipo de redes con una tecnología de acceso radio optimizada aparece el término Long Term Evolution, LTE,[3GPP LTE (2008), Bestak (2008)] y SAE (Service Architecture Evolution) para el núcleo de la red o red troncal, cuyo objetivo es nuevamente aumentar la capacidad y la tasa de transmisión de datos, la eficiencia del espectro, el bajo retardo, la interacción con los sistemas UTRAN/GERAN existentes y también aspectos relacionados con la reducción de los costes de instalación y operación de la red. LTE utiliza nuevos esquemas de acceso sobre la interfaz aire como es el esquema OFDMA (Orthogonal Frequency

¹Organismo mundial para coordinar los esfuerzos entre los institutos de estandarización (ETSI en Europa, ARIB y TTC en Japón, TTA en Estados Unidos, CWTS en China y TTA en Corea). Para el desarrollo de los estándares del cdma2000 se constituyó el 3GPP2 formado por los institutos ARIB, TTA y CWTS

Division Multiple Access) en el enlace de bajada y el SC-FDMA (Single Carrier - Frequency Division Multiple Access) en el enlace de subida. Además los esquemas MIMO forman una parte esencial de LTE.

La cuarta generación de redes móviles, 4G, está basada totalmente en la familia de protocolos TCP/IP y define la convergencia entre las redes fijas y las inalámbricas con tasas de transmisión de datos por encima de los 100Mbits/s. Estas redes han de mantener una QoS extremo a extremo y una alta seguridad para permitir ofrecer servicios de cualquier clase en cualquier momento, en cualquier lugar, con el mínimo coste posible. El Wireless World Research Forum (WWRF) define la cuarta generación de redes como una red que opere con la tecnología de Internet combinándola con otros usos y tecnologías tales como Wireless Local Area Network, WLAN, y Mobile WiMAX, estándar IEEE802.16e, que ha sido incluido bajo el paraguas de la IMT-2000 para ofrecer beneficios significativos a la comunidad móvil y así contribuir a la entrega global de servicios de Internet a bajo coste. WiMAX (Worldwide Interoperability for Microwave Access) es una tecnología dentro de las conocidas como tecnologías de última milla, también conocidas como de bucle local. Una de sus características primordiales es dar servicio de banda ancha en zonas donde el despliegue de cable o fibra óptica por la baja densidad de población presenta unos costes por usuario muy elevados (zonas aisladas y rurales).

Por un lado, el estándar IEEE 802.20, MBWA (Mobile Broadband Wireless Acces), estándar diseñado para redes inalámbricas de banda ancha basado en servicios IP móvil y que pretende ser una especificación de la cuarta generación de redes móviles, ha de especificar la capa física y la capa MAC de la interfaz aire basada en la conmutación de paquetes y optimizada para el transporte IP. Por otro lado, las redes inalámbricas Mesh, también denominadas redes acopladas o redes malladas, son aquellas redes en las que se mezclan las dos topologías de las redes inalámbricas: la topología ad-hoc y la topología de infraestructura. Estas redes utilizan la topología de infraestructura básicamente para agregar dispositivos que, a pesar de estar fuera del rango de cobertura de los puntos de acceso, están dentro del rango de cobertura de alguna tarjeta de red que directamente o indirectamente está dentro del rango de cobertura de un punto de acceso. Permiten que las interfaces de red se comuniquen entre sí, independientemente del punto de acceso. La convergencia de estas tecnologías y la posibilidad de comunicación eficiente entre ellas, en un entorno móvil IP, representa el objetivo principal de la cuarta generación de las redes de comunicaciones.

1.2.3. Interfaz radio

La definición de una interfaz radio en un entorno de redes móviles es un proceso iterativo [Prasad et al. V1 (2009)], tanto por consideraciones técnicas puras como por su regulación u otras decisiones y restricciones de otra índole. El diseño de la interfaz radio es un problema multidimensional que depende, en cada caso, de distintos factores como son: el escenario físico y el canal radio, los requisitos de las aplicaciones, las capacidades de los terminales, la carga real de tráfico y el perfil de los abonados. Es por esto que se hace necesario el estudio de las condiciones físicas definidas por la información del estado del canal para la mejor explotación de los recursos radio, así como el modo de operación óptimo de los terminales de usuario, el tamaño del área de cobertura, la capacidad que puede proporcionarse a los usuarios conectados a la red, los requisitos de las aplicaciones en términos de la demanda en tiempo real y las tasas de bits, es decir, requerimientos específicos de calidad de servicio.

La interfaz radio se modela por medio de una pila de protocolos que consta de varias capas y subcapas (Figura 1.1):

1. Capa Física (PHY, Physical layer): dedicada al procesado de las señales de radiofrecuencia, la modulación/demodulación, el multiplexado/demultiplexado, la sincronización temporal y frecuencial, la macrodiversidad y el traspaso, así como el control de potencia.
2. Subcapa de Control de Acceso al Medio (MAC, Medium Access Control): centrada en el canal de acceso, la resolución de la contención de acceso al medio, el multiplexado de los canales lógicos, el mapeado de los canales lógicos y los canales de transporte, la gestión de los flujos de datos de usuario.
3. Subcapa de Control de Enlace (RLC, Radio Link Control): se centra en aspectos como la segmentación y el ensamblado de los paquetes de datos, la corrección de errores y el control de flujo.
4. Subcapa de Gestión de Recursos Radio (RRM, Radio Resource Management): analiza el control de los servicios ofrecidos por la capa de acceso al medio, gestiona los recursos radio de forma eficiente, establece, mantiene y libera el canal radio, así como el control de la QoS en el que se incluyen los algoritmos de asignación de recursos (planificación o scheduling).

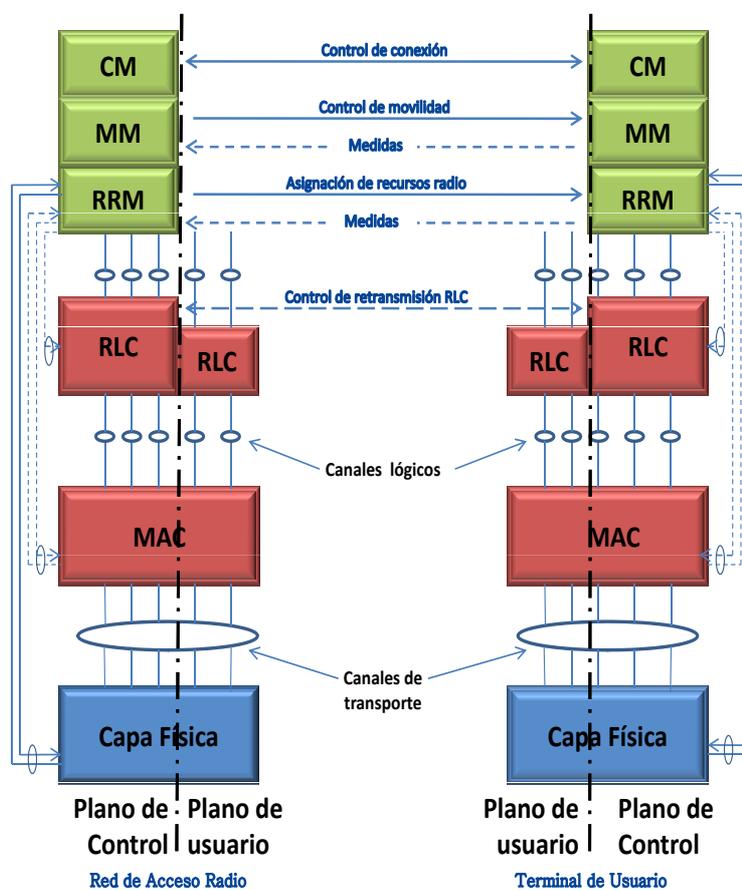


Figura 1.1: Capas y subcapas de la interfaz radio

5. Subcapa de Gestión de Movilidad (MM, Mobility Management): da soporte a la movilidad del usuario que no están relacionados con las funciones del canal radio, también realiza funciones de verificación de identidad de terminal/usuario así como procesos de autenticación.
6. Subcapa de Gestión de Conexión (CM, Connection Management): establece, controla y finaliza la conexión, gestiona el encaminamiento de la llamada/sesión, también se ocupa de los procedimientos de adaptación de la llamada a las características de las redes a las que accede.

Con la introducción de la fase 2G+ de GSM (2.5G), en concreto, con la

incorporación de GPRS, se define e implementa una nueva arquitectura de los sistemas móviles celulares y las funciones básicas de la interfaz radio para el transporte de los paquetes de datos. Así, en [Brasche and Walke (1997)] y también en [Cai and Goodman (1997), Hoff et al. (1998), Bettstetter (1999)] se realiza una descripción detallada de la nueva arquitectura, los servicios portadores y suplementarios, la calidad de servicio y sus parámetros, la gestión de la movilidad, la gestión de la sesión y el encaminamiento. En la nueva arquitectura del sistema se incorporan los nodos SGSN (Serving GPRS node) y GGSN (Gateway GPRS node) como elementos clave que hacen posible la conmutación de paquetes de datos e incorporan funciones primordiales como:

- la entrega de paquetes de datos desde y hacia las estaciones móviles en su área de servicio;
- el encaminamiento de los paquetes, la gestión de la movilidad, la gestión del enlace lógico y las funciones de autenticación;
- el registro de la localización de los usuarios y de los perfiles de los usuarios;
- el encaminamiento de los paquetes en la interfaz entre el *backbone* de la red y la red externa de paquetes de datos.

Es de destacar que estos elementos continúan siendo nodos primordiales de las redes actuales. En particular, en [Bettstetter (1999)] y relacionado con la interfaz radio, se describen los canales lógicos en GPRS, así como el mapeado de los canales lógicos en los canales físicos, la gestión de los recursos radio y los esquemas de codificación del canal. También en este trabajo se proporciona una visión de la arquitectura de protocolos, tanto del plano de señalización como del plano de transmisión. Incluye la interfaz del backbone, la interfaz SGSN-GGSN, el SNDCCP (Subnetwork Dependent Convergence Protocol, realiza las funciones de multiplexación de varias conexiones de la capa de red en una conexión de la capa LLC subyacente y la compresión/descompresión de los datos de usuario e información redundante de las cabeceras), la capa DLC (Data Link Layer) que incluye la subcapas RLC y MAC, la capa física y la interfaz BSS-SGSN (Base Station System-SGSN). Un último punto a destacar es la inter-operabilidad (*interworking*) con redes IP también evaluado y definido en [Park (2002)] que proporciona una visión de la coexistencia entre distintas tecnologías que proporcionan los

servicios adecuados al usuario y que conduce al objetivo de la convergencia de las redes móviles.

En el trabajo de [Kalden et al. (2000)], aparte de la descripción detallada de la tecnología y, en particular, de la interfaz radio y de la capa de protocolos de GPRS, además, se examinan las aplicaciones WWW sobre GPRS por medio de simulaciones del modelo de red en el que se obtienen medidas de rendimiento del caudal conseguido.

En el estudio y análisis de la capa de enlace, LLC, del sistema GPRS podemos encontrar trabajos[Ludwig et al. (1997), Demetrescu (1997)] dedicados al rendimiento de los parámetros que caracterizan la capa RLC/MAC. En concreto en [Ludwig et al. (1997)] se trata con especial interés la evaluación analítica del rendimiento del enlace de bajada de GPRS para distintos esquemas de codificación, fijando el tamaño de los paquetes IP y obteniendo el retardo producido con una alta carga de tráfico. También se deduce una expresión para obtener el retardo de la trama LLC y el cálculo de la probabilidad de retransmisión de tramas RLC/MAC. Ahondando en este último aspecto en [Demetrescu (1997)] se realiza un análisis por medio de esquemas ARQ (Automatic Repeat-Request) selectivos con el objetivo de gestionar los escasos recursos radio y calcular el número óptimo de retransmisiones de las tramas RLC/MAC en función del número de errores en el enlace radio y del tamaño de las tramas. Así mismo proporciona una expresión cerrada para el cálculo de la probabilidad de retransmisión de las tramas LLC. Otros trabajos relacionados con la capa de enlace de GPRS [Ajib et al. (1999), Queseth et al. (1999), Ameigeiras et al. (2000)] presentan mecanismos que permiten decrementar el número de bloques de control de reconocimiento de la subcapa RLC y así reducir el retardo requerido para una entrega de paquetes de datos[Ajib et al. (1999)]. También se muestran algoritmos de adaptación del enlace, eligiendo el mejor esquema de codificación para la transmisión de los datos evaluando tanto la BLER (Tasa de error de bloques) como la relación C/I (Portadora/Interferencia), obteniendo un mejor comportamiento en el throughput que con cualquier esquema de codificación fijo[Queseth et al. (1999)]. No obstante en [Ameigeiras et al. (2000)] se insiste en proponer un algoritmo de adaptación del enlace en una estimación de la BLER calculada a partir de los mensajes de reconocimiento enviados por el receptor, de esta forma la frecuencia de reconocimientos influye en la adaptación del enlace y, por tanto, se puede estimar el throughput medio del usuario. En cuanto a la gestión de la movilidad (MM), diversos trabajos estudian la incorporación de mecanismos de

gestión de la movilidad, a modo de ejemplo Cellular IP junto con GPRS [Ramirez et al. (2002)a].

La evolución de la interfaz radio de GSM/GPRS se presenta de forma detallada en [Furuskär et al. (1999), Molkdar et al. (2002)] y también en [Itkonen et al. (2002)]. Debido al crecimiento explosivo de Internet y la subsecuente demanda de las comunicaciones de datos haciendo uso de dispositivos inalámbricos, esta red móvil evoluciona al sistema EDGE proporcionando la conmutación de circuitos para las comunicaciones de voz mediante el sistema ECSD (Enhanced Circuit Switched Data) y la conmutación de paquetes de datos por medio del EGPRS (Enhanced General Packet Radio Service). En concreto, en el trabajo [Furuskär et al. (1999)] se desarrolla el sistema EDGE, desde la interfaz radio hasta los protocolos utilizados y los servicios portadores ofrecidos por esta tecnología. EGPRS utiliza modulación 8-PSK para los esquemas de codificación del MCS-5 al MCS-9 para permitir tasas de datos más altas que GPRS. Las cabeceras de los radio bloques en EGPRS se codifican de forma robusta. La red EGPRS puede seleccionar de forma automática el esquema de codificación más apropiado dependiendo de las condiciones del enlace radio con el objetivo de maximizar el throughput. En cuanto al rendimiento de EGPRS se estudia el retardo en función del tamaño de paquete utilizado y la tasa media de bits por slot de tiempo, permitiendo tasas de pico más altas con respecto al estándar GSM.

La eficiencia del enlace en EGPRS se estudia, en general, en términos de la tasa de errores de radio-bloques (BLER) y el caudal (throughput) por time slot para los diferentes esquemas de codificación y modulación. La mejora de la capacidad de los datos en EGPRS se fundamenta en las mejoras del nivel de enlace y, por tanto, es de importancia su estudio. En [Molkdar et al. (2002)] se presenta una visión general de EDGE y, en particular, se realiza un especial énfasis en la interfaz aire de su componente EGPRS, tanto de la capa física como la capa LLC (RLC/MAC). En concreto es de destacar que la nueva segmentación de los bloques RLC permite la retransmisión de los bloques erróneos RLC con esquemas de codificación más robustos. También aumentando el tamaño de ventana mayor al utilizado en el proceso ARQ selectivo se llega a reducir la transmisión de bloques inútiles en el protocolo RLC. En el caso de [Ball et al. (2004)] se introduce un nuevo concepto que es la prevención de retransmisiones frecuentes en RLC, donde el emisor trata de evitar las retransmisiones de bloques RLC con el estado de reconocimientos pendientes, para ello realiza un polling rápido y con esta estrategia se logra mejorar el throughput de las comunicaciones en EDGE. Otras mejoras propuestas en

el diseño de la capa de enlace RLC/MAC, es la de [Qiu et al. (2000)], se resumen en propuestas como la realización de un acceso en el enlace de subida más rápido y la asignación de recursos con un nuevo conjunto de canales de control de paquetes y un nuevo identificador de petición de acceso.

1.2.3.1. Evolución de la interfaz radio de GERAN: EDGE2

GERAN (GSM/EDGE Radio Access Network) representa hoy en día el núcleo ó backbone de las comunicaciones móviles de casi tres mil millones de abonados en todo el mundo, proporcionándoles acceso e itinerancia para los servicios de voz y paquetes de datos. La evolución de GERAN es un paso importante para los operadores que desarrollan GSM, GSM/WCDMA y GSM/LTE con el objetivo de soportar los servicios de forma continua. Esta evolución incluye mejoras significativas en las tasas de datos de los usuarios, la eficiencia del espectro y la cobertura, al mismo tiempo que evita serios problemas para dar servicio en las redes GSM existentes. GERAN consta de una estructura de protocolos por capas y el nuevo concepto EGPRS2 en el enlace de subida [Saily et al. (2008)] afecta solamente a los protocolos de la interfaz radio bajo la capa LLC, por tanto, EGPRS2 no requiere nuevas ubicaciones para las estaciones base, ni un nuevo espectro de radio frecuencia, así como tampoco sobre la planificación de células y frecuencias. En el citado trabajo [Saily et al. (2008)] se presenta el concepto de alto rendimiento del enlace ascendente para la evolución de GERAN (EGPRS2 UL) cuya estandarización es sujeto de estudio por el 3GPP, y en el que se presentan los aspectos clave de transmisión como son, por un lado, la utilización de un pulso de espectro ensanchado, optimizado para conseguir los requerimientos de interferencia existentes en la red y, por otro lado, el pulso GMSK lineal existente para escenarios donde el pulso ensanchado no pueda ser ubicable. Los valores de los anchos de pulsos se especifican según reglas de planificación de frecuencia de la red donde los niveles de interferencia co-canal dominan sobre los niveles de canal adyacente.

EGPRS2 en combinación con el enfoque de utilización de doble portadora (dual-carrier) en el enlace de bajada [Ivanov et al. (2008)] rompe la posibilidad de ofrecer mejoras de tasas de datos que duplican y triplican las tasas conseguidas con la utilización de una única portadora en EDGE. Como consecuencia, en las redes actuales se hace necesaria la gestión de los recursos radio de forma inteligente así como la utilización del control de calidad del enlace radio para implementar dinámicamente las asignaciones de las doble-

portadoras sobre los transceptores. En el trabajo de [Ivanov et al. (2008)] se aporta una visión general de la arquitectura de la evolución de GERAN y el modelo de simulación de la red incluyendo el nuevo enfoque dual-carrier de GERAN, obteniendo valores de tasas de 350 a 360 Kbps con terminales de usuarios utilizando 8 time slots y en condiciones óptimas hasta tasas de pico por encima de los 500 Kbps.

De forma paralela en la estandarización de nuevas tecnologías en la interfaz radio, para el acceso radio de los usuarios se abandona el acceso tipo CDMA característico de UMTS y aparece la utilización del concepto SDR (Software Defined Radio) para optimizar el acceso radio. De forma genérica la red completa prevista en todas las subredes implican el concepto "todo IP". Las tasas de pico máximas previstas son de 100 Mbps en enlace descendente y 50 Mbps en enlace ascendente (con un ancho de banda en ambos sentidos de 20Mhz). Los nodos principales dentro de esta implementación son el Evolved Node B (BTS evolucionada) y el System Access Gateway que actuará también como interfaz a Internet, conectado directamente al Evolved Node B. El servidor de los recursos radio, RRM, será otro componente utilizado para facilitar la inter-operabilidad con otras tecnologías.

1.2.4. Introducción a la calidad de servicio en redes celulares

La evolución de la gestión de la calidad de servicio, QoS (Quality of Service), para las redes móviles celulares comenzó con la introducción de GPRS. La principal característica en el marco de gestión de QoS fue el concepto de GPRS como portador de servicio (BS, Bearer Service) que se basa en el concepto del contexto PDP (Packet Data Protocol) que es, en esencia, una configuración de conexión lógica para transportar el tráfico IP (también X.25 en sus orígenes) entre la estación móvil y el gateway (GGSN, Gateway GPRS Support Node). Los estándares del 3GPP han desarrollado el concepto del PDP para soportar nuevos requisitos, de tal forma que la incorporación del PDP en GPRS permitía que una estación móvil tuviese múltiples contextos PDP pero donde todos ellos necesitaban una dirección PDP dedicada. En versiones posteriores se consigue que coexistan varios contextos PDP por cada dirección PDP, cada uno de ellos teniendo su propio perfil de calidad de servicio. El primer contexto PDP abierto por la misma dirección PDP se denomina contexto primario y los subsiguientes contextos se nombran como PDP secundarios. No obstante, la utilización de contextos secundarios necesitan que estén conectados al mismo nombre de punto de acceso (APN, Access

Point Name) junto con el contexto primario. La base de esta arquitectura se fundamenta en la diferenciación del portador de calidad de servicio realizada sobre la base del contexto PDP, esto quiere decir que toda aplicación de un usuario que comparta el mismo contexto PDP posee los mismos atributos de QoS.

Desde un punto de vista del operador móvil, la gestión de la calidad de servicio extremo a extremo es un aspecto de interés para la gestión eficiente de la propia red. Una manera opcional de resolver el problema de gestionar redes complejas es administrando desde un lugar de control centralizado la QoS con el objetivo de conseguir un comportamiento del servicio que corresponda a las especificaciones de calidad acordadas. Así el control centralizado de la QoS se ha de preocupar de la planificación y dimensionado de la red, el servicio ofrecido extremo a extremo, el tráfico e indicadores claves de rendimiento de la red.

1.2.4.1. Parámetros de calidad de servicio en redes celulares

Los parámetros o atributos de QoS que se negocian en un contexto PDP entre la red y la estación móvil determinan el perfil de calidad de servicio que se ofrece al usuario. Los requisitos de calidad de servicio definidos a través de los parámetros de QoS en Release 97/98 del 3GPP y [Bilgic et al. (1999)], son las clases: precedencia, retardo, fiabilidad y throughput (tabla 1.1).

La clase precedencia indica la importancia de mantener el servicio acordado bajo condiciones adversas, por ejemplo qué paquetes se han de descartar si los recursos disponibles disminuyen o se produce congestión en la red. La clase retardo define el retardo de los paquetes de datos por el tránsito de estos a través de la red. La combinación de los modos de transmisión de las distintas capas de protocolos de red soporta los requisitos de rendimiento por medio de la clase fiabilidad. El caudal de datos de usuario se especifica por medio de un conjunto de clases de throughput que caracterizan el ancho de banda requerido para determinados tipos de tráfico.

Con la estandarización de UMTS y, en concreto, en [3GPP Release'99] se introducen los nuevos parámetros o atributos que atienden a las cuatro clases de QoS: Conversacional, Streaming, Interactiva y Background. Estos nuevos parámetros o atributos (tabla 1.2) de QoS contribuyen a la distinción de las cuatro clases y cuyo principal factor para distinguir las clases conversacional, streaming, interactiva y background, es el hecho de cuán sensibles son al retardo y a la variación del retardo. La clase conversacional se caracteriza por

Parámetro de QoS	Descripción
Clase Precedencia	Se definen tres niveles de precedencia del servicio que indican la prioridad para mantener el servicio. El nivel alto de precedencia garantiza el servicio frente a los otros niveles. El nivel normal de precedencia mantiene el servicio por encima de los usuarios con baja prioridad. El nivel bajo de precedencia soporta el servicio después de satisfacer el servicio para los usuarios con prioridad alta o normal.
Clase Retardo	El parámetro de retardo se refiere al retardo extremo a extremo observado en el sistema GPRS. Se especifican tres clases de retardo y una clase best effort.
Clase Fiabilidad	La fiabilidad se define en función de la tasa de error residual por medio de: la probabilidad de pérdida de datos, la probabilidad de datos entregados fuera de secuencia, la probabilidad de datos duplicados en su entrega y la probabilidad de datos corruptos.
Clase Throughput	El caudal de los datos de usuario se especifica por medio de las clases que caracterizan el ancho de banda esperado requerido para un contexto PDP. El throughput se define bien como de pico (tasa máxima a la cual los datos pueden ser transportados a través de la red en un contexto PDP) o bien como throughput medio (valor de tasa media a la cual los datos se espera que se transfieran durante el tiempo de activación de un contexto PDP).

Cuadro 1.1: Parámetros de QoS en 3GPP Release'97/98

utilizar un tráfico altamente sensible al retardo para aplicaciones del tipo voz sobre IP ó vídeo sobre IP. En cambio la clase streaming se utiliza fundamentalmente para el transporte en tiempo real y también sensibles al retardo, sin embargo, en las aplicaciones de audio y vídeo utilizadas en esta clase, se puede compensar la variación del retardo por medio de mecanismos de buffering (memorias locales) en la parte del usuario. Las clases interactiva y background se especifican generalmente para aplicaciones como la navegación en web y, en general, la transmisión de datos como el correo electrónico o la transferencia de archivos. La principal diferencia entre la clase interactiva y la clase background, ambas menos sensibles al retardo, es que la primera de ellas necesita una respuesta del usuario mientras que en la segunda este hecho no es necesario y además se realiza como una tarea en el que el tiempo no es un factor crítico como es, por ejemplo, la recepción de correo electrónico o la transferencia en el enlace de bajada de uno o varios archivos. En estas dos últimas clases se mejora la relación de tasa de errores debido principalmente a la utilización de los esquemas de retransmisión de datos.

Cada parámetro de QoS en el contexto PDP ha de solicitarlo la estación móvil bien por medio de valores ya definidos o bien con valores que le asigna la propia red. Este proceso es el denominado de negociación de los atributos de QoS del contexto PDP. En el supuesto que el terminal de usuario requiera valores suscritos desde la red de los parámetros clase de tráfico o tasa de bits garantizada, solamente se permiten por parte de la red los valores acordados para la clase interactiva y la clase background. En el caso de que la estación móvil solicite cualquier otro valor de parámetro en la activación del contexto PDP, los valores se asignan desde la red, en concreto del HLR (Home Location Register), en este elemento de la arquitectura de la red se pueden configurar los parámetros de QoS para cada abonado. En el caso en que la estación móvil solicite valores numéricos para los atributos de QoS, el propio perfil almacenado en el HLR proporciona para cada abonado los límites con los que cuenta en combinación con el punto de acceso (APN). El perfil de QoS utilizado para ciertos abonados es el mismo para cada uno de ellos, su tratamiento es, por tanto, en función del perfil de usuario establecido para ese conjunto de usuarios.

Parámetro de QoS	Descripción
Clase de Tráfico	Se especifica el tipo de aplicación para la que se optimiza el servicio portador de acceso radio (RAB, Radio Access Bearer). Se refiere a las clases: Conversacional, Streaming, Interactiva y Background.
Orden de entrega	Especifica si el portador proporcionará ó no el servicio en secuencia de las unidades de datos, SDU.
Tamaño máximo de SDU	Define el tamaño máximo permitido de las unidades de datos (SDU), en octetos.
Información del formato SDU	Define los posibles tamaños de las unidades de datos en bits).
Entrega de SDUs erróneas	Indica si las unidades de datos con errores pueden ser entregadas ó no.
BER residual	Especifica la tasa de errores de bits en cada flujo para la entrega de las unidades de datos.
Tasa de errores SDU	Define la relación entre las unidades de datos perdidas o detectadas como erróneas frente a las transmitidas.
Retardo de transferencia	Se define como el tiempo desde que una petición de transferencia de una SDU aun punto de acceso al servicio hasta que se realiza la entrega al otro punto de acceso al servicio, en milisegundos. Indica el retardo máximo para el percentil 95 de la distribución de retardo en todas las unidades de datos entregadas durante el tiempo de vida del servicio portador.
Tasa máxima de bits	Define la tasa máxima de los bits entregados a un punto de acceso al servicio durante un período de tiempo dividido por la duración de dicho período de tiempo (en Kbps).
Tasa garantizada de bits	Indica el número garantizado de bits entregados a un punto de acceso al servicio dentro de un período de tiempo dividido por dicho período de tiempo (en Kbps).
Prioridad de manejo de tráfico	Especifica la importancia relativa para gestionar las unidades de datos pertenecientes al RAB comparado con las unidades de datos de otros portadores.
Prioridad de asignación/retención	Especifica la importancia relativa comparada con otros RABs para la asignación y retención del portador de acceso radio. El atributo de prioridad de asignación/retención es un parámetro del abonado que se negocia desde el equipo terminal.

Cuadro 1.2: Parámetros de QoS definidos para GERAN y UTRAN en 3GPP Release'99

1.2.5. Mecanismos de soporte a la calidad de servicio en redes celulares

Los mecanismos de soporte a la calidad de servicio, tanto en GERAN como en UTRAN, se dividen en dos planos distintos: el plano de control y el plano de usuario. Durante el procedimiento de establecimiento del servicio, las funciones del plano de control tienen en cuenta las funciones propias del control de admisión y del establecimiento de la conexión junto con los mecanismos de reserva de QoS que implica la gestión de los recursos y el control de carga y handover. Los mecanismos del plano de usuario tienen en cuenta la provisión de la QoS una vez establecida las condiciones de conexión, esto supone el acondicionamiento del tráfico, la adaptación del enlace, la planificación de los paquetes de datos y el control de potencia para la reducción de los niveles de interferencia.

Los mecanismos de QoS en el plano de control los enumeramos en los siguientes items:

- Control de admisión
- Mapeado del servicio portador radio
- Gestión de los recursos
- Control de carga y traspaso (handover)

Los mecanismos de QoS en el plano de usuario los dividimos en los siguientes aspectos:

- Adaptación del enlace
- Acondicionamiento del tráfico
- Planificación (*Scheduling*) de paquetes
- Control de potencia

1.2.5.1. Control de admisión

La función del control de admisión es la de aceptar o denegar el acceso a la red a nuevos usuarios. De esta forma este mecanismo trata de evitar situaciones de congestión basándose en medidas de interferencia y disponibilidad de los recursos. El control de admisión se realiza en el acceso inicial del terminal de usuario cuando establece la conexión, también en el traspaso (handover) y en la reconfiguración del portador del servicio de acceso a la red o de la creación y modificación del contexto de flujo de paquetes (PFC, Packet Flow Context) en el subsistema de la estación base (BSS). Cada uno de estos eventos puede dar lugar a distintas respuestas dependiendo de la prioridad del usuario y de la situación de carga de la red.

En el trabajo de [Tsao (2000)b] así como en el de [Kochem et al. (2003)] se proponen sistemas de control de admisión para GPRS que se apoyan en otros mecanismos de soporte a la QoS para este tipo de redes como es: la gestión de los recursos radio y la planificación de paquetes. En particular en [Tsao (2000)b] se realiza por medio de simulaciones la evaluación del mecanismo de control de admisión en una primera fase de establecimiento de la conexión, consiguiendo satisfacer los requisitos de QoS especificados en el contrato de servicio. Por otro lado, en [Kochem et al. (2003)] se propone un algoritmo para GPRS que proporciona las garantías de QoS en la fase de establecimiento de la conexión. Dicho algoritmo se basa en información local dentro de la celda en la que se encuentra el terminal móvil y también de información remota procedente de las celdas adyacentes para poder determinar si se puede o no aceptar la conexión entrante. Con la introducción de una estrategia apropiada de QoS común para GPRS/EDGE y UMTS, en [Müllner et al. (2004)] se realiza un estudio de los mecanismos de QoS que pueden proporcionar beneficios significativos a los usuarios en distintas condiciones de carga de la red. Así se realiza mediante simulaciones el mecanismo de control de admisión regulando el acceso al servicio de la red, admitiendo en función de las clases y los atributos de QoS las conexiones de los terminales de usuario. El control de admisión hace uso particularmente del atributo "prioridad de asignación/retención", de forma que si una conexión es admitida por parte de la red, los usuarios con alta prioridad pueden mantener su conexión frente a otros usuarios con menor prioridad en caso de que exista sobrecarga en la red.

1.2.5.2. Mapeado del servicio portador radio

Los requisitos de QoS que ha de satisfacer la red, han de lograrse a través de los portadores radio. Este mecanismo se consigue seleccionando las condiciones adecuadas de transmisión sobre los canales lógicos elegidos, además de, entre otros factores, los modos de transmisión (con reconocimiento o sin reconocimiento), las modulaciones, los esquemas de codificación y el tipo de corrección de errores. Las retransmisiones de las unidades de datos permite la mejora del atributo de QoS fiabilidad. Este parámetro en la transmisión de los paquetes de datos es fundamental para un transporte eficiente en muchas de las aplicaciones de datos (por ejemplo: e-mail, transferencia de archivos). El atributo clase de tráfico permite seleccionar las condiciones del tráfico además de otros atributos.

Tanto los servicios basados en las clases interactiva y background requieren el uso de RLC en modo con reconocimiento de la transmisión con el objetivo de proporcionar un transporte de la información fiable a expensas de introducir un retardo y throughput variable.

Para los servicios basados en la clase streaming se pueden soportar en modo con reconocimiento o sin reconocimiento de las unidades de datos. El hecho de solicitar una alta fiabilidad en la transmisión de la información por medio de retransmisiones supone el incremento del retardo o la variación del retardo. Frente a esta solución es posible limitar el número de retransmisiones haciendo uso de un esquema de codificación más robusto, lo cual tiene un efecto directo sobre el throughput logrado. En el caso de utilizar el modo de transmisión sin reconocimiento, dependiendo de las condiciones del enlace se puede adaptar el enlace hacia un esquema de codificación que aumente el rendimiento del throughput.

Para las aplicaciones basadas en la clase conversacional normalmente se realizan en modo de transmisión sin reconocimiento. En la interfaz radio haciendo uso de la conmutación de paquetes de datos se satisfacen los requisitos haciendo uso del PDCP (Packet Data Convergence Protocol) que realiza la adaptación de las cabeceras IP/UDP/RTP.

1.2.5.3. Gestión de los recursos

Un mecanismo clave en la interfaz radio de cualquier red de comunicación es la gestión de los recursos radio. Dada la importancia para la presente tesis de este mecanismo se tratará en el capítulo siguiente de forma detallada.

1.2.5.4. Control de carga y traspaso (handover)

Los traspasos o handover se pueden producir básicamente por tres causas: la primera de ellas por el propio movimiento del terminal móvil desde una área geográfica a otra y, por tanto, su comunicación ha de mantenerse en el cambio de una celda a otra adyacente; la segunda de las razones es por condiciones deficientes de la comunicación en la celda actual; y la tercera de las razones es porque se produce congestión en la celda actual. El mecanismo de control de carga supervisa el tráfico en las celdas que conforman la red para evitar, en la medida de lo posible, situaciones de congestión o sobrecarga. El control de carga actúa de forma que si la red detecta situaciones de sobrecarga realiza traspasos de las comunicaciones a celdas adyacentes y, por tanto, ha de seleccionar qué conexiones son las que ha de trasladar. En función de los perfiles de QoS acordados previamente, se tomarán decisiones por parte de la red para realizar los traspasos, así los usuarios con perfiles de baja prioridad serán los primeros en ser trasladados.

1.2.5.5. Adaptación del enlace

Los enlaces en las comunicaciones móviles se ven afectados por las condiciones del enlace radio, esto quiere decir que los valores de parámetros como el retardo, el caudal o la tasa de errores están variando continuamente. El mecanismo de QoS que trata de satisfacer los requisitos de QoS ante las variaciones de estos parámetros se denomina la adaptación del enlace. Estas variaciones se compensan eligiendo los esquemas de codificación y modulación más adecuados a las circunstancias. En malas condiciones del enlace radio la elección es de los esquemas de codificación más robustos que aumentan la redundancia y, por tanto, se puede mejorar la corrección de errores que por otro lado reduce el caudal resultante.

Distintos algoritmos [Queseth et al. (1999), Gozalvez et al. (2004)] se usan para la adaptación del enlace con el objetivo de seleccionar los esquemas de codificación más adecuados en función de las condiciones del enlace radio. Esta selección del esquema de codificación se basa en las estimaciones de calidad del canal. En un mecanismo típico de adaptación del enlace, el receptor mide la relación C/I (portadora/interferencia)[Gozalvez et al. (2000)] o la tasa de bits erróneos y realiza una estimación de la calidad del canal promediando las medidas obtenidas en un intervalo de tiempo determinado.

La retransmisión de datos es eficiente en EGPRS fundamentalmente por

la utilización de la redundancia incremental combinada con la adaptación del enlace. Tanto en [Featherstone et al. (2001), Molldar et al. (2001)] como en [Furuskar2 et al. (1999)] se introduce la redundancia incremental como mejora del rendimiento de la capa física a través del modo con reconocimiento de la subcapa RLC. La redundancia incremental consiste en el ajuste de la tasa de datos a las condiciones reales del canal por medio de información redundante en la transmisión. Los esquemas de modulación y codificación pueden cambiarse en las retransmisiones subsiguientes.

Trabajos como los de [Leung et al. (2001)] realizan un estudio de cómo combinar la adaptación del enlace y el control de potencia para lograr una tasa de errores menor y reducir la variación de errores aumentando la calidad del transporte, en concreto, de música codificada MPEG-4. También en [Kodikara et al. (2004)a] se realiza un análisis de la adaptación del enlace para la clase streaming sobre canales EGPRS. Al mismo tiempo en [Kodikara et al. (2004)b] investigan el desarrollo de técnicas para la mejora del rendimiento de las comunicaciones de vídeo en tiempo real proponiendo y haciendo un análisis de un nuevo algoritmo que mejora el enlace radio.

1.2.5.6. Acondicionamiento del tráfico

El hecho de transmitir paquetes por la interfaz radio implica que cada paquete de datos experimente efectos distintos, el mecanismo de acondicionamiento de tráfico vela por la conformidad entre la QoS negociada entre la red y el usuario y, según el servicio acordado, el tráfico experimentado por los datos. Esto implica la necesidad del control del tráfico a través de cuatro funciones básicas: control de policía (policing), marcado (marker), conformado (shaper) y eliminación de paquetes (dropper). El primero de los mecanismos, policing comprueba que los paquetes de datos correspondan al perfil de QoS y se reenvía, se marca, se conforma o se elimina en función, todo ello, del perfil del flujo de datos al que pertenezcan los paquetes. El marcador es el responsable de etiquetar los paquetes de acuerdo al trato que debe tener cada uno de ellos.

La estación móvil se puede observar en la figura 1.2 como un conjunto formado por un equipo terminal, TE, y un terminal móvil, TM. El TE es un host Internet que intercambia mensajes de la capa de aplicación junto con mensajes de señalización que transportan requerimientos de calidad de servicio con otros nodos. El mapeador/traductor juega dos papeles: el primero es la traducción de los parámetros de calidad de servicio de las redes externas y

los atributos de calidad de servicio de la red que determinan las características de los contextos GPRS; y el segundo papel es el de clasificar los paquetes en la interfaz de ingreso y la selección del contexto PDP correcto para transportar cada paquete. Otro elemento es el monitor cuya función es la de medir la cantidad de tráfico y sus características. La salida de este elemento se utiliza por la función policial para determinar sus acciones asegurando que el tráfico en la interfaz de ingreso no exceda del perfil de la calidad de servicio negociada. Esta función ha de ser configurable de forma que pueda tomar distintas acciones según los tipos de paquetes.

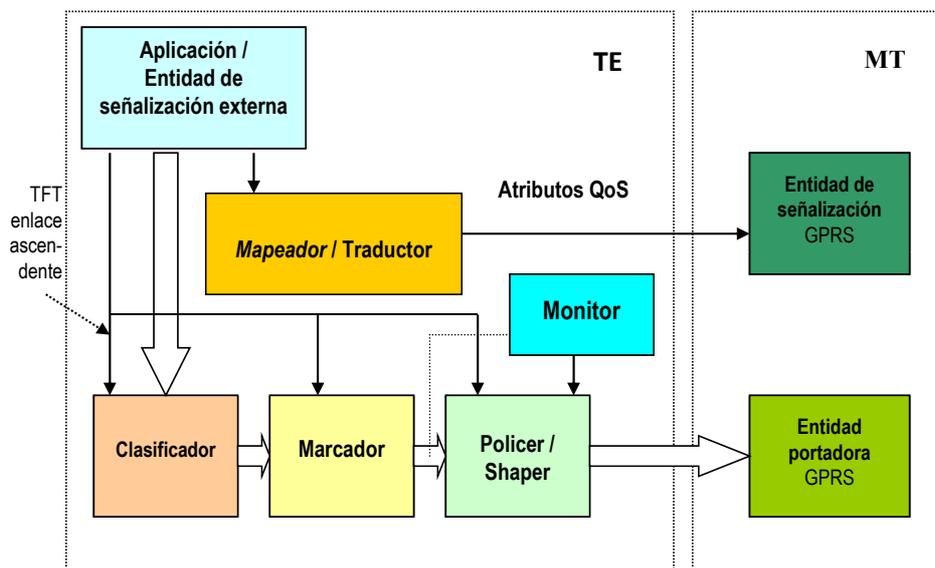


Figura 1.2: Elementos de QoS en el equipo terminal

De forma alternativa, un protocolo de señalización como es el caso de RSVP (Resource Reservation Protocol)[Ramírez et al. (2000)] se puede utilizar para intercambiar parámetros de calidad de servicio para una sesión determinada. Independientemente del método de señalización, el TE se espera que determine la calidad de servicio y los TFTs para el tráfico en el enlace descendente.

El proceso de traducción de los requerimientos de calidad de servicio externos señalados por la entidad de aplicación o una entidad de señalización especializada, como es el caso de RSVP, en atributos de calidad de servicio

en GPRS se realiza por el TE. Por otro lado, los nodos de GPRS realizan operaciones específicas para proporcionar el servicio de acuerdo con los atributos de calidad de servicio señalados por parte del TE. Estos atributos los gestiona el TE tanto para el enlace ascendente como para el descendente. Dado que el equipo terminal realiza la clasificación de todos los paquetes para el tráfico en el enlace ascendente, no es necesario una TFT en el enlace ascendente. La TFT en el enlace descendente se determina por medio del TE y se le indica al GGSN. El intercambio de señalización entre TE y GGSN es parte de la activación del contexto PDP o renegociación. El TE determina cuándo activar o renegociar un contexto PDP.

1.2.5.7. Planificación (*Scheduling*) de paquetes

La planificación de los paquetes es uno de los mecanismos básicos en la provisión de la calidad de servicio. Desde el punto de vista de la comunicación en la interfaz radio, la gestión eficiente de elegir cómo y en qué orden se han de reenviar los radio bloques o paquetes que llegan a un nodo de la red, se convierte en un factor clave para evitar los cuellos de botella que se puedan generar, además de satisfacer los requisitos acordados del servicio proporcionando justicia entre los flujos de datos. Para ello los algoritmos de planificación (*scheduling*) han de tener en cuenta los perfiles de QoS con el objetivo de dar prioridad entre los servicios y los usuarios con distintos perfiles de abonados, teniendo en cuenta entre otros parámetros, el atributo de prioridad de asignación/retención. Así, un servicio de clase conversacional tendrá la máxima prioridad en la política de planificación. Los servicios de la clase *streaming* necesitan ciertas garantías de calidad de servicio en cuanto al *throughput*, asegurando cierto ancho de banda para los canales que soportan este tipo de servicio. En cambio para las clases interactiva y *background* aprovecharán ciertos espacios temporales de silencio para reenviar los datos disponibles.

Distintos algoritmos de scheduling para GPRS, EGPRS y redes inalámbricas, en general, han sido analizados [Ramirez et al. (2002)b] en la literatura científica. Desde la planificación de paquetes FIFO (First In First Out) ó FCFS (First Come First Served) pasando por el algoritmo LBFS (Least Bits left First Served)[Cavalcanti et al. (2002)] donde los paquetes más largos pueden tolerar mayores retardos y los paquetes pequeños podrían transmitirse primero; el algoritmo LTFS (Least Time left First Served) en el que se intenta estimar el tiempo que resta antes de que éste expire, siguiendo el

criterio de QoS para cada paquete y transmitiendo el paquete con el margen más corto para satisfacer los requisitos de QoS, de tal forma que aquellos paquetes que se estiman que no puedan conseguir la QoS demandada se les asigna una prioridad menor [Johanson et al. (1998)]. Otros algoritmos como el RR (Round Robin) [Calin et al. (1999), Tang et al. (2004)] ó el SPS (Static Priority Scheduling), el EDF (Earliest Deadline First) y el MED (Modified Earliest Deadline) han sido evaluados en [Sau et al. (1998), Pang et al. (1999)] obteniendo comparativas entre los distintos esquemas de planificación. En [Tsao (2000)a] los autores argumentan cómo desafortunadamente los algoritmos GPS (Generalized Processor Sharing) y EDD (Earliest Due Date) no pueden aplicarse directamente en las redes inalámbricas debido a los errores producidos por la localización de los terminales en entornos móviles, así que exploran políticas de planificación como extensión de las anteriores como son el Delay-EDD y el Jitter-EDD, demostrando que se pueden conseguir garantías en los parámetros de QoS a corto y largo plazo.

Otros algoritmos de planificación que gestionan los radio bloques en el sistema EGPRS, aparte de las políticas FIFO y RR, se evalúan en el esquema HRF (High Rate First) donde el planificador elige el flujo con la tasa de transmisión estimada más alta en función de la información de calidad del canal que proporciona el terminal móvil; por otro lado el algoritmo SRPT (Shortest Remaining Processing Time) en el que el planificador sirve el flujo con el tiempo de servicio restante más corto [Balachandran et al. (2001)]. También en [Cai et al. (2000), Chang et al. (2001)] se realizan simulaciones sobre el algoritmo WRR (Weighted Round Robin) haciendo uso de la capacidad multislots de los usuarios con requisitos de servicio prioritarios, concluyendo que no es un algoritmo efectivo para soportar servicios diferenciados. Esquemas como el algoritmo WFQ (Weighted Fair Queueing)[Lee et al. (2002)], ampliamente utilizado en redes fijas, ha sido propuesto para su utilización en combinación con mecanismos de intercambio de canales entre flujos de tráfico para satisfacer los requisitos de QoS en redes móviles.

[Stuckmann et al. (2003)] proponen y analizan dos algoritmos en dos niveles distintos de planificación basados en el Round Robin, el primero el DPARR (Displaced Pending Acknowledge Round Robin) y el segundo el LQDWRR (Link Quality-based Deficit Weighted Round Robin). Se basan en la utilización de la información de reconocimiento de los radio bloques en la capa RLC para la planificación de los TBF (Temporary Block Flow) y el uso de la información de la calidad del enlace para la planificación adaptativa de los TBF. Otro trabajo [Yang et al. (2001)] propone el algo-

ritmo DSCM (Dynamic Scheduling for Mobile Communications) basado en el estándar SCFQ (Self-Clocked Fair Queuing) que ajusta de forma dinámica la tasa de servicio para una conexión concreta en función de la calidad del canal. En [Bada et al. (2002)] evalúan los algoritmos MLT (Minimum Laxity Threshold), basado en el concepto de la cantidad de tiempo que el planificador queda inactivo para poderse dedicar a otros flujos menos prioritarios, y el algoritmo MED, concluyendo que el mejor comportamiento de planificación de los radio bloques corresponde al MLT frente al algoritmo MED, si bien la complejidad de implementación del primero es mayor. También [Cavalcanti et al. (2002)] proponen dos algoritmos de planificación demostrando que cuando se combina un esquema de reuso más ajustado en un sistema de estación base y un algoritmo de scheduling mejorado se puede obtener una ganancia de la capacidad de hasta el 450%. En [Vieira et al. (2006)] se presentan tres mecanismos del denominado EQoS (Enhanced QoS) control de admisión, planificación y control de calidad que trabajando de forma conjunta permiten el tratamiento diferenciado de las clases de servicio 3GPP. Introducen el concepto de comportamiento de la red por medio de los KPIs (key performance indicators), indicadores que cambian en distintos escenarios con distintos tipos de tráfico y niveles de carga de la red.

1.2.5.8. Control de potencia

El mecanismo de control de potencia, por un lado, se utiliza en el enlace ascendente para reducir el consumo de potencia del terminal móvil y, por otro lado, intenta mantener el nivel de recepción y la calidad recibida a través de reducir los niveles de interferencia mientras se mantiene la conexión establecida con el perfil de QoS acordado. Esto quiere decir que a través de niveles umbrales se va cambiando la potencia del transmisor para reducir el consumo de potencia. En [Simonsson et al. (2000)] se expone un trabajo en el que se combina las funciones correspondientes a minimizar la energía consumida por los terminales y las técnicas de planificación de los recursos. Así esta combinación de scheduling y control de potencia se utiliza para minimizar el uso de la energía o la interferencia emitida.

1.3. Objetivos de la presente tesis doctoral

Con el objetivo de lograr la optimización del sistema móvil celular EDGE (actualmente está en funcionamiento por un grupo numeroso de operadoras) así como mejorar la eficiencia de la asignación de los recursos en esta red móvil celular de radio paquetes, nos centramos en proponer nuevos modelos de forma analítica que permitan mejorar el rendimiento de la asignación de los canales a los usuarios mejorando las tasas de pico de las que disponen actualmente. Así en un entorno de referencia como es el sistema EDGE, la gestión de los recursos radio por parte de la estación base se centrará en la ordenación de estos y la localización de cada uno de ellos para utilizar distintas políticas que permitan maximizar el uso de estos recursos ofreciendo la calidad de servicio (QoS) requerida por los usuarios.

En los modelos que se proponen, utilizando distribuciones estadísticas que definan el comportamiento de los procesos de llegadas y servicio de los tipos de aplicaciones (voz, datos, imágenes, ...) y resolviendo los sistemas de tráfico correspondientes a cada modelo, proporcionamos tanto para los servicios de voz como para los servicios de paquetes de datos, nuevos métodos de asignación de canales que mejoran la transmisión de los paquetes de datos manteniendo el grado de servicio del tráfico de voz. Al utilizar diversas estrategias de asignación de los recursos podremos llegar a conclusiones que permitan deducir de forma rigurosa qué pasos ha de seguir el operador de red para planificar de forma adecuada la asignación de los canales de sus usuarios de acuerdo a los requerimientos de QoS de cada uno de ellos.

En la tesis que se propone además se sintetizarán los modelos propuestos por medio de expresiones analíticas cerradas que permitan obtener las probabilidades de encontrar un grupo de recursos radio consecutivos en el tiempo, de tal forma que puedan simplificar el método de resolución empleado y planteado actualmente en la literatura científica. Este hecho permite que puedan ser asignados los recursos a uno o varios usuarios para satisfacer los requisitos de los servicios del tráfico de voz y de datos que gestionan los operadores.

Por un lado, es necesario realizar estrategias de reserva de recursos que, tanto para las llamadas de la propia celda en la que se permanece como para aquellas que proceden de un traspaso de otras celdas y, en función de las prioridades que se estimen para los distintos tráficos, bien para el servicio de voz o bien para el de los datos, se proporcionen resultados satisfactorios

que se correspondan con los parámetros de QoS establecidos previamente. No obstante es necesario plantear que el tráfico de datos puede disponer de sistemas con buffering que, en concreto, correspondan a modelos de espera y pérdida con colas limitadas para minimizar la pérdida de paquetes.

Capítulo 2

Modelado de los recursos radio en redes celulares de paquetes de datos

2.1. Introducción

Uno de los mayores retos para los sistemas de comunicaciones móviles en las nuevas generaciones es la de proporcionar servicios a distintos tipos de terminales y aplicaciones. El objetivo principal es el de asegurar la movilidad sin hilos para el usuario y esto se traduce en la necesidad de utilizar distintas estrategias que puedan asegurar que los recursos siempre estén disponibles tanto para la red como para los terminales móviles. Los esquemas de asignación de los recursos junto con la planificación de estos suponen la mejora de la comunicación entre los abonados de la red móvil. La gestión eficiente de los recursos radio es necesaria con el objeto de proporcionar la calidad de servicio demandada por las aplicaciones utilizadas en los terminales móviles. Las redes móviles se caracterizan por una topología dinámica y los enlaces sufren variaciones en la calidad de la transmisión con el tiempo; esto hace que, para cubrir esta variabilidad del canal inalámbrico, se han de desarrollar y mejorar técnicas tanto de planificación de los recursos como de modelos que permitan la asignación eficiente de los recursos. Por medio del modelado de los esquemas de asignación de los recursos podremos analizar los parámetros que ofrezcan la calidad de servicio requerida por los usuarios de la red.

2.2. La asignación de los recursos radio en las redes móviles celulares

En un sistema de comunicaciones móvil, los terminales móviles se desplazan en un área de servicio y este desplazamiento sucede de vez en cuando, por tanto, requerirán servicios de comunicación extremo a extremo. La perspectiva desde la cual se puede abordar el problema de la asignación de los recursos a los terminales que solicitan una comunicación puede plantearse, bien desde la visión por parte del proveedor del servicio o bien por parte del usuario o terminal móvil. El operador o proveedor del servicio dispone y proporciona los recursos de comunicación a los usuarios que lo solicitan. Las dos partes interesadas tienen intereses contrapuestos, por un lado, el proveedor del servicio tiene como principal objetivo obtener el máximo beneficio a través del suministro del servicio a un número mayor de terminales - implica una mayor capacidad del sistema- y, por otro lado, el usuario desea obtener la mejor calidad de servicio por el coste que le supone el empleo de la red de comunicación. Es evidente que cuanto mayor sea el número de usuarios que soliciten el servicio estos requerirán más recursos y, por tanto, se limitará la capacidad del sistema. Dado que el operador es el que requiere un mayor coste de inversión en el sistema, es éste quién regula el proceso de diseño de la red desde su propia visión. La formulación del problema estará entonces en diseñar el sistema de comunicación donde se soporte el número de usuarios para unos requisitos de calidad del servicio determinados.

El diseño de la red móvil implica el estudio de la infraestructura de la propia red fija en la que se consideran: el número de estaciones base o puntos de acceso al servicio dónde van ubicados; qué capacidad ha de proporcionarse en función del número de puntos de acceso y, en general, lo que implica la planificación de la red móvil. Otro aspecto es el estudio de la asignación de los recursos radio a partir de un diseño de la infraestructura de red. Esto implica el análisis y el estudio de cómo se asignarán los recursos inalámbricos para satisfacer la demanda de los usuarios de la red móvil.

Un planteamiento formal del problema de la asignación de los recursos en la red móvil celular expuesta en [Zander et al. (2001)], expone elementos del modelado de la asignación de los recursos en las redes móviles celulares. En concreto, se expone que los terminales implicados en las comunicaciones que gestiona la red serán los denominados como terminales activos, es decir, aquellos que requieren una conexión en un determinado instante. Por tanto

los terminales activos se enumeran en el conjunto M :

$$M = \{1, 2, 3, \dots, m\}$$

Este conjunto de terminales activos cambian en el tiempo y, por tanto, variará el tamaño de M . La distribución estadística de M dependerá del comportamiento de los usuarios, es decir, del tráfico que generen y la duración de las sesiones. Además de los terminales se enumeran los puertos de acceso (estaciones base) por el conjunto $B = \{1, 2, 3, \dots, b\}$ y los tipos de señal que se establecen en la conexión, enumerada por el conjunto $C = \{1, 2, 3, \dots, c\}$. Para realizar una conexión o enlace en el sistema es necesario un puerto de acceso del conjunto B , un tipo de señal del conjunto C y una potencia de transmisión para realizar la comunicación entre el terminal y el puerto de acceso. Si en la asignación de recursos radio en el enlace se asume que todos los transmisores utilizan la misma potencia constante, la asignación de los recursos se puede simplificar en los siguientes conjuntos:

$$M^{(k)} = \{j\}$$

siendo j el terminal al que se le ha asignado un tipo de señal $k = 0, 1, 2, 3, \dots, c$.

$$B^{(k)} = \{i\}$$

siendo i el puerto de acceso radio al que se le ha asignado un tipo de señal $k = 0, 1, 2, 3, \dots, c$.

El valor de $K = 0$ representa aquel terminal al cual no se le asignado ningún canal. Los conjuntos $M^{(k)}$ normalmente son disjuntos, dado que cada terminal solo utiliza un tipo de señal determinada, mientras que los conjuntos $B^{(k)}$ no lo son puesto que una estación base o puerto de acceso sirve a múltiples terminales utilizando distintas señales. Con los niveles de potencia de las señales y las interferencias se puede modelar las ganancias de los enlaces entre transmisores y receptores: $G_{i,j}$ (variable aleatoria) representa la ganancia en potencia del puerto de acceso i y el terminal j , donde la potencia recibida por el receptor j es:

$$P_{rx,j} = P_{tx,j} \cdot G_{i,j}$$

y $P_{tx,j}$ es la potencia del transmisor j . En consecuencia las ganancias de los enlaces en forma matricial de dimensión $B \times N$ es:

$$G = \begin{pmatrix} G_{11} & G_{12} & \dots & G_{1M} \\ G_{21} & G_{22} & \dots & G_{2M} \\ \dots & \dots & \dots & \dots \\ G_{B1} & G_{B2} & \dots & G_{BM} \end{pmatrix}$$

Esta matriz representa las condiciones instantáneas de propagación del sistema puesto que sus valores pueden variar en el tiempo. La tarea de los esquemas de asignación de los recursos es encontrar las asignaciones de los enlaces para los cuales la calidad de servicio sea la más adecuada en las posibles comunicaciones.

Las tendencias de la evolución de la arquitectura definida por 3GPP [The Third Generation Partnership Project, 3GPP en <http://www.3gpp.org>] se encaminan hacia una topología abierta distribuida. Este hecho es de máxima importancia para la gestión de los recursos radio en las redes de comunicaciones inalámbricas [Mihovska et al. (2009)]. Las funciones del control de la red radio (RNC, Radio Network Controller) se divide en entidades funcionales que se clasifican en distintos bloques en función del ámbito y el plano en el que se ubiquen, sea el plano de control o sea el plano de usuario (Figura 2.1). Atendiendo a este esquema, distintas propuestas han sido planteadas por el 3GPP que descompone las funciones de control de la red radio y las traduce en dos nuevos tipos de entidades de red como son: los servidores de control radio y los servidores del plano de control. En definitiva, las arquitecturas abiertas adoptadas para los nuevos sistemas de comunicaciones se diseñan para el uso de mecanismos distribuidos en la gestión de la red de acceso radio.

La heterogeneidad de los escenarios actuales para el soporte de las interacciones entre los distintos sistemas, implica la mejora de los mecanismos de la gestión de los recursos radio entre otros aspectos de la red.

El incremento de la demanda por parte de los abonados de una red móvil inalámbrica fiable para el transporte de voz y datos ha motivado a los proveedores de servicio actualizar sus redes para facilitar una mayor capacidad de gestión de los recursos con mejores prestaciones. El objetivo de los proveedores de servicio es la integración de un amplio abanico de servicios en una infraestructura de red común. La heterogeneidad de los escenarios actuales produce que las soluciones para soportar interacciones de los sistemas sean complejos y múltiples, esto puede significar retardos en la ejecución y degradación de la calidad de servicio percibida por los usuarios, reducción del caudal o el incremento de la carga del tráfico en la red. Es por esto que

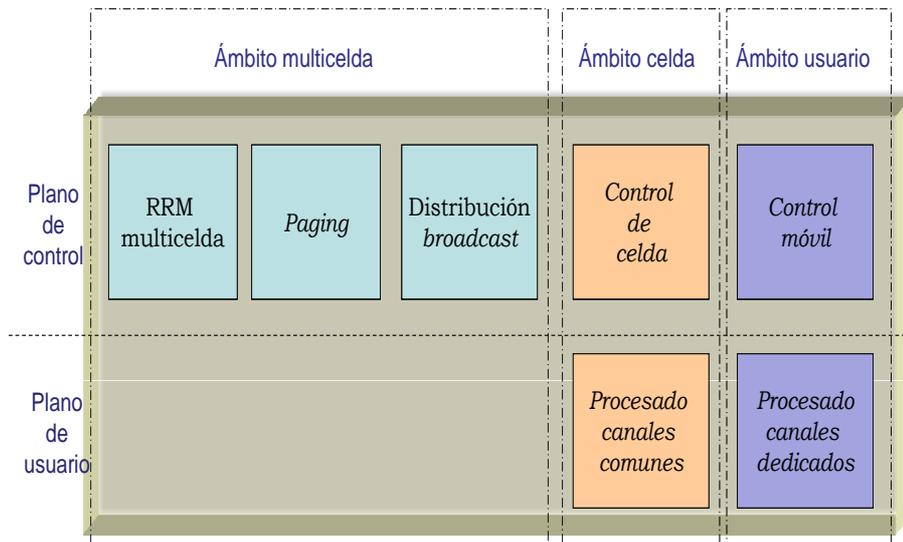


Figura 2.1: Funciones del Controlador de Red Radio

los mecanismos de la gestión de los recursos radio deben ser estudiados desde una perspectiva más amplia. La interacción entre distintas tecnologías de acceso es esencial para los operadores, de forma tal que puedan asegurar la continuidad del servicio, mejoras en las capacidades de las redes y mayores beneficios para los operadores.

La mayoría de los objetivos de la gestión de los recursos radio se centran en conseguir una alta capacidad y un mayor throughput para una calidad de servicio determinada. Entre las tareas de la gestión de los recursos se dividen cuatro áreas de estudio distintos como son: la asignación de los canales, el control de admisión, el control de potencia y el traspaso. El control de admisión es posible cuando una nueva llamada intenta ser aceptada en una celda dada solo cuando la capacidad de dicha celda está disponible para manejar dicha llamada, en caso contrario la llamada se bloqueará. El mismo caso se produce con las llamadas procedentes de un traspaso de una celda a otra. El control de potencia es un punto crítico para asegurar que la potencia del terminal móvil no caiga por debajo de un cierto umbral y asegurar principalmente el consumo de batería por parte del terminal móvil. Todos los sistemas móviles celulares incorporan el envío de mensajes de control en el enlace de subida

para gestionar su potencia.

En escenarios heterogéneos con distintas tecnologías radio, es necesario optimizar la gestión de los recursos radio, RRM, así se distinguen varios mecanismos: la *interworking* RRM, la RRM de cooperación, la RRM común, la *joint* RRM, la RRM concurrente y la RRM por capas. Por un lado, la gestión de recursos en cuanto a su funcionamiento entre distintas tecnologías (*interworking*) se basa en distintos niveles de acoplamiento. Por otro lado, la gestión de recursos radio combinada representa un enfoque de cooperación donde los recursos se manejan de forma centralizada.

Una gestión de recursos común es una solución propuesta por los grupos de GERAN y UTRAN del 3GPP para hacer cooperar a las redes GSM/GPRS/EDGE y UMTS. La RRM combinada es un enfoque de cooperación entre las distintas tecnologías donde los recursos se gestionan de forma centralizada para toda RAT (tecnología de acceso radio) por medio de una única unidad funcional. En este modelo se distinguen distintas unidades: la IMU, Interface Monitoring Unit, que controla los recursos utilizados y aquellos que están disponibles en todas las interfaces del sistema; la RMU, Resource Management Unit, que obtiene los datos enviados por la IMU de tal forma que hace un seguimiento de la carga de tráfico en los elementos para encontrar el escenario correcto ante la situación de congestión, para ello utilizará la herramienta apropiada para ejercer la toma de decisiones. Se trata pues de un sistema constituido por distintas unidades que controlan el tráfico además de prever y reorganizar los atajos para reaccionar ante situaciones de congestión en sectores con sobrecarga. La monitorización en tiempo real por medio de los indicadores de rendimiento claves (KPI, Key Performance Indicators) es uno de los mecanismos principales en la gestión de los recursos. Una vez se selecciona una técnica para afrontar un evento de posible congestión, bien se obtiene la información a través de una base de datos almacenada con casos similares a los que se presenten, o bien se realiza por un enfoque basado en un refinamiento que optimice la técnica de gestión de los recursos (RMT, Resource Management Technique) para conseguir los objetivos del operador de la red.

La gestión de los recursos radio común (CRRM, Common Radio Resource Management)[Skehill et al. (2007)] es un mecanismo de distribución del tráfico entre estos sistemas, ofreciendo la posibilidad de incrementar la capacidad de la red y la calidad de servicio percibida por el usuario con la consiguiente reducción de costes. Para un determinado operador de red el conjunto de recursos radio se reparte en repositorios de recursos. Un servidor

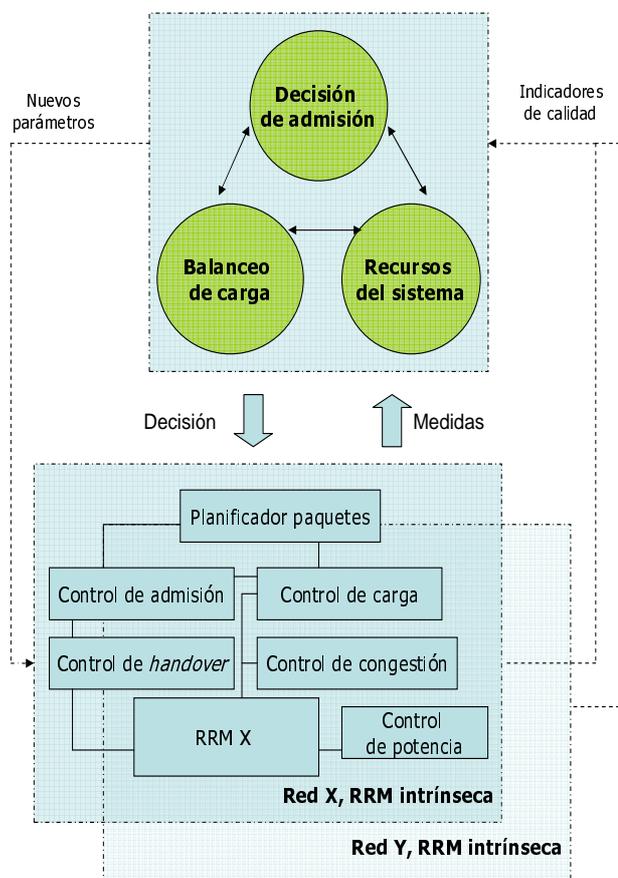


Figura 2.2: Interacción para la gestión de recursos radio común

CRRM viene a representar un nodo lógico dentro de la arquitectura GERAN y UTRAN que toma continuamente información medible desde cada uno de las plataformas (RRM intrínseca a cada tecnología) y toma decisiones para optimizar la gestión de la red heterogénea (Figura 2.2).

La *Joint* RRM, JRRM, permite soportar el trabajo entre distintas tecnologías de acceso, principalmente gestionadas por un controlador central, gestionando las subredes acopladas y, sobre todo, proporcionando una mayor capacidad al sistema. Este enfoque de la gestión da lugar a la planificación conjunta de los recursos y también al control de admisión conjunto con objeto de conseguir la mayor eficiencia posible y satisfacer los requisitos de calidad

de servicio establecidos.

La arquitectura de la JRRM es bastante similar a la de la RRM común, no solo se restringe a las redes móviles celulares sino que además permite el acceso radio sobre otras redes inalámbricas como las WLANs. De ahí que permita el acceso de un terminal a varias RATs manteniendo de forma simultánea varios enlaces sobre las RATs.

Por otro lado, la RRM concurrente es otro mecanismo para la gestión eficiente de recursos radio entre distintas redes de acceso radio, RANs. Al contrario que la JRRM es una entidad local la que asigna los recursos dentro de cada red de acceso y estas entidades locales han de cooperar de forma distribuida para conseguir un uso eficiente de los recursos.

La gestión de los recursos por capas (Layered RRM) se basa en la incorporación de una capa correspondiente al nivel de enlace que es común a todas las capas que subsisten en cada una de las tecnologías de acceso (RAT), de tal manera que la capa común tendrá funciones de puente (*bridge*) entre las distintas tecnologías de acceso.

2.2.1. Fundamentos de la asignación de los recursos radio

Desde un punto de vista de la asignación de los canales, para cada una de las celdas existe, por un lado, una división entre la asignación estática de las frecuencias desde la estación base y, por otro lado, la asignación variable o dinámica de las distintas portadoras que puede gestionar la estación base para establecer las comunicaciones con las estaciones móviles activas en esa celda.

Otra visión - es la que nos concierne¹- trata de asignar los recursos disponibles para cada una de las portadoras de frecuencia entre la estación base y los terminales móviles en el transporte de la información.

En primer lugar hemos de destacar que en una red móvil en la que se emplea tanto la conmutación de circuitos - para cursar el tráfico de voz - como la conmutación de paquetes - para cursar el tráfico de datos, se utilizan distintos algoritmos de asignación de los canales. Por tanto, se han de distinguir dos procesos distintos cuales son: 1, la interacción entre los canales haciendo uso de la conmutación de circuitos y los canales que hacen uso de la conmutación de paquetes, y 2, la interacción o la coexistencia de los usuarios

¹Para la red móvil objeto de esta tesis nos referimos a la red EDGE y los recursos vienen representados por los *timeslots*.

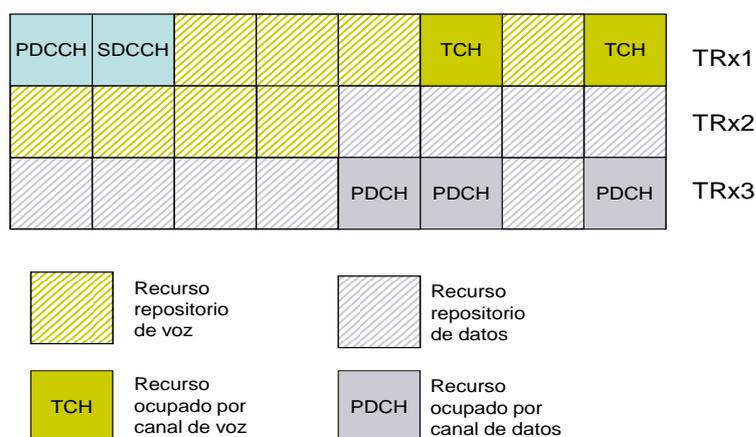


Figura 2.3: Asignación de canales mediante repositorios

haciendo uso de los mismos recursos para el transporte de datos en la red.

Es posible que se compartan todos los recursos disponibles entre el tráfico de datos (conmutación de paquetes) y el tráfico de voz (conmutación de circuitos), y también es factible dividir estos recursos en repositorios separados. Tanto en un caso como en otro, el algoritmo de asignación de los canales selecciona los recursos que se utilizarán en cada conexión además de permitir que un determinado canal sea el que pueda emplear uno o varios recursos (*time – slots*) con ciertos parámetros de frecuencia.

2.2.2. Modelos de asignación de canales

El primer modelo de asignación de los canales que se presenta es el que hace uso de repositorios separados de recursos. En este caso se mostrarán dos repositorios o *pools* (Figura 2.3).

Este método de asignación se basa en dividir los recursos en dos repositorios donde irán alojados los dos tipos de canales, sin contar con los canales de señalización (PDCCH² y SDCCH³), tal como se muestran en la Figura 2.3.

²PCCCH - Packet Common Control Channel, canal formado por distintos canales lógicos de control: el PRACH (Packet Random Access Channel), PPCH (Packet Paging Channel), PAGCH (Packet Access Grant Channel) and PNCH (Packet Notification Channel).

³Stand-alone Dedicated Control Channel (SDCCH) canal que se utiliza para proporcionar una conexión fiable para la señalización y los mensajes cortos.

Esto significa que se dispone de un modelo de reserva de recursos dedicados en función del tipo de canal al que vaya destinado. Las conexiones de voz se asignan a los canales que pertenecen al repositorio de conmutación de circuitos y las conexiones de datos se asignarán a los canales dentro del repositorio de conmutación de paquetes. Los beneficios de tener dos repositorios separados se puede resumir en que la interacción entre los tipos de canales no existe pero, en cambio, en función de la demanda del tráfico de un tipo, voz, o de otro, datos, la ocupación de los recursos podría verse desaprovechada. Para ello es necesario ajustar los servicios de voz, conmutación de circuitos, que precisan de una asignación fija de los *time – slots*, con tráfico formado por un flujo continuo en el tiempo, frente a los servicios de datos, conmutación de paquetes, que requieren una asignación dinámica de los recursos y de corta duración, siendo en general un tráfico denominado “a ráfagas”. El algoritmo de asignación de los canales ha de optimizarse para las dos clases de tráfico dentro de cada uno de los repositorios de forma tal que produzca la mínima interacción entre los dos repositorios.

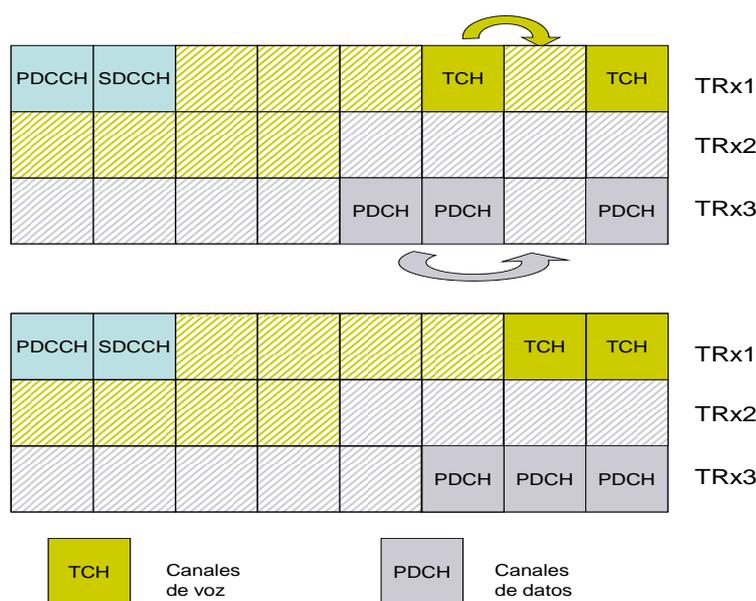


Figura 2.4: Método de *repacking*

Uno de los principales inconvenientes de este modelo es la necesidad de realizar el método de *repacking* continuamente para poder optimizar la uti-

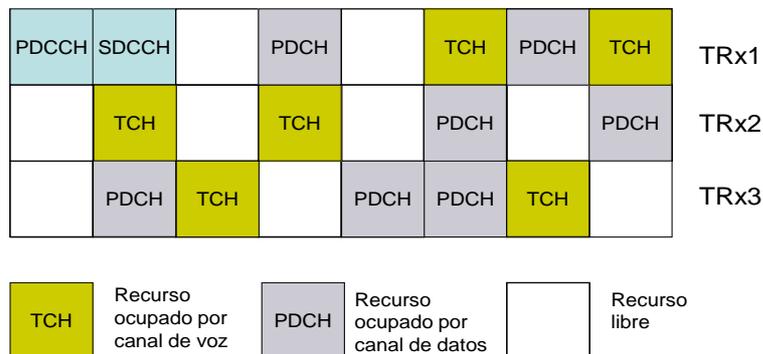


Figura 2.5: Asignación de canales mediante un único repositorio

lización de los canales a un extremo del repositorio. Esto quiere decir que los recursos estarán disponibles de forma consecutiva para los usuarios. Este método de repacking se puede observar en la Figura 2.4. La implementación del esquema *repacking* hace uso del mecanismo *intracell-handover* (traspaso dentro de la misma célula) para la asignación a un time-slot contiguo, de tal forma que los recursos que se utilizan por los usuarios ocupen posiciones contiguas, una vez los time-slots van quedando inactivos, bien por el traspaso a otra celda de la llamada o bien por la finalización de la llamada o de la transmisión de los datos. Este método, en general, puede incluir la asignación contigua entre distintas portadoras de frecuencia dentro de la misma celda. El gran inconveniente de este método es que si se producen un gran número de traspasos dentro de la misma celda durante una llamada se pueden producir efectos negativos en cuanto a la calidad del servicio. En particular para períodos de expulsión de una llamada por encima de los 450ms de duración puede ocurrir que, cada vez que una llamada se reubica a otra posición (nuevo time-slot recién liberado y contiguo a otro ocupado) se producen deficiencias en el servicio ofrecido al usuario.

Otro modelo para compartir los recursos consiste en formar un único repositorio de canales en los que tanto el tráfico de voz como el de datos hacen uso de los recursos disponibles en la celda. Las conexiones han de competir por la utilización de los recursos de la celda (Figura 2.5).

Este método para la asignación de los recursos posee un mayor número de lugares posibles en los que ubicar los canales y proporciona una mayor libertad para la selección del canal. No obstante en este método es mucho

más compleja la optimización de la asignación de los canales tanto para la conmutación de circuitos como para la conmutación de paquetes. En general, los recursos disponibles obedecen a un repositorio fragmentado en el que los canales no ocupan posiciones consecutivas y, por tanto, con los terminales móviles que disponen de la capacidad *multislot*, en particular, su potencialidad para la transmisión de datos se ve afectada negativamente (menor *throughput* o mayor retraso en la transmisión).

Tal como se ha comentado con anterioridad, el otro proceso que se ha de tener en cuenta en la gestión eficiente de los recursos es la interacción o la coexistencia de los usuarios haciendo uso de los mismos recursos para el transporte de datos. Para ello lo más adecuado es seleccionar recursos consecutivos que sean menos utilizados. Si existe un número X de terminales que hacen uso de ese slot de tiempo, un nuevo terminal que quiera hacer uso de ese recurso obtendrá la $1/X + 1$ parte de la capacidad de ese time-slot. Si las conexiones de datos tienen distintas prioridades es posible la asignación a los terminales móviles en función de esas preferencias previamente establecidas. Además es posible la reserva de ciertos recursos para poder garantizar cierto caudal en la transmisión de datos. Este último esquema es el de reserva de recursos con el objetivo de maximizar el *throughput*.

2.3. Trabajos anteriores acerca de los métodos de asignación de canales

En esta sección presentamos trabajos realizados anteriormente para la asignación de recursos en las redes de paquetes de datos, en particular, para redes móviles celulares TDMA. El primero de ellos en exponer el concepto de *multi – slot packet data channel* para usar la capacidad máxima posible en la transferencia de datos y el concepto de *master – slave* para acomodar en los time-slots canales de control/señalización (*paging*, acceso aleatorio y de reserva) es el trabajo de [Turina et al. (1996)], así como canales únicamente dedicados para la transferencia de datos. Más tarde, en [Chen et al. (1996)] se presentan dos algoritmos de asignación de canales, estos son: (a) el de división de los repositorios para los canales de voz y los canales de datos; (b) el algoritmo de uso compartido por ambos tipos de tráfico, de forma que se reduzca el bloqueo y la probabilidad de finalización forzada de las llamadas, basándose en que durante los períodos de congestión las llamadas de voz

hagan uso de parte de los time-slots dedicados al tráfico de datos.

A partir de estos trabajos se presentan algoritmos de asignación de los recursos radio, fundamentalmente para GPRS y EDGE. En el trabajo de [Kennedy et al. (1999)] se evalúa un algoritmo de asignación de los recursos radio híbrido en el que las transferencias de datos utilizan los períodos inactivos de llamadas de voz GSM además de utilizar un número de canales dedicados para datos. De tal manera que si se aumenta el número de canales dedicados para datos, la probabilidad de bloqueo aumenta y el número medio de canales disponibles para datos no se ve incrementado significativamente, además de implicar un efecto similar en el *throughput* del sistema. Por el contrario, con el aumento de canales dedicados a datos, la desviación estándar del número de canales disponible para datos disminuye, lo cual implica una mejora en la calidad de servicio, en concreto para el retardo de paquetes.

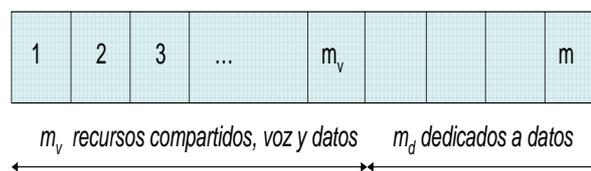


Figura 2.6: Asignación recursos de voz GSM y datos GPRS (Ni and Häggman)

[Ni et al. (1999)] parten de que el diseño de GPRS para la transmisión de datos toma los recursos radio del *pool* de canales que no se utilicen para el tráfico de voz, con lo cual el número de canales asignados para el tráfico de datos es una variable aleatoria dependiente del tráfico de voz. En este trabajo se evalúa, por medio de un método de aproximación, el rendimiento de GPRS usando un único time-slot para datos puesto que deducen que con el servicio multi-slot la probabilidad de bloqueo aumenta y el retardo se hace mayor. Los canales de datos PDCHs se liberan tan pronto como se pueda en el caso de que la demanda de recursos por parte de los servicios de conmutación de circuitos aumenten. De las dos posibilidades de liberación de dichos PDCHs - 1. liberación inmediata del canal, implica la interrupción de la transmisión de datos y 2. retardar la liberación del canal, con lo cual se puede esperar a que se transmitan los datos para que después se asigne el canal a la llamada de voz - en este caso se asume el primer caso de liberación inmediata. Así se utilizan canales dedicados para datos y el resto compartidos para voz y para

datos (Figura 2.6). Dicho trabajo asume que en el caso en que la red no pueda proveer de recursos a una petición con su tasa de transmisión, la red ha de negociar y acordar con el usuario la nueva tasa de transmisión que la red pueda proporcionar.

En cuanto al estudio de la reducción de las tasas de bits haciendo uso de un time-slot para cada usuario que se incorpore al sistema y en caso de necesitarse más recursos para la transmisión de datos, el número que se propone por parte de [Al Agha et al. (2000)] es de $x/2$, siendo el número máximo que se pueda asignar x de 8 usuarios por time-slot en una sesión multicast en una celda. Este número se puede reducir a la mitad con objeto de reenviar paquetes perdidos de un usuario. En este estudio se garantiza que la estación base, BS, asigne un time-slot por usuario si el número de usuarios aumenta, y es la BS la que multiplexa a los usuarios sobre un time-slot.

Otros esquemas de asignación de recursos se presentan en artículos como los de [Zhang et al. (2000)], [Itkonen et al. (2002)] y [Jaupitre et al. (2002)]. El primero de ellos, modela la gestión de los time-slots haciendo una reserva de recursos sólo para el tráfico de datos, otros sólo para el tráfico de voz y los restantes para uso compartido tanto de voz como de datos. Este modelado permite obtener un throughput efectivo estimado por time-slot, el tráfico total estimado para voz y datos, el número mínimo estimado de time-slots de voz dedicados para soportar cierto grado de servicio, GoS, el número mínimo estimado de time-slots para datos y, por último, determina si el caudal de datos percibido por el usuario satisface la QoS designada previamente. [Itkonen et al. (2002)] proponen criterios de planificación para obtener una capacidad combinada de la interfaz aire GSM+GPRS/EDGE haciendo uso de una asignación de los time-slots. En concreto realizan medidas de entornos en dos celdas, una de ellas en una celda exterior en calles de Helsinki y otra celda en un entorno interior dentro de un edificio público. Para una portadora se configuran 7 time-slots por trama donde se reservan 3 time-slots para tráfico de voz y el resto se asignan indistintamente para tráfico de voz o de datos en función de la demanda. Este estudio contrasta las medidas obtenidas con los valores teóricos que no se distancian en cuanto a disponibilidad del número de time-slots para canales de datos así como la probabilidad de tener cierto número de canales de voz ocupados, tanto en entornos exteriores como interiores. Por último, en [Jaupitre et al. (2002)] se analiza la eficiencia de compartir los recursos entre el tráfico de voz y de datos por medio de una configuración basada en utilizar: portadoras en la que los recursos están reservados solo para tráfico de voz; otras portadoras

en las que se reservan la mitad de los recursos para tráfico de voz y la otra mitad para tráfico de datos; y portadoras en las que se comparten los recursos tanto para voz como para datos. Esta estrategia ha hecho posible analizar por medio de un número variable de flujos de datos dos clases de tráfico: *best effort* y *premium*, en particular el *throughput* para el tráfico unidireccional y bidireccional en modo estático y dinámico en el enlace ascendente.

2.4. Criterios de calidad de servicio y reglas de dimensionado

El éxito de la evolución de las redes celulares de radio paquetes está íntimamente relacionado con la capacidad de ofrecer un conjunto de servicios con un coste económicamente razonable y de manera fiable. Los requisitos fundamentales para que esto se produzca son los mecanismos de la gestión del tráfico, capaces de proporcionar el soporte necesario de calidad del servicio y los procedimientos de ingeniería de tráfico. La gestión del tráfico y la ingeniería de tráfico son funciones complementarias y mutuamente dependientes. La ingeniería y la gestión de tráfico incluyen funciones necesarias para planificar, diseñar, proyectar, dimensionar, desarrollar y supervisar las redes en condiciones óptimas de acuerdo a la demanda de servicios, los márgenes de beneficios de la explotación, la calidad de la prestación y el entorno regulatorio y comercial.

2.4.1. Conceptos de ingeniería de tráfico

Los procedimientos de ingeniería de tráfico aseguran que la red se diseñe y se implemente de manera eficiente en términos tecnológicos y económicos. Estos procedimientos se basan en la relación tráfico-rendimiento, la capacidad de los enlaces de la red, la demanda de tráfico y el rendimiento obtenido, todo ello con el adecuado dimensionado de los elementos de red para soportar el tráfico ofrecido a ésta.

En particular para las redes de conmutación de paquetes el objetivo se centra en definir procedimientos sencillos que se puedan aplicar igualmente en las redes que utilizan la conmutación de circuitos. Para las redes móviles de paquetes de datos la cantidad de recursos radio disponibles es un factor crítico en cuanto al rendimiento que se pueda conseguir. Las reglas de la

ingeniería de tráfico deben ser sencillas y normalmente basadas en el número de usuarios y el volumen de tráfico ofrecido durante la hora cargada, todo ello con objeto de asegurar una aplicación directa en la práctica. Sin embargo, la relación entre el rendimiento y el tráfico ofrecido se puede estimar en función del número de recursos ocupados.

La gestión del tráfico, por otro lado, basada en los requisitos de calidad de servicio demandados por los usuarios, supone consecuentemente la definición del grado de satisfacción que estos obtienen del servicio solicitado. Así la disponibilidad de la red y sus elementos, el caudal (transferencia de datos efectiva), la pérdida de paquetes de datos, el retardo y la variación del retardo, son los parámetros básicos que se modelan para permitir que, en función del tipo de tráfico o concretamente de la aplicación del usuario (voz, correo electrónico, sesiones interactivas, navegación web, transferencia de datos, video-conferencia, comercio electrónico, multicast, ...), varíe el grado de importancia de cada uno de los parámetros de calidad de servicio. A través de los modelos de tráfico para cada una de las aplicaciones se puede diseñar, implementar y mejorar los sistemas de comunicaciones.

Para planificar los recursos y dimensionar la red con distintos tipos de tráfico, se necesita modelar cada tipo de tráfico. En esta fase de modelado, se necesita describir los parámetros que son de interés para el análisis. Es en este sentido que interesa mucho más que hacer un modelo exacto del tráfico, realizar un modelo de los parámetros de tráfico que tienen influencia en el rendimiento de la red.

2.4.2. Modelos de colas y caracterización del sistema móvil celular

En el estudio y la caracterización de los sistemas de comunicaciones se definen distribuciones del tráfico por medio de los modelos de colas. Un sistema de colas se determina por medio del proceso de llegadas al sistema, el proceso de servicio y la disciplina de cola utilizada. En concreto en un sistema móvil celular de radio paquetes el proceso de llegadas se determinará bien para los paquetes (IP en este caso) si se analiza la capa de red, o bien para los radio bloques en la capa de enlace (RLC/MAC).

En función del tipo de modelo [Hoymann et al. (2002)] sea una M/M/n, procesos ON/OFF u otros [Vornefeld (2002)], podemos obtener distintos resultados y compararlos para un sistema de comunicaciones. Así, por excelen-

cia, el modelo utilizado para las redes que utilizan conmutación de circuitos se basa en el modelo de pérdidas de Erlang para su dimensionado. Esto significa que el proceso de llegadas de las llamadas sigue un proceso de Poisson con una tasa media de llegadas λ , donde la probabilidad de que lleguen k llegadas en un intervalo de tiempo Δt es:

$$P(X = k) = \frac{(\lambda \Delta t)^k}{k!} \cdot e^{-\lambda \Delta t}$$

para $k \geq 0$.

El tiempo entre llegadas consecutivas se modela por medio de una distribución exponencial:

$$P(T \leq t) = 1 - e^{-\lambda t}$$

Para modelar la duración de la llamada se utiliza una distribución exponencial cuya función es:

$$f(t) = \mu \cdot e^{-\mu t}$$

donde $1/\mu$ es la duración media de la llamada.

En el caso de las redes móviles se necesita relacionar los parámetros de tráfico a distintos tamaños de celdas y la movilidad de los usuarios. Con el objetivo de realizar el análisis del tráfico, se hace necesario el análisis de parámetros como: el tiempo medio de permanencia en la celda, la intensidad de trasposos (*handovers*), la probabilidad de bloqueo de *handovers*, el número medio de trasposos y la probabilidad de pérdida de las llamadas. Tanto en [Hong et al. (1986)] como en [Hlavacs et al. (2002)] podemos observar modelos que atienden, en el primero, a un modelo de telefonía móvil celular usando conmutación de circuitos y, en el segundo, al modelado de la gestión de recursos para distintos tipos de tráfico. En el caso de entornos multiclase es necesario restringir el número de llamadas simultáneas para cada clase de tráfico, por tanto, se definen las limitaciones para las llamadas de cada clase k siguiendo las siguientes relaciones:

$$\sum_{k=1}^K c_k > C$$

para $0 \leq i_k \leq c_k \leq C$, $k = 1, 2, \dots, K$. Donde i_k es el número de llamadas de clase de tráfico k , c_k es el límite del número de canales que pueden asignarse a esa clase al mismo tiempo, C es el número total de canales en la celda y K es el número de clases de tráfico. Si esta expresión no se cumple entonces

se puede separar en grupos de K cadenas de Markov independientes unidimensionales. En caso contrario el proceso de llegadas será la superposición de procesos de Poisson de distintas clases de tráfico. De esta forma se puede tratar con una tasa total de llegadas a la celda:

$$\lambda = \sum_{k=1}^K (\lambda_{n,k} + \lambda_{h,k}),$$

siendo $\lambda_{n,k}$ y $\lambda_{h,k}$ las tasas de llegadas de llamadas de clase k en la propia celda y la de llegadas de llamadas de clase k provenientes de traspaos (*handovers*) procedentes de celdas vecinas respectivamente. El tiempo de permanencia de la llamada se distribuye por la función:

$$f(t) = \sum_{j=1}^K \frac{\lambda_j}{\lambda} \mu_j \cdot e^{-\mu_j t}$$

donde $\lambda_k = \lambda_{n,k} + \lambda_{h,k}$ y $\mu_k = \mu_{c,k} + \mu_{h,k}$ siendo las tasas de llegadas de llamadas y las tasas de finalización para el tráfico k , respectivamente. El tiempo de permanencia total dependerá directamente del tráfico total ofrecido A que viene a ser: $\sum_{j=1}^K A_j$. En el caso que $K = 2$ se tratará de una cadena de Markov binomial (Figura 2.7) de dos dimensiones para dos clases de tráfico.

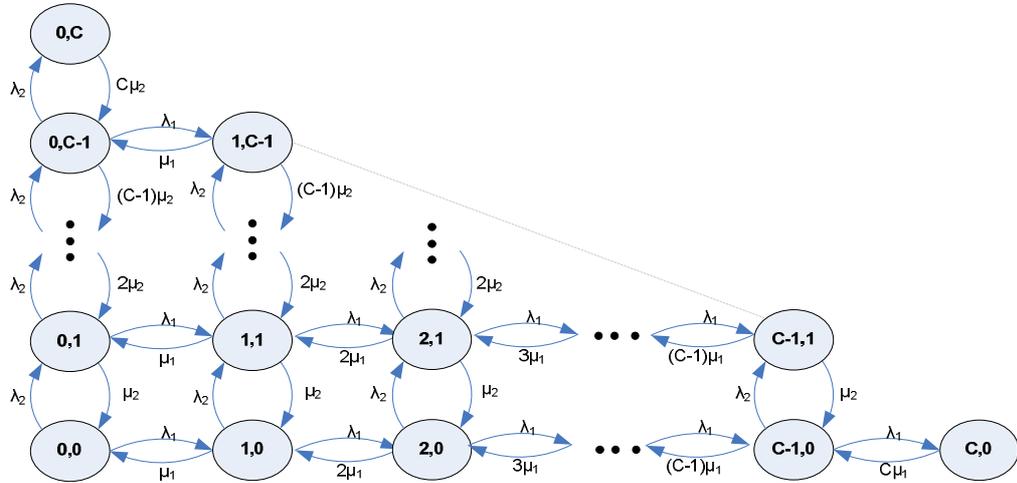


Figura 2.7: Cadena de Markov binomial para dos clases de tráfico

Por un lado, en el tráfico propio de paquetes de datos, los procesos se caracterizan por llegadas de paquetes IP, en concreto, por medio del tiempo

entre llegadas de los datagramas IP y el tamaño de cada datagrama. Por otro lado, el proceso de servicio de los radio bloques (RLC/MAC) se caracteriza por medio del estudio del proceso de servicio del radio bloque y su tiempo de servicio. Los modelos analíticos tratan normalmente de encontrar una respuesta aceptable al problema de la exactitud del modelo frente a la “tratabilidad” analítica de dicho modelo. En este compromiso se manejan distintas consideraciones y es el grado de aplicación de los distintos modelos los que tratan de estudiar y mejorar la eficiencia de los sistemas de comunicaciones. Modelos que se utilizan en la evaluación del tráfico de paquetes se basan en aplicar modelos como el de flujo de fluidos (FFM, *Fluid Flow Model*) desde trabajos como [Tan et al. (1995)] hasta en [Timus et al. (2011)], o modelos basados en procesos de llegadas MMAP (*Marked Markovian Arrival Processes*) [Irnich et al. (2002)].

En el primer modelo FFM trata de predecir la función de distribución del tamaño del buffer bajo una carga de tráfico determinada. Después de calcular la cantidad media de tráfico en cola, se calcula el tiempo medio de espera en cola, en el caso de tráfico IP será la espera de los datagramas IP. La tasa de llegadas de los datos por un número de fuentes se compara con el agua que cae dentro de un depósito (representa el buffer) que se vacía a una tasa constante. La unidad de datos se supone que sea infinitamente pequeña. Comparando las fuentes de tráfico con los grifos de agua que de forma asíncrona conmutan como fuentes ON/OFF. Los tiempos de permanencia en los estados ON y OFF se distribuyen exponencialmente (Figura 2.8). Siendo que en el estado ON una fuente trasmite datos a una tasa constante. A estas fuentes ON/OFF también se las denomina IRP (*Interrupted Rate Process*) que en el caso de N fuentes iguales superpuestas conforman un proceso NIRP que pueden describirse por medio de una cadena de Markov unidimensional con un número total de estados $N + 1$ (Figura 2.9).

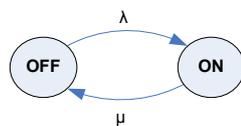


Figura 2.8: Cadena de Markov para un proceso IRP

En [Vornfeld (2003)] se propone un modelado analítico para la evaluación del rendimiento del sistema EGPRS, haciendo uso de una caracterización de las fuentes de tráfico por medio de los procesos MMAP como entrada al

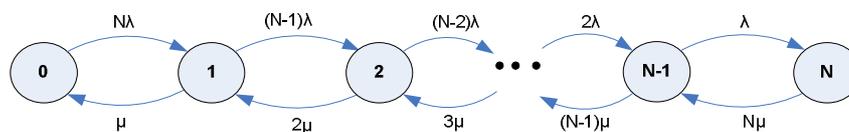


Figura 2.9: Cadena de Markov unidimensional para un proceso NIRP

modelo de colas. Este análisis del modelo conduce a la deducción de que un factor clave es la asignación individual de las distribuciones del tiempo de servicio que vienen determinadas por las condiciones de propagación radio.

Para caracterizar el tráfico del sistema se requieren aplicaciones para tomar distintas medias de rendimiento. Los parámetros medidos a partir de estas aplicaciones [Irnich et al. (2002), Hoymann et al. (2002)] permiten evaluar el sistema. Las medidas como: el caudal de paquetes IP por usuario medidos durante períodos de transmisión, los tiempos de respuesta de la aplicación (diferencia entre el tiempo cuando un usuario solicita una página web o un correo electrónico y el tiempo cuando se recibe completamente la información solicitada), el retardo del datagrama IP (retardo extremo-extremo), la tasa de bloqueo de una sesión o la utilización de los canales de paquetes de datos (PDCH) que implica conocer el número de radio bloques utilizados tanto para datos como para control, son factores relevantes de rendimiento del sistema [Stuckmann et al. (2002)].

Las reglas de ingeniería de tráfico con respecto al dimensionado de los canales, concretamente de datos, necesarios para cada una de las celdas son estudiadas desde dos puntos de vista. El primero de ellos trata de estimar el número de canales en función de la demanda de los usuarios, es decir, una asignación dinámica del número de canales necesarios. La segunda perspectiva para resolver esta cuestión es por medio de la reserva de un número fijo de canales de datos (PDCHs) para el potencial transporte de la información que soliciten los usuarios. Tanto una visión como la otra han sido estudiadas en [Stuckmann et al. (2001)] y proponen cuatro reglas sencillas que logran satisfacer los requisitos necesarios para el soporte del tráfico de datos. Estas reglas son: 1, definir la calidad de servicio deseada, por medio del throughput del enlace descendente por ejemplo; 2, estimar el número de usuarios por celda; 3, definir el tráfico ofrecido (IP) por usuario y calcular el tráfico ofrecido por celda; y 4, establecer el número de canales en función de los parámetros obtenidos anteriormente y con una probabilidad de bloqueo determinada. De esta forma se puede observar que desde el punto de vista de los operadores

es clave que las reglas concluyan en recetas relativamente sencillas a las que recurrir para el dimensionado de sus redes.

2.5. Conclusiones

En este capítulo hemos presentado dos secciones en las que, por un lado, se muestran los conceptos, funciones y mecanismos de la asignación de los recursos de las redes móviles celulares y, por otro lado, el modelado y criterios de QoS en los mecanismos de asignación de los recursos para estas redes. De forma concreta, se han descrito los tipos de gestión de recursos radio en las redes actuales con escenarios heterogéneos y las distintas tecnologías asociadas.

En cuanto a los modelos de asignación de los recursos se puede concluir que, en función del tipo de tráfico, voz o datos, los esquemas de asignación de los canales pueden basarse en la compartición de los recursos por medio de repositorios comunes de recursos o, de forma alternativa, distinguir el tipo de repositorio para cada clase de tráfico. Así, se han presentado distintos esquemas para EGPRS/EDGE de la literatura científica.

Por último se introducen aspectos sobre el modelado de los esquemas de asignación de los recursos radio por medio de los modelos de colas. Estos modelos permiten evaluar y estimar valores de los parámetros de QoS de los sistemas, así como reglas de dimensionado que permiten la gestión eficiente de la red.

Capítulo 3

Asignación compartida de recursos en la red móvil EDGE

3.1. Introducción

Este capítulo presenta una metodología que se sirve de la capacidad *multislot* de los terminales móviles para optimizar la asignación de los canales por parte de la red. En concreto, esta capacidad multislot viene condicionada a que la transmisión del tráfico de datos para un mismo usuario en un misma llamada ha de tener el canal correspondiente canal ubicado en un conjunto de recursos (time-slots) contiguos en el tiempo. A modo de ejemplo, si se dispusiese de tres canales en una celda con las tres portadoras correspondientes, en una situación tal como se muestra en la Figura 3.1 y se solicitase por parte de una aplicación (usuario o terminal) una capacidad que implicase el uso de varios recursos (time-slots) para la transmisión de sus datos, podría ocurrir que no sería posible asignar en ninguna de las tres frecuencias nuevos canales con capacidad multislot (2 ó más time-slots) puesto que no hay posibilidad de ubicar dicha llamada en time-slots consecutivos libres para alguna de las frecuencias del repositorio. Este ejemplo también es válido si únicamente esta situación se centrara en una determinada portadora. Por ejemplo la portadora TRx2, tal como se muestra en la Figura 3.1), dispone de cuatro time-slots libres pero no consecutivos. La capacidad de un usuario para transmitir los datos al cuádruple de la velocidad de transmisión con un solo time-slot, no es posible por la fragmentación de los time-slots en ocupados y no ocupados de forma disjunta en el tiempo. Este hecho implica que en el momento descrito,

no se podrían asignar, por ejemplo, dos time-slots a dos terminales para cada uno de ellos. Esto significaría que el usuario no dispondría del doble de velocidad de transmisión para transmitir sus datos, así como la combinación de un usuario haciendo uso de tres time-slots, velocidad triple con respecto a otro usuario en el que se le asigne el cuarto time-slot libre.

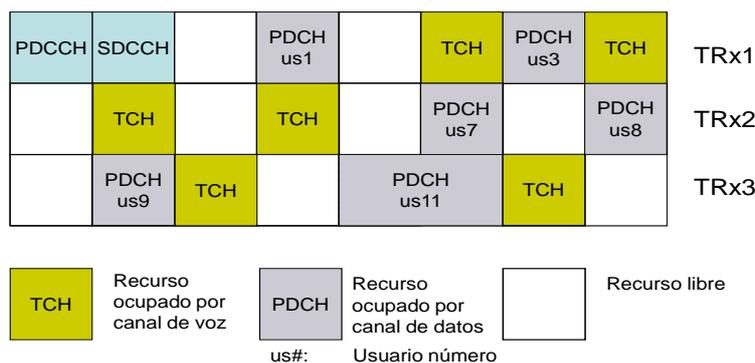


Figura 3.1: Situación momentánea de imposibilidad de capacidad multislot para tráfico de datos

Aquí se presenta y analiza una estrategia inicial para asignar de una manera eficiente, no solo evitar el problema comentado anteriormente sino incrementar considerablemente la probabilidad de ofrecer unas tasas de pico mayores con respecto a usar una asignación aleatoria de los recursos. Al método de asignación lo hemos denominado compartido por ráfagas. En este método se propone una estrategia de asignación de los recursos y un análisis de los estados y las probabilidades de permanecer en cada uno de los estados por los que pasa el sistema. De esta forma con los valores obtenidos se puede estimar los recursos disponibles que, por parte de la red, se requieren para distribuirlos de forma más eficiente entre los usuarios a los que proporciona el servicio.

3.2. Modelo de asignación compartida por ráfagas

El método de asignación de los canales que en esta sección se presenta, corresponde a un esquema en el que se requiere que el número de recursos disponibles y dedicados a la transmisión de datos sea el máximo posible. Es

por esto que cuando se asignen los canales para voz o para el tráfico de datos, estos recursos se localicen por medio de una política que permita optimizar la capacidad de los usuarios de datos.

En primer lugar destacamos que la asignación de los canales será concretamente para un número de recursos igual a ocho time-slots, con lo cual hemos de resolver tres cuestiones cuales son:

1. Cómo identificar los recursos.
2. Cómo se asigna el recurso.
3. Qué ocurre cuando se libere un recurso.

Contestando a la primera cuestión y dado que hay ocho time-slots en una trama [3GPP GERAN (2011)], cada uno de los recursos se indica por medio de los números de identificación "ID numbers", desde el número ID 0 hasta el número ID 7. Cada time-slot se denominará como ocupado cuando una llamada de voz o de datos lo esté usando, en cambio habrá "huecos" en los time-slots no utilizados que podrán ser ocupados posteriormente por nuevas entradas en el sistema. Podemos considerar que para indicar la ocupación o disponibilidad de los time-slots, utilizaremos la notación en binario y con longitud de los recursos n , en base 2, tenemos pues que una posición toma el valor 1 si el time-slot con ese número de identificador está ocupado y 0 en caso contrario. Además, definiremos el Estado r para $r = 0, \dots, 2^n - 1$, por el mismo número en base 10. Por ejemplo, si $n = 8$, el estado 5 es [00000101]. Por tanto, nos referimos al Estado r indistintamente por r o por $[r_{n-1}, \dots, r_1, r_0]$, siendo $r_i = \{0, 1\}$, esto es:

$$\text{Estado } r := \sum_{\text{numID}=0}^7 2^{\text{numID}} \cdot b(\text{numID}),$$

donde:

- $b(\text{numID}) = 1$
cuando el time-slot con ese número de identificador esté ocupado y 0 en caso contrario.

En la Figura 3.2 podemos observar el estado 5, ocupados los time-slots con numID 0 y 2), y el estado 7, ocupados los time-slots con numID 0, 1 y 2.

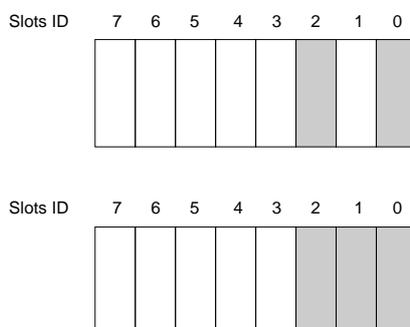


Figura 3.2: Ejemplo de asignación de time-slots

En cuanto a la segunda cuestión, cómo se asignan uno o varios recursos, emplearemos el método conocido como *First Fit* [Ivanovich et al. (2002)]. Este método consiste en la asignación de una nueva llamada al time-slot libre con el número de identificador más pequeño. En la Figura 3.2, inicialmente disponemos de dos time-slots ocupados, ID 0 e ID 2, de tal forma que la nueva asignación ocupará el time-slot con el menor ID que corresponderá en este caso al time-slot número ID 1, tal como aparece en la Figura 3.2, corresponde al estado 7 en el que los tres time-slots con menor número de identificador están ocupados.

No obstante, independientemente del lugar que ocupe la llamada en los posibles time-slots, cuando uno de ellos se libere, se producirá un hueco o slot libre sin ningún tipo de reordenación de los time-slots ocupados.

Este método evita fundamentalmente la excesiva fragmentación de los time-slots consecutivos. A priori, se puede estimar que la probabilidad de encontrar time-slots libres consecutivos es alta y de esta manera un operador puede disponer potencialmente de lugares en los que asignar los canales de las estaciones móviles por medio de la capacidad *multislot* para el tráfico de datos procedente de un usuario.

3.2.1. Análisis del modelo

En el modelo propuesto de asignación de recursos, time-slots, atendiendo a el trabajos de [Ivanovich et al. (2002)], para el tráfico de voz usando conmutación de circuitos se parte de dos presunciones: la primera de ellas es que los tiempos de llegadas son independientes entre sí y se describen por medio de una distribución exponencial negativa con media $1/\lambda$, es decir, implica

un proceso de llegadas de Poisson. Siendo λ el número medio de llamadas que entran al sistema por unidad de tiempo. La segunda presunción es que el tiempo de duración de la llamada es aleatoria y se describe por medio de una distribución exponencial cuya media es $1/\mu$, siendo μ la tasa media de servicio de llamadas por unidad de tiempo [Itkonen et al. (2002)].

El mecanismo de asignación de los recursos *First Fit* puede modelarse por medio de un diagrama de estados. Cada situación en el modelo puede representarse usando una cadena de Markov. El modelo visto como una cadena estocástica discreta puede construirse atendiendo al desarrollo futuro de la cadena y solo depende del estado actual y no de cómo se ha llegado al estado actual. Por tanto, este modelo corresponde a un modelo *markoviano* que puede resolverse por medio de un sistema de ecuaciones correspondiente a la transición de los estados posibles por los que pasa el sistema.

En el caso de utilizar ocho time-slots disponemos de 256 estados y sus posibles transiciones. Para obtener las probabilidades de permanecer en cada uno de los estados se ha de resolver el sistema de ecuaciones de transición correspondiente. De forma analítica a partir de las ecuaciones de transición se puede generalizar la resolución del sistema con un número de time-slots n disponibles, no necesariamente $n = 8$ time-slots. Para el estado $r = 0, \dots, 2^n - 1$, la ecuación de transición de dicho estado r se expresa por medio de (3.1):

$$\begin{aligned}
 [b(i_r < n)\lambda + i_r\mu]P(r) = \\
 = b(i_r < n)\mu \sum_{\substack{j=k_r \\ j:r_j=0}}^{n-1} P(2^j + r) + b(k_r > 0)\lambda \sum_{j=0}^{k_r-1} P(r - 2^j)
 \end{aligned}
 \tag{3.1}$$

donde:

- i_r es el número de time-slots ocupados del estado r (es decir, $i_r = \sum_{j=0}^{n-1} r_j$)
- k_r es el menor número ID que está libre, igualmente el número de time-slots consecutivos ocupados al principio de la trama, ($k_r = 0$ siempre que $r_0 = 0$)
- $b(Expression)$ es 1 si la *Expression* es verdadera ó 0 en caso contrario.

Para continuar con el mismo ejemplo anterior, si consideramos $n = 8$ la ecuación de transición para el estado 5 ó ([00000101]) se reduce a la expresión (3.2):

$$\begin{aligned}
 [\lambda + 2\mu] \cdot P(5) = & \\
 \mu \cdot [P(7) + P(13) + P(21) + P(37) + & \\
 + P(69) + P(133)] + \lambda \cdot P(4) &
 \end{aligned}
 \tag{3.2}$$

La primera parte de la ecuación determina que el Estado 5 bien puede transitar a dos estados distintos si una de sus llamadas se libera o bien se convierte en un nuevo estado si llega una nueva llamada. La segunda parte de la ecuación de transición denota que el Estado 5 procede de cualquier llamada que se libera de los estados 7, 13, 21, 37, 69 y 133 ó de cualquier llamada que se produzca procedente del estado 4. Estas transiciones se muestran en la Figura 3.3.

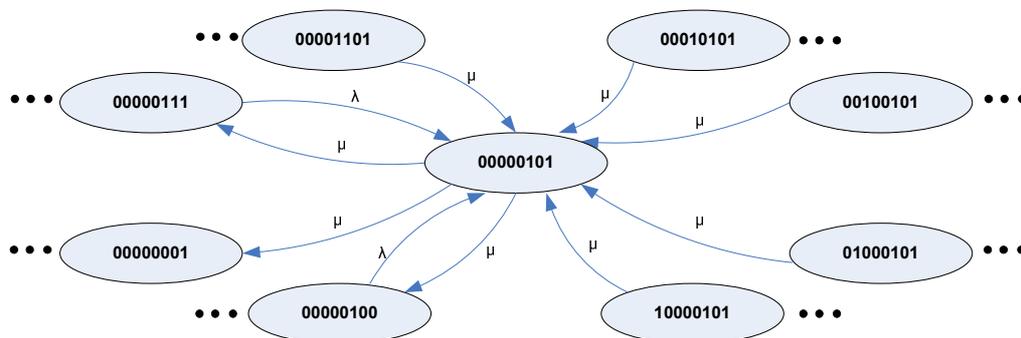


Figura 3.3: Diagrama de transición para el estado 5 con 8 time-slots

Con el objetivo de resolver el sistema de ecuaciones de transición o, de la misma forma, encontrar la probabilidad de cada uno de los estados, se ha de fijar la *Probabilidad de Bloqueo* y el tráfico que se ofrece a los recursos. En el caso usual de una probabilidad de bloqueo del 1% y tráfico ofrecido de 3,13 Erlangs, se obtiene para cada estado r su correspondiente probabilidad $P(r)$. En las Figuras 3.4 y 3.5 se puede observar, tras la resolución del sistema de ecuaciones, los valores de las probabilidades para cada uno de los estados.

Si centramos la atención en el último estado (255) podemos observar que el valor de bloqueo del sistema, es decir, de todos los time-slots ocupados corresponde justamente a los valores de 0,01 para el caso del 1 % y de 0,02 para el caso del 2 % respectivamente.

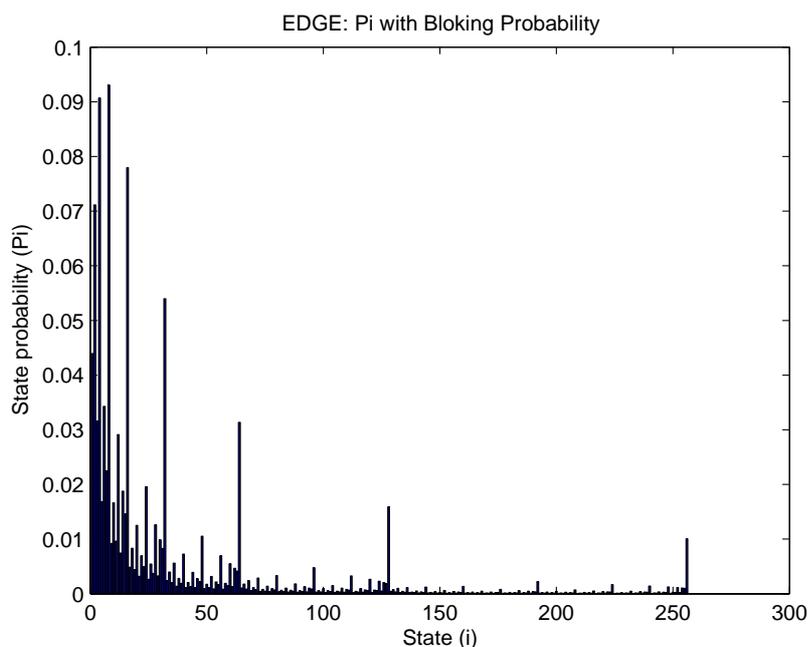


Figura 3.4: Probabilidades de encontrar estados con una probabilidad de bloqueo del 1 %

3.2.2. Concepto de ráfagas: programación y análisis

El propósito de introducir un nuevo concepto en el modelo que se estudia, a partir del conocimiento de las probabilidades de permanecer en cada uno de los estados del sistema y calculadas a partir de la resolución del sistema de ecuaciones, se está en disposición de describir dos maneras de determinar la probabilidad de encontrar un número dado de time-slots libres consecutivos. Esto significa la potencialidad de encontrar la disponibilidad de conseguir conocer el número estimado de time-slots para el transporte del tráfico de datos. A este número de time-slots libres consecutivos lo denominamos ráfaga (*burst*), es decir, la ráfaga consiste en un grupo de recursos contiguos libres

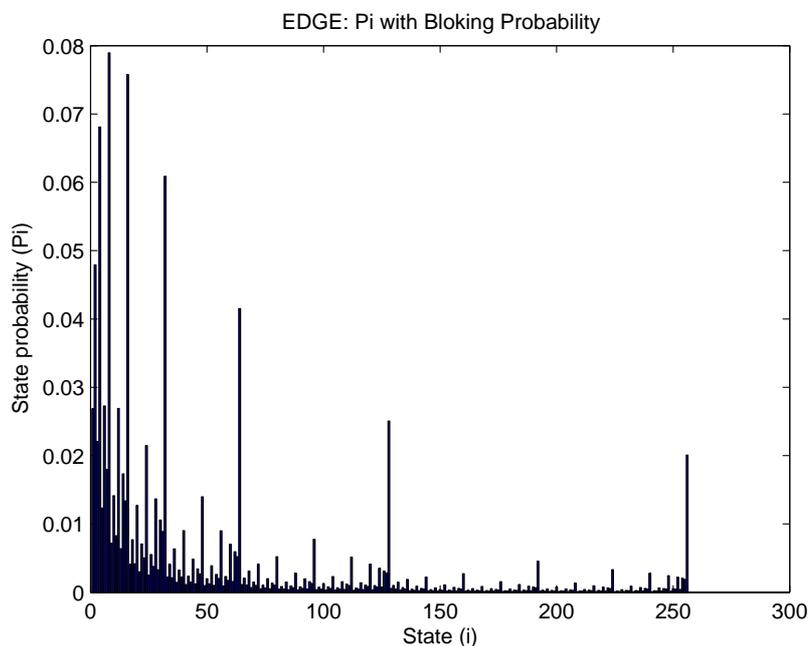


Figura 3.5: Probabilidades de encontrar estados con probabilidad de bloqueo del 2 %

disponibles para el nuevo tráfico entrante al sistema. Así, cuando el primer time-slot de una ráfaga se asigne a un usuario, éste podrá utilizar los time-slots consecutivos libres con objeto de aumentar su *throughput*, haciendo uso de la capacidad multi-slot de los terminales móviles para la transmisión de datos. Este hecho es sumamente relevante puesto que si el operador de red es capaz de conocer la cantidad de time-slots libres consecutivos de los que dispone en cada momento, en consecuencia podrá mejorar la gestión de la calidad de servicio de las aplicaciones que hacen uso de esos recursos.

Analizando cada uno de los posibles estados por los que pasa el sistema, cada uno de los estados consta de un número concreto de ráfagas de diferentes longitudes. Por ejemplo, para el caso de ocho time-slots, el estado 0 posee una ráfaga de ocho, es decir, 8 time-slots consecutivos libres, el estado 1 tiene una ráfaga de 7 o el estado 2, Figura 3.6, posee dos ráfagas: una ráfaga de 6 y otra ráfaga justo de 1 time-slot.

Para determinar, por tanto, la probabilidad de encontrar una ráfaga de una longitud determinada, será necesario conocer qué estados i contienen

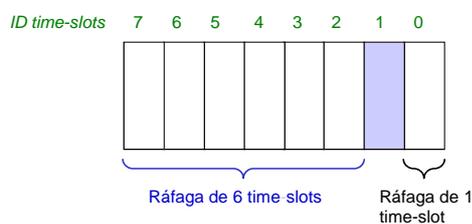


Figura 3.6: Estado 2 con ráfagas de 6 y 1 time-slots libres consecutivos

ráfagas de esa longitud y las respectivas probabilidades P_i de estos estados i . Es necesario conocer cuántas ráfagas y de qué longitudes se dispone con un número n de time-slots disponibles, además de saber cómo contribuyen a la probabilidad global de tener ráfagas de una determinada longitud. Con esta información el operador de red dispondrá de más posibilidades de usar la capacidad multislot para asignar los correspondientes canales a un usuario con determinados requisitos de QoS.

3.2.2.1. Programación de las ráfagas a partir de las probabilidades de los estados

Dado un número de time-slots disponibles, se programa un algoritmo que cuente las ráfagas que pueden encontrarse en todos los posibles estados y sus respectivas longitudes, utilizando posteriormente esta información en forma matricial. Este algoritmo calcula la probabilidad de encontrar una ráfaga de un determinado número de time-slots libres consecutivos por medio de esta matriz, cuyo contenido consiste en almacenar todas las posibles ráfagas para cada uno de los estados por los que puede pasar el sistema.

Puesto que se parte de las probabilidades P_i de permanecer en cada uno de los estados, el proceso consiste en sumar todas las probabilidades de aquellos estados de la matriz cuyas ráfagas coincidan con una longitud dada. Para dicho cálculo, las probabilidades de tener ráfagas de longitud dada, se pueden hallar de dos formas distintas:

1. Probabilidad de encontrar una ráfaga de una longitud exacta dada m . Esto consiste en obtener la suma de todas las probabilidades de aquellos estados en la matriz que cumplan que las ráfagas sean exactamente de longitud m . Hay que tener en cuenta que si un determinado estado tiene dos o más ráfagas de esta longitud m , el valor de la probabilidad del estado contribuirá una sola vez a la suma global.

2. Probabilidad de encontrar una ráfaga de al menos la longitud dada m . Esto significa que se cuentan las ráfagas con valores mayores o iguales a m time-slots libres consecutivos (probabilidades acumulativas). Por ejemplo, si un estado posee una ráfaga de longitud 4, ese estado contribuirá a la probabilidad global de encontrar ráfagas de longitud desde 1 hasta 4. Igual que en el caso anterior, incluso si un estado tiene dos ráfagas de al menos m time-slots libres consecutivos, este estado contribuye solo una vez a la suma global de probabilidades.

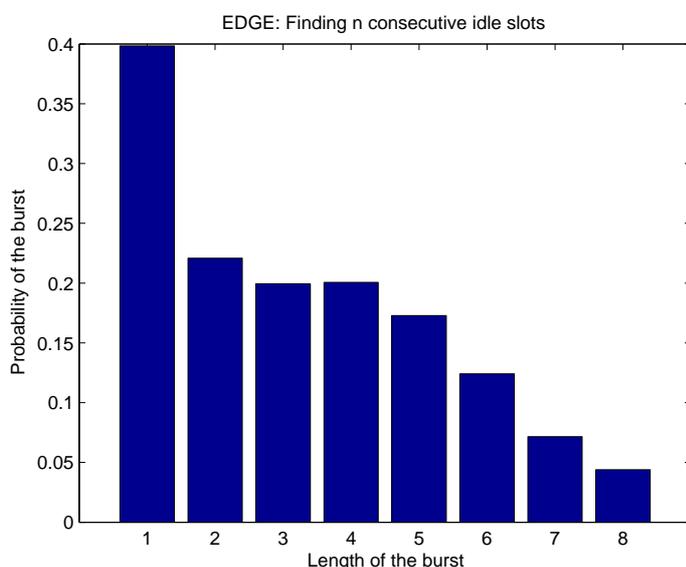


Figura 3.7: Probabilidad de encontrar ráfagas con probabilidad de bloqueo del 1 %

De acuerdo con cada uno de los casos, es posible calcular el máximo *throughput* cuando el terminal móvil posee un número de time-slots disponibles para asignar sus canales de datos. Algunos resultados para el caso de encontrar ráfagas de longitud fija se muestran en la Figuras 3.7 y 3.8, cuando el número de time-slots disponibles es $n = 8$. En la figura 3.7 se pueden observar las longitudes medias de las ráfagas fijando la probabilidad de bloqueo al 1 %, mientras que en la figura 3.8, la probabilidad de bloqueo es del 2 %. Con esta longitud máxima de ráfaga de 8 time-slots, denominamos ráfagas cortas a aquellas cuya longitud sea menor o igual a 3 y largas a aquellas ráfagas mayores o iguales a 4. Así, la gran diferencia que se puede observar radica

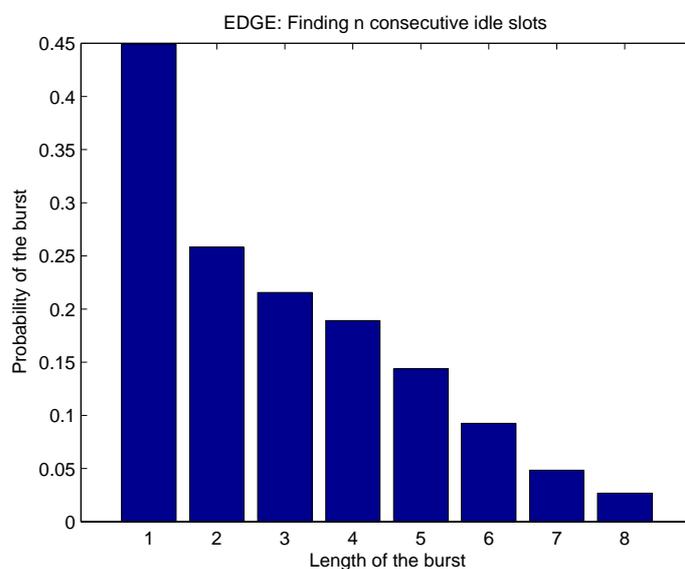


Figura 3.8: Probabilidad de encontrar ráfagas con probabilidad de bloqueo del 2 %

en el hecho de que con probabilidad de bloqueo del 1 %, la probabilidad de encontrar ráfagas largas es mayor que con probabilidad de bloqueo del 2 %. Esto significa que, al contrario de lo que se piensa a priori, las tasas de pico más altas se consiguen con una probabilidad de bloqueo menor (1 %). También incluso cuando la capacidad del sistema está cargado con tráfico de voz, la probabilidad de transmitir datos es alta con mayores tasas de pico que con probabilidad de bloqueo del 2 %. Observando nuevamente las Figuras 3.7 y 3.8, para ráfagas largas las tasas de pico son mayores con probabilidad de bloqueo del 1 % frente a las del 2 %. Es de resaltar que este hecho implica un incremento sustancial de las tasas de datos de los usuarios en el primer caso ($PB = 1\%$).

En suma, para un tráfico de voz bajo, la probabilidad de encontrar tasas altas y tasas medias de datos es alta. El tráfico de datos también puede ser mayor con probabilidad de bloqueo del 1 %, por tanto disponemos de menor tráfico cursado y más recursos disponibles para el tráfico de datos. En concreto para el caso en el que las ráfagas son de longitud mayores o iguales a 5 y probabilidad de bloqueo del 1 %, el tráfico de datos puede ser mayor que para el caso con probabilidad de bloqueo del 2 %, diferencias que se pueden

observar en la Figura 3.9.

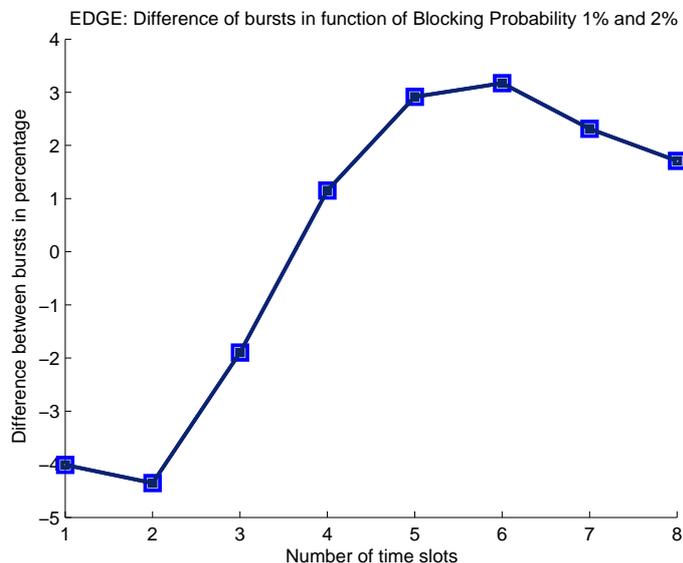


Figura 3.9: Diferencia entre tasas de pico con probabilidad de bloqueo del 1% frente al 2%

En cuanto a la probabilidad acumulativa de encontrar ráfagas con longitudes de al menos un valor m , Figura 3.10, se puede observar de cuánto tiempo dispone el operador de red para estimar el número de time-slots consecutivos que puede asignar a los usuarios. Por ejemplo, conocer cuánto tiempo dispone de tres o más time-slots libres consecutivos, Figura 3.10, se puede observar que dicho tiempo es del 80%. De la misma forma se puede deducir que la mayor parte del tiempo es posible encontrar al menos un time-slot libre disponible.

3.2.2.2. Cálculo de las ráfagas a partir de las ecuaciones de transición de estados

En esta sección presentamos cómo contar y enumerar todas las secuencias de n time-slots con al menos una ráfaga de m time-slots libres consecutivos. Dados n y m , podemos enumerar todas las secuencias de forma recursiva. Denotamos:

- S como un conjunto de enteros

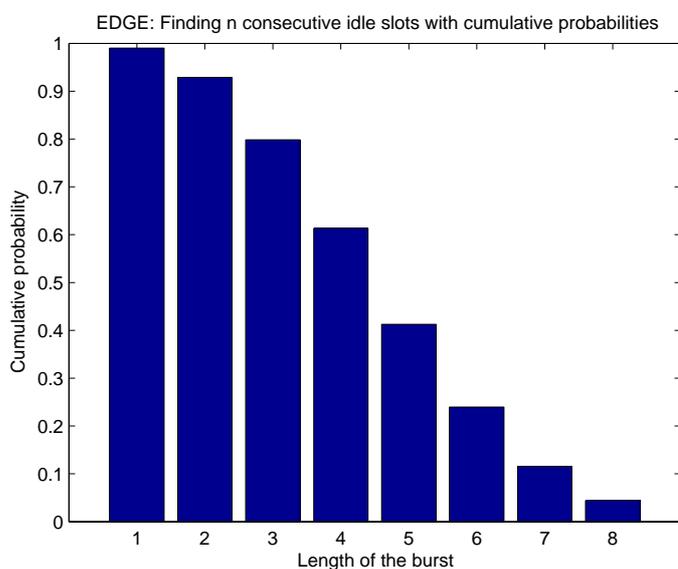


Figura 3.10: Probabilidad acumulativa de encontrar ráfagas con probabilidad de bloqueo del 1 %

- k como un entero
- $k + S$ el conjunto que consta de cada entero de S más k
- S y T dos conjuntos, tal que $S \setminus T$ contiene los elementos de S que no están en T .

A modo de ejemplo, si tomamos $S = \{0, 1\}$ y $k = 4$, $k + S = \{4, 5\}$. Por otro lado, si $S = \{0, 1, 2, 4\}$ y $T = \{1, 4\}$, $S \setminus T = \{0, 2\}$.

También nombraremos a:

- $Q(n, m)$ todos los índices de los estados de n time-slots con al menos una ráfaga de longitud m
- $C(n, m)$ complementario de Q , es decir, todos los índices de los estados de n time-slots sin ráfagas de longitud m .

Una solución eficiente para la numeración de $Q(n, m)$ para valores pequeños de m se puede expresar de forma recursiva de la siguiente manera:

$$\begin{aligned}
Q(n, m) = & \cup_{i=0}^{2^t-1} \{(1+2i)2^m\} \cup \{2^t + C(t, m)\} \cup \\
& \cup_{i=1}^t \cup_{j=0}^{2^{t-i}-1} \{(1+2j)2^{i+m} + 2^{i-1} + C(i-1, m)\},
\end{aligned} \tag{3.3}$$

para $n > m + 1$ y $t = n - m - 1$. Esta expresión es recursiva, por tanto puede ser escrita incluyendo solo los elementos previos $Q(i, m)$ sabiendo que $Q(i, m)$ y $C(i, m)$ son complementarios. Así, se puede reemplazar cada $C(i, m)$ que aparece en (3.3) por:

$$C(i, m) = \cup_{k=0}^{2^i-1} \{k\} \setminus Q(i, m).$$

Los valores de contorno de $Q(n, m)$ para (3.3) son:

$$Q(n, m) = \begin{cases} \emptyset & \text{si } n < m \\ \{0\} & \text{si } n = m \\ \{1, 2^m\} & \text{si } n = m + 1. \end{cases}$$

La ecuación (3.3) enumera todos los elementos de $Q(n, m)$ buscando la primera ráfaga de longitud m comenzando por la parte derecha de la secuencia, el símil con números binarios serían los menos significativos de una secuencia binaria, de tal forma que se asegura que las uniones en (3.3) son disjuntas.

Así, $\cup_{i=0}^{2^t-1} \{(1+2i)2^m\}$ incorpora todos los índices que poseen la primera ráfaga de m time-slots libres consecutivos a partir del time-slot ID 0 hasta el ID $m - 1$. Esta expresión contempla que pueda haber eventualmente más ráfagas de longitud m que aparezcan próximas por la parte más significativa de la secuencia (zona derecha de la secuencia).

De forma similar, la segunda unión $2^t + C(t, m)$ incorpora todos los índices que tienen la primera ráfaga de longitud m desde el time-slot con ID $n - m$ hasta el ID $n - 1$. En consecuencia no hay más ráfagas de longitud m que aparezcan más tarde en la secuencia y es así porque se necesita el conjunto $C(t, m)$, de índices sin ráfagas de longitud m en el resto de la secuencia.

La unión doble establece todos los índices donde la primera ráfaga de longitud m (comenzado desde la parte derecha de la secuencia) aparecen en la mitad de la secuencia. Es decir, existe una ráfaga de longitud m con los

time-slots ocupados a la derecha de la secuencia después de haber encontrado la ráfaga y antes de encontrar la ráfaga. El índice i de la doble unión recorre todas las posibilidades de un time-slot ocupado en el time-slot ID $i - 1$, m time-slots desde el ID i hasta el ID $i + m - 1$ y un time-slot ocupado en el time-slot con ID $i + m$. Luego no existen ráfagas de longitud m antes de la posición i y es por lo que aparece $C(i - 1, m)$ en la doble unión.

Además no hay restricciones por las que puedan aparecer después del time-slot ocupado en el time-slot con ID $i + m$ y también teniendo en cuenta todas las posibles subsecuencias después del time-slot ocupado.

La representación gráfica de la expresión (3.3) para un caso concreto se muestra en la Figura 3.11, en las que se representan todos los estados con $n = 8$ conteniendo ráfagas de longitud $m = 2$ en el time slot cuyo ID es 4 y 5 (representados en un sombreado gris claro frente al gris oscuro que representan los time-slots ocupados).

Todos los estados de la Figura 3.11 se enumeran por medio de la expresión (3.3) en la doble unión para $i = 4$ excepto para cuatro de ellos: los estados 73 y 201 que contienen una ráfaga previa con longitud $m = 2$ contemplada en la doble unión cuando $i = 1$, y de forma similar, los estados 76 y 204 que se cuentan antes dentro de la primera unión.

En (3.3) para computar $Q(n, m)$ se puede simplificar en muchos casos, algunas veces sin necesidad de utilizar el método recursivo. Por ejemplo, para cualquier $2m \geq n > m + 1$, en tal caso, existe solo una ráfaga de m time-slots libres consecutivos y no hay ráfagas con longitudes mayores, así no es necesario contar los elementos previos $Q(i, m)$ una vez tengamos localizadas las ráfagas de longitud m . Esto implica que la expresión cerrada (3.4) a la que llegamos para $Q(n, m)$ es:

$$Q(n, m) = \bigcup_{i=0}^{2^t-1} \{(1 + 2i)2^m\} \cup \bigcup_{i=0}^{2^t-1} \{2^t + i\} \cup \bigcup_{i=1}^t \bigcup_{k=0}^{2^{i-1}-1} \bigcup_{j=0}^{2^{t-i}-1} \{(1 + 2j)2^{i+m} + 2^{i-1} + k\} \quad (3.4)$$

para $2m \geq n > m + 1, t = n - m - 1$.

cuyo cardinal es: $|Q(n, m)| = 2^{t+1} + t \cdot 2^{t-1}$.

Un ejemplo que muestra estas recurrencias para el caso donde $n = 8$,

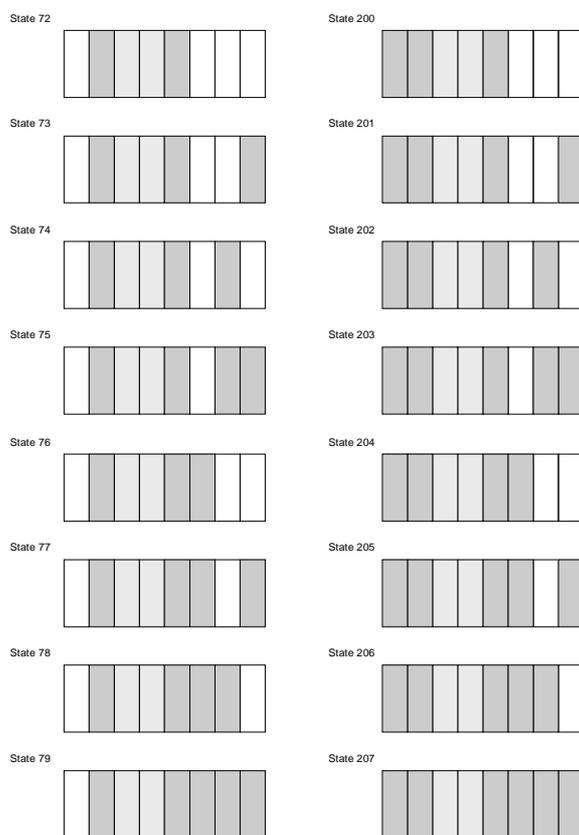


Figura 3.11: Representación de los estados con ráfagas de $m=2$ (gris claro) en posiciones central-izquierda

si $m = 8$: $Q(8, 8) = \{0\}$, cuyo cardinal es $|Q(8, 8)| = 1$;

$m = 7$: $Q(8, 7) = \{1, 128\}$ cuyo cardinal es $|Q(8, 7)| = 2$;

$m = 6$: $Q(8, 6) = \{2, 3, 64, 129, 192\}$ cuyo cardinal es $|Q(8, 6)| = 5$;

$m = 5$: $Q(8, 5) = \{4, 5, 6, 7, 32, 65, 96, 130, 131, 160, 193, 224\}$ cuyo cardinal es $|Q(8, 5)| = 12$;

$m = 4$:

$$Q(8, 4) = \{8, 9, 10, 11, 12, 13, 14, 15, 16, 33, 48, 66, 67, 80, 97, 112, 132, 133, 134, 135, 144, 161, 176, 194, 195, 208, 225, 240\}.$$

cuyo cardinal es $|Q(8, 4)| = 28$.

En el caso de $m = 3$ cuyo cardinal es $|Q(8, 3)| = 61$, los valores son:

$$Q(8, 3) = \{8, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, \\ 34, 35, 40, 49, 56, 68, 69, 70, 71, 72, 81, 88, 98, 99, 104, \\ 113, 120, 136, 137, 138, 139, 140, 141, 142, 143, 145, 152, \\ 162, 163, 168, 177, 184, 196, 197, 198, 199, 200, 209, 216, \\ 226, 227, 232, 241, 248, \}.$$

En el caso de $m = 2$ cuyo cardinal es $|Q(8, 2)| = 120$, los valores son:

$$Q(8, 2) = \{4, 9, 12, 18, 19, 20, 25, 28, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, \\ 42, 43, 44, 45, 46, 47, 48, 49, 50, 51, 52, 53, 54, 55, 56, 57, 58, 59, \\ 60, 61, 62, 63, 68, 72, 73, 74, 75, 76, 77, 78, 79, 82, 83, 84, 89, 92, \\ 100, 101, 102, 103, 105, 108, 114, 115, 116, 121, 124, 132, 137, 140, \\ 144, 145, 146, 147, 148, 149, 150, 151, 152, 153, 154, 155, 156, 157, \\ 158, 159, 164, 165, 166, 167, 169, 172, 178, 179, 180, 185, 188, 196, \\ 200, 201, 202, 203, 204, 205, 206, 207, 210, 211, 212, 217, 220, 228, \\ 229, 230, 231, 233, 236, 242, 243, 244, 249, 252\}.$$

El objetivo de estas expresiones no es más que, finalmente, obtener la suma global de todas las probabilidades de los estados que encontramos con $Q(n, m)$, así conseguiremos la probabilidad de encontrar una ráfaga de longitud m en una secuencia de longitud n . En concreto, un ejemplo en el que los valores obtenidos de las probabilidades de encontrar una determinada ráfaga de longitud m , se muestran gráficamente en la sección anterior en la Figura 3.7 y la Figura 3.8.

3.3. Esquema de asignación repacking

3.3.1. Modelo repacking

Recordemos que el método de *repacking* optimiza la utilización de los canales a un extremo del repositorio, lo que significa que los recursos estarán

disponibles para los usuarios de forma consecutiva. Este método hace uso del *intracell handover* (traspaso dentro de la misma célula) consistente en la re-asignación al time-slot libre con el menor número de identificador, todo ello para aquellas llamadas que ocupen posiciones no contiguas. Este método, en general, puede incluir la asignación contigua entre distintas portadoras de frecuencia dentro de la misma celda. El gran inconveniente de este método es que si se produce un gran número de trasposos dentro de la misma celda durante una llamada se pueden producir efectos negativos en cuanto a la calidad del servicio. En particular para períodos de expulsión de una llamada por encima de los $450ms$ de duración puede ocurrir que cada vez que un usuario en una llamada se reubica a otra posición se producen deficiencias en el servicio ofrecido al usuario.

Este modelo determina que solo pueda existir con n recursos $n+1$ posibles estados. Para el caso de $n = 8$ los estados serán: 0, 1, 2, 3, 4, 5, 6, 7 y 8, correspondiendo con los time-slots ocupados en cada momento. Gráficamente este modelo se muestra en la Figura 3.12.

De forma análoga a como hemos descrito para el modelo secuencial, podemos formular las siguientes expresiones que definen el mecanismo de asignación *repacking*:

$$\begin{aligned}
 \lambda P_0 - \mu P_1 &= 0 \\
 -\lambda P_0 + (\lambda + \mu) P_1 - 2\mu P_2 &= 0 \\
 -\lambda P_1 + (\lambda + 2\mu) P_2 - 3\mu P_3 &= 0 \\
 -\lambda P_2 + (\lambda + 3\mu) P_3 - 4\mu P_4 &= 0 \\
 -\lambda P_3 + (\lambda + 4\mu) P_4 - 5\mu P_5 &= 0 \\
 -\lambda P_4 + (\lambda + 5\mu) P_5 - 6\mu P_6 &= 0 \\
 -\lambda P_5 + (\lambda + 6\mu) P_6 - 7\mu P_7 &= 0 \\
 -\lambda P_6 + (\lambda + 7\mu) P_7 - 8\mu P_8 &= 0 \\
 -\lambda P_7 + 8\mu P_8 &= 0
 \end{aligned}$$

Podemos generalizar que con un número de time-slots n y para cada estado r se puede concluir que expresión que permite calcular las probabilidades de permanecer en cualquiera de los posibles estados por los que pasa el sistema es (3.5):

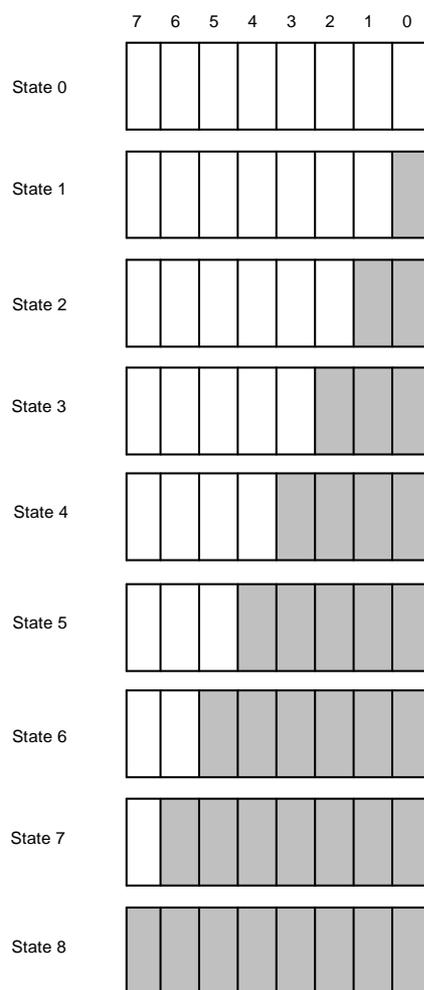


Figura 3.12: Representación de los estados en el método repacking

$$\lambda P_r \varphi(r+1) + r \mu P_r \varphi(r-1) = \lambda P_{r-1} \varphi(r) + (r+1) \mu P_{r+1} \varphi(r) \quad (3.5)$$

donde sea $\Psi = 1..n$ el conjunto de posibles estados y por medio de la función $\varphi(r)$ se indica si el estado r es posible o no, es decir, $\varphi(r)$ toma dos valores posibles, el valor 1 cuando el estado es factible y el valor 0 en caso contrario. En concreto, los posibles estados en cada una de las ecuaciones son el anterior y posterior, excepto en el caso de la ecuación referente al primer estado, 0, que no tiene anterior, y el último estado n que no tiene posterior.

A partir de estas ecuaciones y satisfaciendo la ecuación donde la suma de todas las probabilidades es la unidad, calcularemos las probabilidades de estar en cada uno de los $n + 1$ estados.

La situación final (Figura 3.12) implica que de forma transitoria, cuando se produzca un hueco en el conjunto de time-slots a asignar, es necesario comprobar que todos los time-slots asignados sean contiguos. En concreto, en la mayoría de los casos esto no es así, por tanto, es necesario realizar el traspaso dentro de la misma celda, *intracell handover*. Esto significa que en la misma celda ha de haber un nuevo lugar para la asignación. Esta situación se muestra en la Figura 3.13.

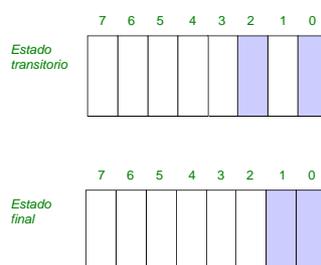


Figura 3.13: Proceso de intra-cell handover

3.3.2. Resultados

Se puede observar en la Figura 3.14 que para $n = 8$ la función de probabilidad es fija y mucho más probable encontrar seis time-slots consecutivos libres con una Probabilidad de Bloqueo del 1% que en el caso en el que el bloqueo sea del 2% (Figura 3.15) donde es más probable encontrar cuatro time-slots libres que en el caso de bloqueo del 1%. No obstante, en ambos casos es visible que la probabilidad de encontrar ráfagas de 5 time-slots es la mayor posible.

Con este método, podemos suponer que el mecanismo de sincronización puede tener problemas en la relación portadora/interferencia (C/I) con las celdas vecinas, en asignación de frecuencias próximas. Sin embargo, las estaciones base no se sincronizan con el mismo reloj y, por tanto, este hecho permite que los traspasos dentro de la misma celda (*intracell handover*) no afecten a celdas próximas o vecinas.

En otro orden de cosas, una cuestión es el hecho de que se necesitan un número importante de traspasos (Figura (3.13)) dentro de la celda para

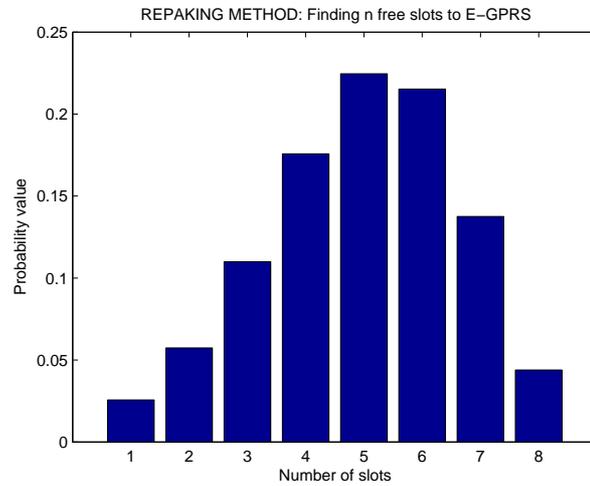


Figura 3.14: Distribución de probabilidad con una probabilidad de bloqueo del 1% usando el método repacking

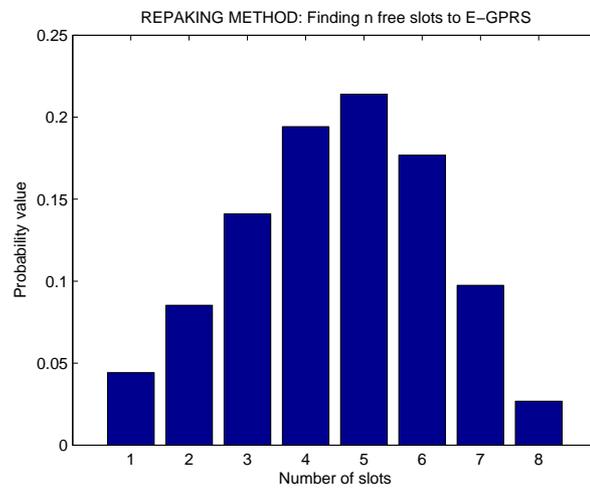


Figura 3.15: Distribución de probabilidad con una probabilidad de bloqueo del 2% usando el método repacking

obtener un mapa de asignación compacto como es el método *repacking*. Esto supone lograr los estados mostrados en la Figura 3.12 y si disponemos de una media de k llamadas por unidad de tiempo y hemos calculado la tasa de repacking o el número de trasposos en la celda, podemos calcular el número de intracell handovers por llamada. Por ejemplo, si tenemos k llamadas por unidad de tiempo, la tasa de intracell handovers se calcula con las probabilidades obtenidas anteriormente y las tasas de terminaciones de llamadas. Así es posible tener una idea de cuanto se carga el sistema con el mecanismo *repacking* partiendo del número medio de llamadas por unidad de tiempo.

3.4. Conclusiones

Se puede concluir que, por un lado, la localización para el tráfico de voz de los time-slots de forma consecutiva en el tiempo (método *First Fit*) implica un mecanismo que facilita la ubicación posterior del tráfico de datos en posiciones también contiguas en el tiempo. Por otro lado, el modelado de este mecanismo permite conocer las probabilidades de tener estados con time-slots libres consecutivos, es decir, ráfagas de recursos que, distribuidos por el operador de red en función de la demanda de los usuarios, significa mejorar la eficiencia y el rendimiento medido en términos de QoS.

De forma obvia, con los resultados obtenidos y, de forma general, se concluye que las probabilidades de encontrar ráfagas cortas de 1, 2 y 3 time-slots libres consecutivos, suponen el mayor porcentaje de tiempo del sistema. Este hecho significa la posibilidad de gestionar los recursos directamente atendiendo a las aplicaciones que requieran durante más tiempo un ancho de banda menor y en términos de planificación (*scheduling*), la posibilidad de optar por mecanismos que favorezcan durante más tiempo los paquetes/radio bloques más cortos con objeto de reducir el retardo del sistema.

Es de resaltar que fijada la probabilidad de bloqueo del sistema, si ésta es del 1% se obtienen ráfagas para la transmisión de datos mayores que en el caso de fijar el bloqueo del sistema al 2%. Implica que con menor probabilidad de bloqueo mayores recursos disponibles y mayor posibilidad de emplear la máxima capacidad multislot posible del terminal de usuario para conseguir un *throughput* de pico mayor.

Asimismo se ha presentado cómo empleando un método recursivo podemos contar y enumerar todas las secuencias de n time-slots con al menos un

ráfaga (time-slots libres consecutivos) de valor m .

Por último se presentan en el método de *repacking* los resultados de probabilidades que arrojan probabilidades mayores de tener 4 y 5 time-slots libres consecutivos y, por tanto, mayor probabilidad de aprovechar la capacidad multislot, si bien hay que tener en cuenta que este método produce inconvenientes cuando el número de trasposos dentro de la misma celda (in-tracell handover) aumenta.

Capítulo 4

Asignación con reserva de recursos para el sistema EDGE

4.1. Introducción

La demanda cada vez mayor de la transmisión de datos por parte de los usuarios a través de dispositivos móviles, ha constituido uno de las demandas más desarrolladas en la última década. En este contexto la adecuación de la integración en las redes móviles celulares de los servicios de voz y de datos es una cuestión a resolver de la forma más óptima posible. En el sistema EDGE la coexistencia de un número considerable de conexiones lógicas y la multiplexación de los paquetes de datos en los canales físicos es justamente el centro del interés de realizar la asignación de estos canales de datos y mejorar el caudal de este tipo de servicio. Esta asignación de los canales sin degradar y manteniendo el grado de servicio de las comunicaciones de voz, implican nuevos métodos de gestionar los recursos para satisfacer los requisitos solicitados por los usuarios.

En este capítulo presentamos el modelo de asignación de recursos para tráfico de voz y tráfico de datos, teniendo en cuenta que, haciendo uso de la reserva de recursos para el tráfico de voz, podemos satisfacer la calidad de servicio demandada por los usuarios independientemente del tipo de tráfico.

4.2. Reserva de recursos radio en EDGE

Dado que en el capítulo anterior hemos presentado los conceptos de ráfagas y el modelo básico de asignación compartido por ráfagas, en esta sección desarrollamos estos conceptos teniendo en cuenta que un conjunto de time-slots están reservados para un tipo de tráfico específico. En concreto se trata de repartir el número de recursos, de tal manera, que algunos de los recursos sean compartidos por el tráfico de voz y el tráfico de datos, y por otro lado, se disponga de recursos exclusivamente para un tipo de tráfico.

El problema consiste en encontrar time-slots disponibles para la transmisión de los datos pero además que estos recursos se encuentren de manera consecutiva en el tiempo para aumentar la tasa de transferencia asignada a los usuarios.

4.2.1. Esquema de asignación de recursos compartida con reserva para tráfico de voz

El esquema de asignación de recursos compartidos junto con la reserva de recursos para el tráfico de voz consiste, básicamente, en dividir los recursos en dos tipos distintos: los time-slots reservados para las llamadas de voz y los time-slots compartidos tanto para el tráfico de voz como para el tráfico de datos (Figura 4.1). A partir de esta distinción, la ocupación de los recursos dependerá del tipo de tráfico entrante. Así, para el tráfico de datos se localizarán de forma correlativa aquellos time-slots libres que tengan el menor número de identificador. La ocupación de los time-slots para el tráfico de datos dependerá, por un lado, de la exigencia de transmisión concertada con la red y, por otro lado, de la capacidad multislot del terminal móvil. El número de time-slots que ocupen los datos estará en función de la disponibilidad en cuanto al número de recursos consecutivos libres. Esto implicará la ocupación secuencial de un número de recursos con números de identificador consecutivos a lo largo de la trama.

Es de resaltar que la situación de fragmentación de los recursos es la que se trata de minimizar, de tal forma que cuando una estación móvil requiere para la transmisión de los datos contar con k recursos, se tiene que disponer de los k time-slots consecutivos libres requeridos por ese usuario para satisfacer los requisitos de QoS. En el caso en el que los time-slots libres no sean consecutivos, la asignación a ese usuario implicará una reducción del

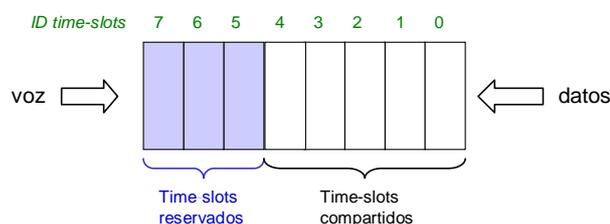


Figura 4.1: Asignación de time-slots con reserva para tráfico de voz

throughput proporcional al número de time-slots de los que se prescindía.

En lo que respecta a las llamadas de voz, éstas ocuparán por cada una de las llamadas el time-slot con el número de identificador mayor disponible. Una situación en la que se produce un traspaso dentro de la misma portadora es la que, tras una liberación de un time-slot reservado para voz bien porque ha terminado la llamada o bien porque se ha realizado un traspaso a otra celda, habiendo al menos una llamada de voz ocupando algún time-slot de los denominados compartidos, se producirá entonces un traspaso en la misma celda (*intracell handover*) y se pasará a ocupar el recurso reservado recién liberado por la llamada de voz que ocupe el time-slot con el número menor de identificador de los recursos compartidos, de esta forma se produce un desplazamiento hacia la izquierda, según la Figura 4.1, agrupando las llamadas de voz en los recursos con números de identificador mayores.

4.2.2. Modelo analítico

En esta sección se describe el modelo analítico correspondiente al esquema de asignación compartida haciendo uso de la reserva de recursos para el tráfico de voz. Así, se pueden deducir las ecuaciones de transición de estados para este modelo, de tal forma que, independientemente del número de time-slots que componga cada una de las tramas, las ecuaciones son completamente genéricas y no limitadas a un número concreto de time-slots reservados para el tráfico de voz ni a un número determinado de time-slots compartidos para los tráficos de datos y de voz.

Sea una trama con una secuencia de longitud n , en la que una posición se codifica con:

- 0 si el time-slot con el correspondiente número de identificador está libre,

- 1 en el caso que el time-slot está ocupado con tráfico de datos y
- 2 si el time-slot está ocupado con tráfico de voz.

En consecuencia, un estado concreto estará definido por medio de una secuencia $[r_{n-1}, \dots, r_1, r_0]$ siendo $r_i = \{0, 1, 2\}$. Es de resaltar que no todas las 3^n secuencias de este tipo son los estados posibles puesto que algunos de los recursos están reservados únicamente para tráfico de voz. Si m es el número de time-slots compartidos para voz y datos, y $n - m$ es el número de time-slots reservados para el tráfico de voz, en el estado definido por $[r_{n-1}, \dots, r_m, r_{m-1}, \dots, r_1, r_0]$, los time-slots con números de identificadores desde el valor 0 hasta el valor $m - 1$ corresponden a los time-slots compartidos y los recursos con número de identificador que van desde el valor m hasta el $n - 1$ corresponden con los time-slots reservados para voz.

El número de estados posibles con este esquema es:

$$Num_{estados} = 2^n + 3^m - 2^m.$$

Una forma de verificar esta expresión es contar todos los posibles estados en los que ninguna llamada de voz ocupa los time-slots compartidos, es decir, 2^n . Esto se debe a que un time-slot con un número de identificador cuyo rango va desde 0 hasta $m - 1$ toma solo dos posibles valores 0 ó 1 y cualquier time-slot cuyo número de identificador esté comprendido entre m y $n - 1$ será bien 0 ó 2. Luego para contar todos los posibles estados en los que los time-slots reservados están ocupados y existe tráfico de voz en los time-slots compartidos, implica que tendremos un total de 3^m estados posibles en esta situación. Los time-slots con números de identificador entre m y $n - 1$ tendrán todos ellos el valor igual a 2 y los time-slots con números de identificador entre 0 y $m - 1$ han de ser bien 0 ó bien 1 ó bien 2, es decir, time-slot desocupado, ocupado con tráfico de datos y ocupado con tráfico de voz, respectivamente. Esto significa que cuando todos los time-slots reservados están ocupados y todos los time-slots compartidos toman los valores de 0 ó 1, contamos doblemente 2^m estados.

Ahora hemos de distinguir entre dos bloques de ecuaciones de transición. El primer bloque corresponde a las ecuaciones de transición para los estados en los que no están ocupados todos los time-slots reservados y, por extensión, aquellos estados en los cuales hay al menos un time-slot reservado libre. El segundo bloque corresponde con las ecuaciones de transición de aquellos estados en los que todos los time-slots reservados para el tráfico de voz están ocupados y, posiblemente, también con tráfico de voz ocupando parte de los time-slots compartidos además de tráfico de datos.

Enumeramos los estados con números decimales. Primero, denotamos los estados en los que no están ocupados todos los time-slots reservados y los representamos como números en base 2. Primero consideramos que una posición toma un valor 1 si el time-slot con el número de identificador correspondiente está ocupado, bien sea de tráfico de voz o bien sea tráfico de datos, y 0 en caso contrario, tal como se muestra a continuación:

Número estado	Vector	
	$n-m$	m
0	[00 ... 00	0 ... 00]
1	[00 ... 000	... 01]
2	[00 ... 000	... 10]
3	[00 ... 000	... 11]
\vdots	\vdots	\vdots
$2^m - 1$	[00 ... 001	... 11]
2^m	[00 ... 020	... 00]
$2^m + 1$	[00 ... 020	... 01]
\vdots	\vdots	\vdots
$2^{n-1} - 1$	[02 ... 221	... 11]
2^{n-1}	[20 ... 000	... 00]
$2^{n-1} + 1$	[20 ... 000	... 01]
\vdots	\vdots	\vdots
$2^n - 2^m - 2$	[22 ... 201	... 10]
$2^n - 2^m - 1$	[22 ... 201	... 11]

Luego se enumeran los estados en los cuales todos los recursos reservados están ocupados, comenzando por el estado $2^n - 2^m$, que corresponde a $[2, \dots, 2, 0, \dots, 0]$. Consideramos las posiciones compartidas en base 3 puesto que los time-slots pueden estar libres, ocupados por datos u ocupados por voz. Así, una posición tomará el valor 2 si el time-slot con el correspondiente número de identificador está ocupado por una llamada de voz, el valor 1 si está ocupado por datos y 0 si el time-slot está libre, tal como se muestra en la siguiente tabla:

Número estado	Vector
$2^n - 2^m$	$[2 \dots 2 \overbrace{0 \dots 0}^{n-m} 00]$
$2^n - 2^m + 1$	$[2 \dots 20 \dots 01]$
$2^n - 2^m + 2$	$[2 \dots 20 \dots 02]$
\vdots	\vdots
$2^n - 2^{m+1} + 3^m$	$[2 \dots 22 \dots 21]$
$2^n - 2^m + 3^m - 1$	$[2 \dots 22 \dots 22]$

Conviene resaltar que esta clasificación de los estados en dos bloques no es totalmente estricta pues, en algunos casos, hay estados de un bloque que se pueden convertir en estados del otro bloque. El caso en que una llamada se libera o llega una nueva llamada, por ejemplo, si una llamada de voz se libera en el estado $[2, \dots, 2, 0, \dots, 0]$ desde el segundo bloque - se asume que ésta es la llamada que ocupa la posición gráficamente más hacia la izquierda (en notación binaria sería la posición más significativa) - el estado se convertiría en $[0, 2 \dots, 2, 0, \dots, 0]$, que representa el estado $2^{n-1} - 2^m$ y uno de los estados del primer bloque. Este tipo de cambio de base se representa en las ecuaciones de transición por medio de un término auxiliar K , que solo se necesitaría en el caso de contorno.

4.2.3. Resolución del modelo

El sistema de ecuaciones que resuelve el modelo expuesto, supone la distinción entre el tráfico de voz y de datos, lo cual implica un desarrollo mayor de las ecuaciones presentadas en el modelo básico sin reserva de time-slots. Para ello es necesario definir las notaciones generales en las que un estado r , tendrá $i_{r,v}$ y $i_{r,d}$ que representan el número total de time-slots ocupados con el tráfico de voz y el tráfico de datos respectivamente. De forma análoga, sea $k_{r,v}$ y $k_{r,d}$ el número de time-slots consecutivos ocupados de la trama en sentido izquierda-derecha para tráfico de voz y tráfico de datos respectivamente.

Para todos los estados r en los que no todos los recursos reservados para voz están ocupados, el subsistema de ecuaciones de transición corresponde a la siguiente expresión:

$$[\lambda_v + b(i_{r,d} < m) \cdot \lambda_d + i_{r,v} \cdot \mu_v + i_{r,d} \cdot \mu_d]P(r) =$$

$$\begin{aligned}
&= b(i_{r,v} < n - m - 1) \cdot \mu_v \cdot \sum_{\substack{j=k_{r,v} \\ j:r_{n-1-j}=0}}^{n-m-1} P(2^{n-1-j} + r) + \\
&+ b(i_{r,v} = n - m - 1) \cdot \mu_v \cdot P(2^m(2^{n-m} - 1) + K) + \\
&+ b(i_{r,d} < m) \cdot \mu_d \cdot \sum_{\substack{j=k_{r,d} \\ j:r_j=0}}^{m-1} P(2^j + r) + \\
&+ b(k_{r,v} > 0) \cdot \lambda_v \cdot \sum_{j=0}^{k_{r,v}-1} P(r - 2^{n-1-j}) + \\
&+ b(k_{r,d} > 0) \cdot \lambda_d \cdot \sum_{j=0}^{k_{r,d}-1} P(r - 2^j),
\end{aligned}$$

donde

$$K = \sum_{j=0}^{m-1} r_j \cdot 3^j.$$

En resumen, en el caso específico en el que no exista ningún time-slot reservado ($n = m$), las ecuaciones de transición son muy similares a las expresadas en el capítulo anterior. No son exactamente iguales puesto que hay que distinguir entre los distintos tipos de tráfico en este segundo modelo y considerar 3^n estados distintos y no 2^n , como en el modelo básico.

A modo de ejemplo las ecuaciones para el caso de $n = 4$ y $m = 2$, se puede determinar que el estado 9 se representa como $[2001] \sim [V00D]$ y su ecuación de transición es:

$$\begin{aligned}
&[\lambda_v + \lambda_d + \mu_v + \mu_d]P(9) = \\
&\mu_v \cdot P(13) + \mu_d \cdot P(11) + \lambda_v \cdot P(1) + \lambda_d \cdot P(8),
\end{aligned}$$

La primera parte de la ecuación determina que el estado 9 puede transitar a cuatro estados distintos: dos son posibles si una llamada de cualquier tipo se libera y los otros dos estados corresponden a una llamada de cualquier tipo que pueda llegar. La segunda parte de la ecuación de transición denota que el estado 9 puede adquirirse siempre que una llamada de voz sea liberada procedente del estado 13, así como una llamada de datos se libere procedente

del estado 11, también si llega una llamada estando bien en el estado 1 o bien datos estando en el estado 8 (Fig. 4.2).

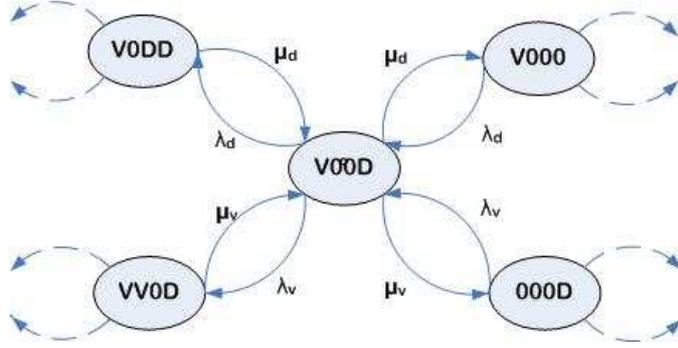


Figura 4.2: Diagrama de transición para el estado 9

Llegados a este punto se definen las ecuaciones de los estados en los cuales todos los time-slots están ocupados. Es necesario una notación adicional para este bloque de ecuaciones: primero, en el caso que se realice un traspaso en la misma celda (*intracell - handover*) cada vez que un recurso reservado se libere (si hubiera time-slots compartidos que están ocupados con tráfico de voz), se necesita conocer el número de identificador más pequeño de un time-slot ocupado por una llamada de voz puesto que la llamada en ese recurso se asignará al time-slot reservado que se haya liberado. Para un estado cualquiera r , sea l_r , el número de identificador correspondiente a ese estado. En segundo lugar, sea $\lambda_{v,p}$ la tasa de pérdidas de las llamadas de voz para los time-slots reservados, dada una Probabilidad de Bloqueo para los time-slots reservados, BP_R , será: $\lambda_{v,p} = \lambda_v \cdot BP_R$.

En consecuencia, teniendo en cuenta la notación expuesta, para los estados r en los que todos los time-slots estén ocupados, la relación de ecuaciones de transición correspondiente será:

$$\begin{aligned}
 & [b(i_{r,d} + i_{r,v} < n - 1)(\lambda_v + \lambda_d) + i_{r,v} \cdot \mu_v + i_{r,d} \cdot \mu_d]P(r) = \\
 & = b(i_{r,v} + i_{r,d} < n - 1) \left[\sum_{\substack{j=k_{r,d} \\ j:r_j=0}}^{m-1} (\mu_d P(3^j + r) + \mu_v P(2 \cdot 3^j + r)) \right] +
 \end{aligned}$$

$$\begin{aligned}
& \left. +b(k_{r,d} < l_r) \cdot \sum_{\substack{j=k_{r,d} \\ j:r_j=0}}^{l_r-1} (n-m) \cdot \mu_v \cdot P(2 \cdot 3^j + r) \right] + \\
& +b(k_{r,d} > 0) \cdot \lambda_d \cdot \sum_{\substack{j=0 \\ j:r_j=1}}^{k_{r,d}-1} P(r - 3^j) + \\
& +b(k_{r,v} > n-m) \cdot \lambda_{v,p} \cdot \sum_{\substack{j=n-m-1 \\ j:r_{n-j}=2}}^{k_{r,v}} P(r - 2 \cdot 3^{n-j}) + \\
& +b(i_{r,v} = n-m-1) \cdot \lambda_v \cdot \sum_{j=m}^{n-1} P(K - 2^j),
\end{aligned}$$

where

$$K = \sum_{j=0}^{m-1} r_j \cdot 2^j + \sum_{j=m}^{n-1} 2^j.$$

Para ilustrar estas ecuaciones, podemos poner como ejemplo con $n = 4$ y $m = 2$, el estado 20 se representa como $[2222] \sim [VVVV]$ y su ecuación de transición es:

$$4 \cdot \mu_v \cdot P(20) = \lambda_v \cdot (P(18) + P(14))$$

La ecuación correspondiente al estado 20 aparece bien por una llegada de una llamada de voz procedente del estado 14 ($[2220]$) ó bien del estado 18 ($[2202]$). La primera parte de la ecuación determina que el estado 20 pueda convertirse en cuatro estados si una de las llamadas de voz se libera, no necesariamente en distintos estados puesto que en este caso se realizaría el traspaso dentro de la misma celda (*intracell - handover*). Es de resaltar que $[2022]$ no es un estado posible puesto que hay un time-slot reservado libre y se realizaría el traspaso dentro de la misma celda. El estado posible correspondiente para esta situación sería el $[2220]$, esto se debe a que la llamada de voz con el número menor de identificador se traslada para ocupar el recurso que ocupaba la llamada liberada en los time-slots reservados. Así, si las llamadas ocupan números de identificador 0, 2 ó 3 y estos se liberan, el estado 20 transitará al estado 14 con un factor de probabilidad de $3\mu_v$ y

cuando la llamada ocupe el time-slot cuyo número de identificador es el 1 el estado 20 se puede convertir en el estado 18 (Fig. 4.3).

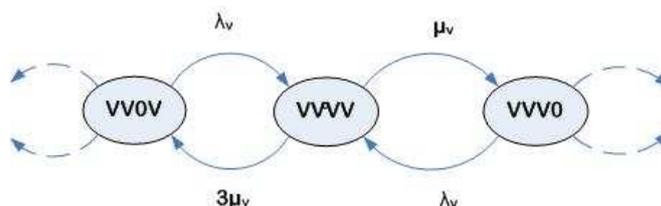


Figura 4.3: Diagrama de transición para el estado 20

4.2.4. Medidas de rendimiento

Con objeto de obtener medidas del *throughput* en distintos escenarios con el modelo expuesto, en primer lugar es necesario calcular las probabilidades de encontrar ráfagas con reservas de time-slots para el tráfico de datos.

El primer escenario corresponde con disponer de todos los recursos para compartir tanto para voz como para datos ($n = 8$), es decir, sin reserva de recursos que utilizando la notación previa: $n - m = 0$. Los siguientes escenarios corresponden con la reserva de 1, 2, 3 ó 4 time-slots reservados para el tráfico de voz respectivamente. Para cada uno de los casos se determina la longitud media de ráfagas con una probabilidad de bloqueo del 1% y del 5%.

En cada uno de los casos es posible calcular el máximo *throughput* cuando un terminal móvil dispone de un número de time-slots para asignar a sus canales de datos. En las figuras Fig. 4.4 y Fig. 4.5 se ilustran algunos resultados en los que se trata de encontrar las ráfagas de una longitud dada con $n = 8$ time-slots.

La Figura 4.4 muestra los cinco escenarios señalados anteriormente en una única gráfica con una probabilidad de bloqueo del 1%, así como en la Figura 4.5 la probabilidad de bloqueo es del 5%.

Si se definen todas las ráfagas con una longitud menor o igual a 3 las ráfagas cortas y todas las ráfagas de longitud mayor o igual a 4 como largas, la diferencia más significativa entre los primeros escenarios en la Figura 4.4 y la Figura 4.5 es que la probabilidad de encontrar ráfagas largas es mayor con una probabilidad de bloqueo menor, del 1%, frente a una probabilidad

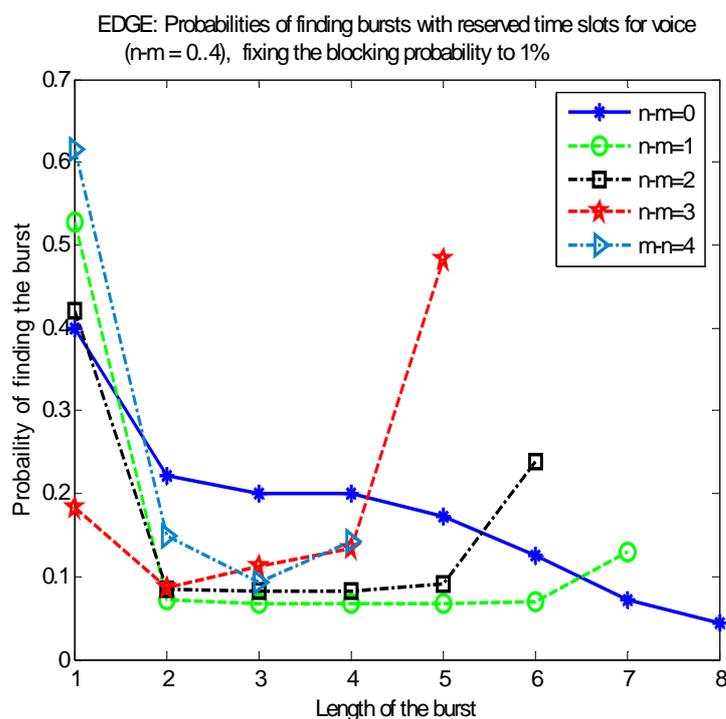


Figura 4.4: Escenarios con $n-m$ desde 0 hasta 4 time-slots reservados y probabilidad de bloqueo del 1%

de bloqueo mayor, del 5%. En consecuencia, las tasas de pico mayores se obtienen con una menor probabilidad de bloqueo, es decir, cuando la capacidad del sistema para tráfico de voz y tráfico de datos es alta, también existe una alta probabilidad de transmitir datos con una tasa de pico mayor.

En las figuras 4.4 y 4.5 también se puede observar que las tasas de pico para las ráfagas largas son mayores con una probabilidad de bloqueo menor que para el caso de probabilidad de bloqueo del 5%. Esto implica un considerable incremento de la transmisión de datos del usuario y es de destacar que con menor probabilidad de bloqueo incrementan las tasa de datos de los usuarios.

Si se examinan los escenarios con reserva de recursos para el tráfico de voz en la figura 4.4 ($n - m > 0$), en todos los casos existe una característica común: la ráfaga mayor se concentra en la ráfaga de longitud 1 y la próxima ráfaga más larga corresponde al valor de ráfaga mayor posible, es decir, de

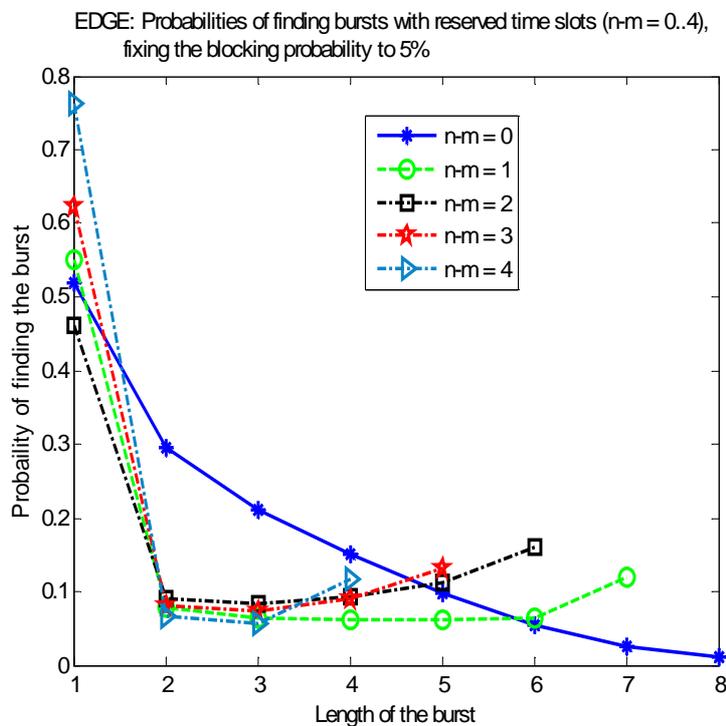


Figura 4.5: Escenarios con $n-m$ desde 0 hasta 4 time-slots reservados y probabilidad de bloqueo del 5%

valor m . No obstante, las probabilidades de encontrar time-slots consecutivos libres para los casos intermedios ($2 \leq x \leq m - 1$) son similares y menores que aquellos cuyas longitudes de ráfagas oscilan entre 1 y m .

Por otro lado, incidiendo en los casos con time-slots reservados para voz según la Figura 4.4, se puede apreciar que si el número de recursos reservados para el tráfico de voz se incrementa, la probabilidad de encontrar ráfagas con una longitud igual a 1 también se incrementa. Esto sucede fundamentalmente porque el número de time-slots compartidos para datos y voz disminuye, así se reduce el número de posibles time-slots libres consecutivos. De hecho, estas probabilidades se concentran en la ráfaga de longitud 1, con la excepción del caso de $n - m = 3$ time-slots reservados, en los que la probabilidad más alta corresponde a la ráfaga de longitud 5.

En la Figura 4.5 también se muestran los cuatro escenarios con la reserva

de recursos pero con la probabilidad de bloqueo del 5%. Se puede observar que las probabilidades de encontrar ráfagas con longitud de 1 son mayores que el caso anterior donde la probabilidad de bloqueo era del 1%, independientemente del número de time-slots reservados. Por el contrario, las probabilidades de encontrar ráfagas con longitud la mayor posible, m , son ligeramente menores que en el caso de probabilidad de bloqueo del 1%. En cuanto a los casos intermedios en los que las longitudes de las ráfagas cuyo rango es $n - m = 2..m - 1$, las probabilidades de encontrar time-slots libres consecutivos de una longitud dada son similares tanto para probabilidad de bloqueo del 1% como del 5%.

4.2.5. Esquema de asignación de recursos compartida con reserva para tráfico de datos

Históricamente el operador de red ha requerido de canales para el tráfico de voz puesto que la prioridad era mantener las comunicaciones de voz frente al transporte de datos. Sin embargo, cada vez más se hace necesario garantizar el transporte de datos y, por tanto, disponer de recursos radio donde localizar los canales de datos.

De forma análoga al esquema presentado anteriormente para la reserva de recursos para el tráfico de voz, el esquema de asignación de recursos compartidos junto con la reserva de recursos para el tráfico de datos consiste, básicamente, en dividir los recursos en los time-slots compartidos para ambos tipos de tráfico, voz y datos, y los time-slots reservados para el tráfico de datos (Figura 4.6).

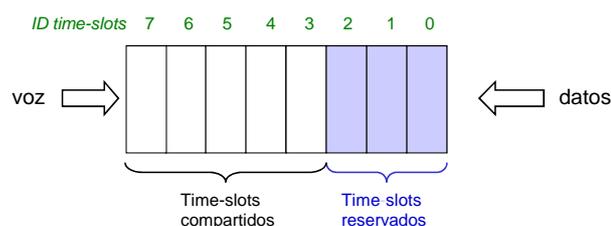


Figura 4.6: Asignación de time-slots con reserva para tráfico de datos

De acuerdo a esta demanda de canales de datos para el transporte de la información por parte de los terminales móviles, el objetivo de esta sección es la de comparar ciertos escenarios y analizar la garantía de QoS para el tráfico

de datos. En concreto en las Figuras 4.7 y 4.8 se muestra la distribución de los time-slots con uno y dos recursos reservados para los paquetes de datos y el resto de recursos compartidos para el tráfico de voz y el tráfico de datos.

Haciendo uso de pocos recursos reservados para el tráfico de datos, 1 y 2 time-slots en este caso, se pueden obtener grandes capacidades para los usuarios. Esto implica que disminuyendo el tráfico de voz cursado, se puede incrementar las tasas de pico de los usuarios/aplicaciones de datos. A partir del modelo original de la asignación de canales sin reserva de recursos, podemos contrastarla con la reserva de 1 y 2 time-slots para el tráfico de datos.

Para el supuesto en el que se reserve un solo time-slot para el tráfico de datos, en la Tabla 4.1 se puede apreciar los valores calculados una vez se fija la probabilidad de bloqueo al 1%. En la tabla 4.2 los valores mostrados corresponden a una probabilidad de bloqueo del 2%. En ambos casos el hecho destacable es que, para la mayoría de los valores de longitud de ráfagas, la probabilidad de encontrar un número n de time-slots libres consecutivos es mayor cuando se hace la reserva del time-slot que sin reserva de recursos para el tráfico de datos. Esto implica que se obtienen mayores tasas de pico para los usuarios y más oportunidades de asignación de los recursos disponiendo de mayores ráfagas. El operador de red tendrá mayores posibilidades de negociar los parámetros de QoS con el usuario y éste aprovechar la capacidad multislot de los terminales móviles.

En las Tablas 4.3 y 4.4 se muestra la comparativa entre los valores de las probabilidades de encontrar ráfagas de longitud n sin reserva de recursos frente a utilizar la reserva de 2 time-slots para el tráfico de datos.

En el primer caso (Tabla 4.3) se fija la probabilidad de bloqueo al 1% y en el segundo caso (Tabla 4.4) la probabilidad de bloqueo es del 2%. Los resultados si bien son similares a los anteriormente expuestos con un solo

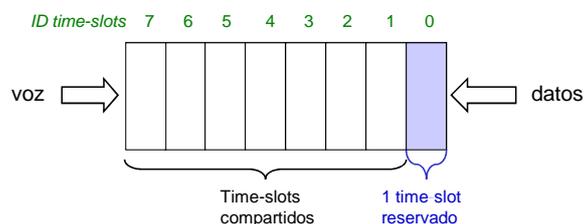


Figura 4.7: 8 time slots con 1 time-slot reservado y 7 time-slots compartidos

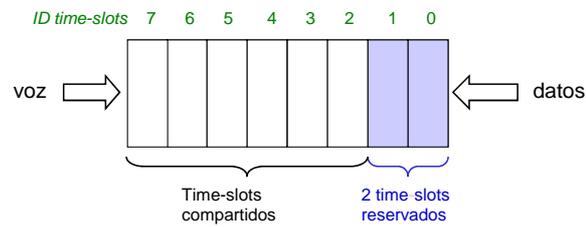


Figura 4.8: 8 time slots con 2 time-slots reservados y 6 time-slots compartidos

Longitud n de la ráfaga	$P(\text{ráfaga } n)$ con 8 time slots	$P(\text{ráfaga } n)$ 7 + 1 time-slot reservado	Incremento (%) 7+1 versus 8 time slots
1	0.0611	0,06282	+ 0.17
2	0.1303	0,14056	+ 1.02
3	0.1853	0,19772	+ 1.24
4	0.2006	0,21951	+ 1.89
5	0.1728	0,17885	+ 0.60
6	0.1242	0,11727	- 0.69
7	0.0716	0,00083	- 7.07
8	0.0439	0,08243	+ 3.85

Cuadro 4.1: Comparativa de probabilidades de encontrar ráfagas sin reserva y con reserva de 1 time-slot con Probabilidad de Bloqueo del 1 %

time-slot reservado para datos - en cuanto a encontrar mayor probabilidad de ráfagas comprendidas entre 2, 3, 4 y 5 time-slots libres consecutivos - no obstante el tanto por ciento de incremento de ráfagas de estas longitudes se incrementa en más del 4 % en unos casos cuando se reservan 2 time-slots frente a reservar solo un time-slot. En particular, en el caso de encontrar ráfagas de longitud 8 se incrementa hasta más de un 6 % en el caso de fijar la probabilidad de bloqueo al 1 % si comparamos las Tablas 4.1 y 4.3

Longitud n de la ráfaga	P(ráfaga n) con 8 time slots	P(ráfaga n) 7 + 1 time-slot reservado	Incremento (%) 7+1 versus 8 time slots
1	0.1013	0.1076	+ 0.6280
2	0.1739	0.1876	+ 1.3651
3	0.2043	0.2161	+ 1.1793
4	0.1891	0.2027	+ 1.3577
5	0.1437	0.1457	+ 0.2036
6	0.0924	0.0858	- 0.6569
7	0.0485	0.0011	- 4.7402
8	0.0268	0.0534	+ 2.6635

Cuadro 4.2: Comparativa de probabilidades de encontrar ráfagas sin reserva y con reserva de 1 time-slot con Probabilidad de Bloqueo del 2%

Longitud n de la ráfaga	P(ráfaga n) con 8 time slots	P(ráfaga n) 6 + 2 time-slots reservados	Incremento (%) 6+2 versus 8 time-slots
1	0.0611	~ 0	- 6.1100
2	0.1303	0.1825	+ 5.2242
3	0.1853	0.2180	+ 3.2676
4	0.2006	0.2580	+ 5.7362
5	0.1728	0.1865	+ 1.3678
6	0.1242	0.0049	- 11.9270
7	0.0716	0.0015	- 7.0096
8	0.0439	0.1486	+ 10.4707

Cuadro 4.3: Comparativa de probabilidades de encontrar ráfagas sin reserva y con reserva de 2 time-slots con Probabilidad de Bloqueo del 1%

4.3. Conclusiones

En este capítulo se han presentado dos esquemas de asignación de recursos para el sistema EDGE haciendo uso del mecanismo de reserva de recursos

Longitud n de la ráfaga	P(ráfaga n) con 8 time slots	P(ráfaga n) 6 + 2 time-slots reservados	Incremento (%) 6+2 versus 8 time-slots
1	0.1013	~ 0	- 10.1300
2	0.1739	0.2605	+ 8.6598
3	0.2043	0.2417	+ 3.7373
4	0.1891	0.2355	+ 4.6370
5	0.1437	0.1510	+ 0.7307
6	0.0924	0.0060	- 8.6378
7	0.0485	0.0021	- 4.6379
8	0.0268	0.1032	+ 7.6409

Cuadro 4.4: Comparativa de probabilidades de encontrar ráfagas sin reserva y con reserva de 2 time-slots con Probabilidad de Bloqueo del 2%

radio, teniendo en cuenta que la ocupación de dichos recursos siempre sigue un orden establecido secuencial, tanto para el tráfico de voz como para el tráfico de datos.

El primer esquema corresponde a la asignación de los recursos haciendo uso, por un lado, de time-slots compartidos para el tráfico de voz y el tráfico de datos y, por otro lado, de la reserva de time-slots para el tráfico de voz. A partir de aquí, se obtienen las expresiones analíticas del modelo que permiten obtener los resultados de longitudes de ráfagas con probabilidades de bloqueo del 1% y 2%. De forma genérica se puede concluir que si la probabilidad de bloqueo aumenta y el número de recursos compartidos para datos y voz disminuye, el número potencial total de recursos libres consecutivos también decrece. Esto demuestra que un incremento en la probabilidad de tener todos los time-slots ocupados reduce la probabilidad de encontrar recursos libres consecutivos, implica tasas de pico menores. En otras palabras, las tasas de pico serán mayores cuanto menor sea la probabilidad de bloqueo del sistema.

En el esquema de asignación de recursos utilizando la reserva de time-slots para el tráfico de datos, de acuerdo a los escenarios presentados (sin reserva de recursos, uno o dos time-slots reservados) se puede resumir en cuatro conclusiones. La primera conclusión es que se incrementa las tasas de pico del tráfico de datos de los usuarios usando un valor pequeño de reserva de

recursos. En segundo lugar, concluimos que independientemente de la probabilidad de bloqueo (1% ó 2%) igualmente se incrementa la probabilidad de encontrar ráfagas para longitudes comprendidas entre 2 y 5 time-slots libres consecutivos. En tercer lugar, con este esquema se garantizan necesidades de los usuarios de datos incluso cuando el tráfico de voz sea alto. Por último, se puede garantizar recursos para los trasposos de tráfico de datos procedentes de otras celdas del sistema móvil.

Como conclusión general, los esquemas presentados permiten optimizar al operador de red la asignación de los recursos de acuerdo al perfil de QoS que requieran los distintos usuarios de la red. Esto significa que, no solo en tiempo es posible ofrecer la QoS proporcionando acceso prioritario a ciertos recursos sino que además se puede proporcionar distintas longitudes de ráfagas a los usuarios. Así se puede asignar ráfagas de longitud mayor a aquellos usuarios o aplicaciones con requisitos de bajo retardo y, por tanto, disponer de tasas de pico altas.

Capítulo 5

Asignación de recursos con reserva y buffering para EDGE

5.1. Introducción

En este capítulo se introduce un nuevo mecanismo al método de asignación de los recursos presentado en el capítulo anterior. En concreto, además de utilizar la reserva de recursos para un determinado tipo de tráfico, se añade como herramienta un *buffer* de longitud limitada con objeto de disminuir la probabilidad de pérdida del tráfico de datos. Así, los paquetes de datos que lleguen al sistema y en el caso en el que todos los recursos que admitan este tipo de tráfico, es decir, aquellos compartidos tanto para el tráfico de voz como el de datos, si están ocupados podrán esperar en la memoria temporal, buffer, hasta la liberación de uno o varios recursos para ser transmitidos.

A partir de este punto se introduce un esquema de la asignación compartida de los recursos usando como técnicas de mejora de las prestaciones del sistema, la reserva de recursos radio para el tráfico de voz y el mecanismo de buffering para el tráfico de datos. A continuación presentamos tanto el modelo analítico de dicho esquema así como la resolución del modelo, los resultados y las correspondientes conclusiones.

5.2. Esquema de asignación de recursos con reserva y buffering

El esquema de asignación compartida de recursos haciendo uso, por un lado, de la reserva de recursos radio para el tráfico de voz y, por otro lado, del mecanismo de buffering para el tráfico de datos, se basa en el esquema presentado en el capítulo anterior. Además se añade una memoria temporal, cola, con política de servicio FIFO que, en función del tipo de tráfico entrante (Figura 5.1), se pueden producir los siguientes casos:

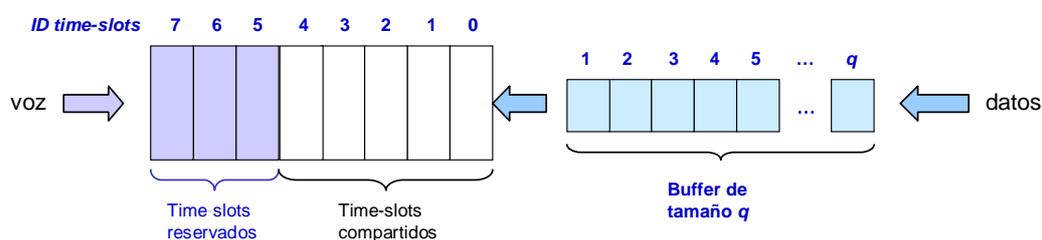


Figura 5.1: Asignación de time-slots con reserva y buffering

1. Si las llamadas entrantes son de voz, ocuparán los time-slots libres con el identificador cuyo valor sea el mayor posible; esto implica que primero se ocuparán los time-slots reservados que estén disponibles y posteriormente los time-slots compartidos libres hasta ese momento;
2. En el caso en el que se produzca la liberación de una llamada de voz de un time-slot reservado, este recurso será ocupado por la llamada, si la hubiese, que esté ocupando un time-slot de los denominados compartidos con el número menor de identificador (*intracell - handover*).
3. Para el tráfico de datos, los time-slots se ocuparán de forma ordenada siguiendo un sentido de ocupación contrario al del tráfico de voz y ocuparán los time-slots con el número de identificador menor situándose en los recursos compartidos.
4. En caso de que todos los time-slots compartidos estén ocupados cuando intente acceder un bloque de datos, estos serán almacenados temporalmente en un buffer hasta encontrar recursos disponibles a tal efecto en los time-slots disponibles.

Utilizando este esquema se pretende obtener una probabilidad de pérdida de datos menor al esquema presentado con anterioridad gracias a la utilización de la cola de datos de dimensión q .

5.2.1. Modelo analítico

En esta sección se describe el modelo analítico correspondiente al esquema de asignación compartida haciendo uso de la reserva de recursos para el tráfico de voz y la cola para el tráfico de datos. De esta forma las ecuaciones de transición de estados para este modelo se pueden expresar independientemente del número de time-slots que componga cada una de las tramas, además de manera genérica, sin determinar el número concreto de time-slots reservados para el tráfico de voz, el tamaño del buffer ó el número de time-slots compartidos por el tráfico de datos y el tráfico de voz.

De forma análoga a como se enumeró en el capítulo anterior los estados posibles, un estado estará definido por una secuencia $[r_{n-1}, \dots, r_1, r_0, q]$ con $r_i = \{0, 1, 2\}$ y donde q representa el tamaño del buffer. es decir, el número de elementos de datos en cola. Por otro lado, m es el número de time-slots compartidos para voz y datos, y $n - m$ es el número de time-slots reservados para el tráfico de voz. En el estado definido por $[r_{n-1}, \dots, r_m, r_{m-1} \dots, r_1, r_0, q]$, los time-slots con números de identificadores desde el valor 0 hasta el valor $m - 1$ corresponden a los time-slots compartidos y los recursos con número de identificador que van desde el valor m hasta el $n - 1$ corresponden con los time-slots reservados para voz. El número de estados posibles con este esquema es:

$$Num_{estados} = 2^n + 3^m - 2^m + q \cdot (2^{n-m} + 2^m - 1).$$

Una forma de verificar esta expresión es contar todos los posibles estados en los que ninguna llamada de voz ocupa los time-slots compartidos, es decir, 2^n . Esto se debe a que un time-slot con un número de identificador cuyo rango va desde 0 hasta $m - 1$ toma solo dos posibles valores 0 ó 1 y cualquier time-slot cuyo número de identificador esté comprendido entre m y $n - 1$ será bien 0 ó 2. Luego para contar todos los posibles estados en los que los time-slots reservados están ocupados y existe tráfico de voz en los time-slots compartidos, implica que tendremos un total de 3^m estados posibles en esta situación. Los time-slots con números de identificador entre m y $n - 1$ tendrán todos ellos el valor igual a 2 y los time-slots con números de identificador entre 0 y $m - 1$ han de ser bien 0 ó bien 1 ó bien 2, es decir, time-slot desocupado, ocupado con tráfico de datos y ocupado con

tráfico de voz, respectivamente. Esto significa que cuando todos los time-slots reservados están ocupados y todos los time-slots compartidos toman los valores de 0 ó 1, contamos doblemente 2^m estados. Todo ello teniendo en cuenta que el buffer no contiene elementos de datos en cola. Para el caso en el que todos los time-slots compartidos estén ocupados bien con voz o con datos (2^m), sabiendo que si están ocupando estos recursos con tráfico de voz obligatoriamente todos los time-slots reservados también han de estar ocupados (2^{n-m}), se ha de añadir el producto de la capacidad del buffer por esta combinación de estados en los que los time-slots compartidos estén totalmente ocupados.

Enumeramos los estados con números decimales, con un ejemplo muy sencillo se puede ilustrar los distintos estados para este esquema. El número de time-slots es $n = 4$ y $m = 2$ time-slots compartidos para voz y datos. Si $q = 1$, los estados se presentan en función de los siguientes rangos. Primero, denotamos los estados en los que no están ocupados totalmente los time-slots reservados y los representamos como números en base 2 (los primeros $2^n - 2^m$ estados, 12 para este ejemplo). Primero consideramos que una posición toma un valor 1 si el time-slot con el número de identificador correspondiente está ocupado, bien sea de tráfico de voz o bien sea tráfico de datos, y 0 en caso contrario. El valor 2 en los time-slots reservados representa la ocupación por el tráfico de voz. El valor separado por una coma indica cuantas posiciones hay ocupadas en el buffer, tal como se muestra a continuación:

Número estado	Vector		
	$\overbrace{\quad}^{n-m}$	$\overbrace{\quad}^m$	$\overbrace{\quad}^q$
0	[00	00	, 0]
1	[0001,	0]	
2	[0010,	0]	
3	[0011,	0]	
4	[0200,	0]	
5	[0201,	0]	
6	[0210,	0]	
7	[0211,	0]	
8	[2000,	0]	
9	[2001,	0]	
10	[2010,	0]	
11	[2011,	0]	

Ahora, denotamos los estados en los que siempre están ocupados totalmente los time-slots reservados y los representamos como números en base 3, esto es, el rango de valores desde el estado $2^n - 2^m$ hasta el estado $3^m + 2^n - 2^m - 1$, (del estado 12 hasta el 20 para este ejemplo).

Número estado	Vector
12	$\left[\overbrace{22}^{n-m} \overbrace{00}^m, \overbrace{0}^q \right]$
13	[2201, 0]
14	[2202, 0]
15	[2210, 0]
16	[2211, 0]
17	[2212, 0]
18	[2220, 0]
19	[2221, 0]
20	[2222, 0]

A continuación enumeramos los estados que incluyen datos en el buffer, es decir, las posiciones de q ocupadas que para este caso es $q = 1$. Esto significa los estados que comprendidos entre el estado $3^m + 2^n - 2^m$ hasta el estado $3^m + 2^n - 2^m + q \cdot (2^{n-m} + 2^m - 1) - 1$, es decir, del estado 21 hasta el 27 para este ejemplo. En este rango se pueden distinguir dos bloques: el primero correspondiente a aquellos estados en los que todos los time-slots compartidos están ocupados por datos y al menos uno de los time-slots reservados está libre; el segundo bloque de estados corresponde a aquellos en los que todos los time-slots compartidos pueden estar ocupados tanto por voz como por datos.

Número estado	Vector
21	$\left[\overbrace{00}^{n-m} \overbrace{11}^m, \overbrace{1}^q \right]$
22	[0211, 1]
23	[2011, 1]
24	[2211, 1]
25	[2212, 1]
26	[2221, 1]
27	[2222, 1]

5.2.2. Resolución del modelo

El sistema de ecuaciones correspondiente al esquema expuesto de tráfico de voz y datos con reserva y buffering, implica la inclusión de todos los estados en los que interviene la cola de datos, por tanto una ampliación de las ecuaciones presentadas en el modelo con reserva de recursos.

Un recorrido del modelo desde lo particular hacia lo general nos lleva a observar por medio de ejemplos las ecuaciones que definen el sistema. A modo de ilustración se puede observar en la Figura 5.2 el grado de complejidad que se puede formar únicamente con los valores de tamaño de trama $n = 3$ time-slots, de los cuales $m = 1$ es el time-slot compartido y $q = 1$ el tamaño del buffer.

En la Figura 5.2 se pueden destacar los casos en los que, en primer lugar, la ocupación de los time-slots reservados para el tráfico de voz se ocupan de forma secuencial y, por tanto, para los estados $[000, 0]$, $[00D, 0] \sim [001, 0]$, $[00D, 1] \sim [001, 1]$ cuando se produzca una llamada entrante de voz la transición sólo podrá ser a los estados $[V00, 0] \sim [200, 0]$, $[V0D, 0] \sim [201, 0]$, $[V0D, 1] \sim [201, 1]$ respectivamente. En segundo lugar, la transición desde el estado $[VVV, 0] \sim [222, 0]$ en el caso en el que cualquier llamada de voz termine, sólo podrá transitar al estado $[VV0, 0] \sim [220, 0]$; esto se debe a que si el recurso que se libera es el primero o el segundo time-slot reservado, implicará de forma automática un traspaso dentro de la misma celda (*intracell - handover*) de la llamada que esté ocupando, en este caso, el time-slot compartido y pasará a ocupar el time-slot reservado, primero o segundo según el caso. En tercer lugar y de forma similar al caso anterior, la liberación de cualquier llamada de voz ocupando un time-slot reservado estando en el estado $[VVV, 1] \sim [222, 1]$ implicará que los datos que están ocupando la posición de la cola pasarán siempre a ocupar, en este caso, el time-slot compartido y no otro puesto que nuevamente el time-slot liberado por la terminación de la llamada implicará un *intracell - handover* siempre que sea el caso de cualquier time-slot reservado el que se libere, por tanto, la llamada que ocupaba el time-slot compartido pasará a ocupar la posición de reserva, es decir, transitará al estado $[VVD, 0] \sim [221, 0]$.

Siguiendo con el ejemplo en el que $n = 4$ time-slots de los cuales $m = 2$ time-slots son los compartidos y $q = 1$ es el tamaño del buffer, la representación de alguno de los estados del sistema y las ecuaciones de transición correspondientes las podemos observar a continuación. Por ejemplo, para el estado 20, $[VVVV, 0] \sim [2222, 0]$, es decir donde tanto los time-slots compar-

tados previos y hacia los posteriores estados en los que puede estar el sistema. Es de destacar que, de forma similar a lo comentado anteriormente, la transición del estado 20, $[VVVV, 0] \sim [2222, 0]$ al estado 18, $[VVV0, 0] \sim [2220, 0]$, estará en función de: bien la terminación de una llamada en los time-slots reservados, por tanto, se producirá una operación *intracell-handover*, o bien el último (número de identificador menor) time-slot de los recursos compartidos libere este recurso porque acaba la llamada de voz. En el caso en que la liberación del recurso con la llamada de voz esté ocupando el time-slot más significativo de los recursos compartidos, la transición desde el estado 20 sería al estado 14, $[VV0V, 0] \sim [2202, 0]$, tal como se muestra en la Figura 5.3.

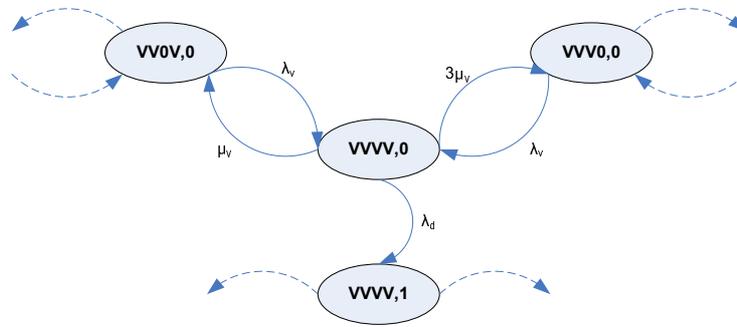


Figura 5.3: Diagrama de transición para el estado 20 ($n=4, m=2, q=1$)

La ecuación que define este diagrama de transición es:

$$[4\mu_v + \lambda_d] \cdot P(20) = \lambda_v \cdot [(P(18) + P(14))],$$

En el caso donde las transiciones parten y llegan al estado 24, $[VVDD, 1] \sim [2211, 1]$, se puede observar en la Figura 5.4 como puede transitar a estados en los que el buffer sigue en la misma situación para el caso en el que las llamadas de voz puedan terminar, o bien la situación en la que se libera uno de los dos time-slots con tráfico de datos y nuevos datos entran procedentes de la cola a ocupar el time-slot desocupado, es decir se pasaría al estado 16, $[VVDD, 0] \sim [2211, 0]$.

La ecuación que define esta situación es:

$$[2\mu_v + 2\mu_d] \cdot P(24) =$$

$$\lambda_d \cdot (P(16) + \lambda_v \cdot [(P(22) + P(23))],$$

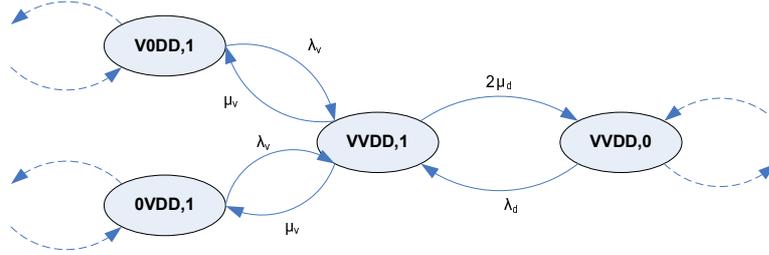


Figura 5.4: Diagrama de transición para el estado 24 ($n=4, m=2, q=1$)

Las notaciones generales en las que un estado r , tendrá $i_{r,v}$ y $i_{r,d}$ que representan el número total de time-slots ocupados con el tráfico de voz y el tráfico de datos respectivamente. De forma análoga, sea $k_{r,v}$ y $k_{r,d}$ el número de time-slots consecutivos ocupados de la trama en sentido izquierda-derecha para tráfico de voz y tráfico de datos respectivamente.

Atendiendo a lo expuesto y de forma genérica, se puede modelar este sistema con las ecuaciones de transición que se resumen en varios casos. En primer lugar, para todos los estados r en los que no todos los recursos reservados para voz están ocupados (los primeros $2^n - 2^m$ estados), el subsistema de ecuaciones de transición corresponde a la siguiente expresión:

$$[\lambda_v + b(i_{r,d} < m) \cdot \lambda_d + i_{r,v} \cdot \mu_v + i_{r,d} \cdot \mu_d]P(r) =$$

$$= b(i_{r,v} < n - m - 1) \cdot \mu_v \cdot \sum_{\substack{j=k_{r,v} \\ j:r_{n-1-j}=0}}^{n-m-1} P(2^{n-1-j} + r) +$$

$$+ b(i_{r,v} = n - m - 1) \cdot \mu_v \cdot P(2^m(2^{n-m} - 1) + K_1) +$$

$$+ b(i_{r,d} = m) \cdot \mu_d \cdot m \cdot P(1 + 3^m + 2^n - 2^m + K_3) +$$

$$+ b(i_{r,d} < m - 1) \cdot [\mu_d \cdot \sum_{j=k_{r,d}; j:r_j=0}^{m-1} P(2^j + r)] +$$

$$\begin{aligned}
& +b(k_{r,v} > 0) \cdot \lambda_v \cdot \sum_{j=0}^{k_{r,v}-1} P(r - 2^{n-1-j}) + \\
& +b(k_{r,d} > 0) \cdot \lambda_d \cdot \sum_{j=0}^{k_{r,d}-1} P(r - 2^j),
\end{aligned}$$

donde:

$$\begin{aligned}
K_1 &= \sum_{j=0}^{m-1} r_j \cdot 3^j; \\
K_3 &= \sum_{\substack{j=m \\ j:r_j \neq 0}}^{n-1} 2^{n-j}.
\end{aligned}$$

Los términos K_1 y K_3 representan cambios de base de los estados en función de la transición que se haya producido. En concreto, K_1 indica aquellos estados en los que existe un time-slot reservado que está libre y pasa a estados donde todos los time-slots reservados están ocupados. En el caso del término K_3 , éste indica el cambio de base de aquellos estados donde todos los time-slots compartidos están ocupados por datos y la cola está vacía y pasa a estados en los que todos los time-slots compartidos se ocupan por datos y en la cola de datos hay posiciones ocupadas.

La siguiente expresión representa las ecuaciones que comprenden desde el estado $2^n - 2^m$ hasta el estado $3^n + 2^n - 2^m - 1$:

$$\begin{aligned}
& [b(i_{r,d} + i_{r,v} < n - 1) \cdot \lambda_v + b(i_{r,d} < m - 1) \cdot \lambda_d + i_{r,v} \cdot \mu_v + i_{r,d} \cdot \mu_d]P(r) = \\
& = b(i_{r,d} + i_{r,v} < n - 1) \cdot [\mu_d \cdot \sum_{\substack{j=k_{r,d} \\ j:r_j=0}}^{n-k_{r,v}-1} P(r + 3^{m-j}) + \\
& + \sum_{\substack{j=k_{r,d} \\ j:r_j=0}}^{n-k_{r,v}-1} P(r + 2 \cdot 3^{m-j}) \cdot \left\{ \begin{array}{ll} \mu_v & \text{si } j \geq L_r + \\ (n - m)\mu_v & \text{si } j < L_r + \end{array} \right\} \\
& + b(k_{r,d} > 0) \cdot \lambda_d \cdot \sum_{\substack{j=k_{r,d} \\ j:r_j=0}}^{k_{r,d}-1} P(r - 3^{m-j}) +
\end{aligned}$$

$$\begin{aligned}
& +b[(k_{r,v} > n - m - 1) \&(L_r \leq m - 1)] \cdot \lambda_v \cdot \sum_{\substack{j=n-k_{r,v} \\ j:r_j=0}}^{m-1} P(r - 2 \cdot 3^{m-j}) + \\
& +b(L_r = m - 1) \cdot \lambda_v \cdot \sum_{j=m}^{n-1} P(K_2 - 2^{n+m-j}) + \\
& +b(i_{r,d} + i_{r,v} = n - 1) \cdot [\mu_d \cdot i_{r,d} \cdot P(K_4) + \mu_v \cdot \sum_{j=0}^{m-1} P(K_4 + 2^{m-j})],
\end{aligned}$$

donde:

$$\begin{aligned}
K_2 &= 1 + \sum_{j=0}^{m-1} r_j \cdot 2^m + \sum_{j=m}^{n-1} 2^{n+m-j}; \\
K_4 &= \sum_{j=0}^{m-1} [(r_j - 1) \cdot 2^{m-j} + 3^m + 2^n - 2^m + 2^{n-m}].
\end{aligned}$$

También los términos K_2 y K_4 representan cambios de base en los estados dependiendo de la transición de estos. El término K_2 indica un cambio de base de un estado en el que, por un lado, todos los time-slots reservados están ocupados y, por otro lado, ningún time-slot de los compartidos contiene tráfico de voz y pasa a un estado en el que hay un time-slot reservado libre. En el caso del término K_4 indica el cambio de base de aquellos estados donde todos los time-slots compartidos están, por un lado, ocupados por voz o datos y, por otro lado, la cola está vacía y pasa a estados en los que todos los time-slots compartidos se ocupan por voz o datos y en la cola de datos hay una posición ocupada.

El término L_r representa el número menor de identificador del time-slot que contiene una llamada de voz de los time-slots denominados compartidos. Este valor de L_r permite conocer cuál ha de ser la llamada que se ha de traspasar (*intracell - handover*) en caso de que cualquier time-slot reservado se haya liberado.

El conjunto de ecuaciones que comprende el rango de estados que va desde el estado $3^n + 2^n - 2^m$ hasta el estado $3^n + 2^n - 2^m + q \cdot (2^{n-m} + 2^m - 1)$, se concreta en la siguiente expresión:

$$[b(B < q) \cdot \lambda_d + b(i_{r,v} < n - m - 1) \cdot \lambda_v + i_{r,v} \cdot \mu_v + i_{r,d} \cdot \mu_d]P(r) =$$

$$\begin{aligned}
&= b(B > 1) \cdot \lambda_d P(r - (2^{n-m} + 2^m - 1)) + \\
&+ b(B = 1) \cdot \lambda_d \left\{ \begin{array}{ll} P(2^m + \sum_{\substack{i=m \\ r_i \neq 0}}^{n-1} 2^{n+m-i}) & si : i_{r,v} < n - m - 1 \\ P(1 + 2^n - 2^m + \sum_{i=0}^{m-1} r_i \cdot 3^{m-i}) & si : i_{r,v} \geq n - m - 1 \end{array} \right\} + \\
&+ b[(k_{r,v} > 0) \& (L_r > m - 1)] \cdot \lambda_v \cdot \sum_{\substack{j=n-k_{r,v} \\ j:r_j=0}}^{n-1} P(r + 2^{n-j}) + \\
&+ b(B < q) \cdot [i_{r,d} \cdot \mu_d P(r + (2^{n-m} + 2^m - 1)) + \\
&b(i_{r,v} \geq n - m - 1) \cdot (n - m + 1) \cdot \mu_v \cdot \sum_{\substack{j=0 \\ j:r_j=0}}^{m-1} P(r + 2^{m-j} + 2^{n-m} + 2^m - 1)] + \\
&+ b(i_{r,v} < n - m - 1) \cdot \mu_v \cdot \sum_{\substack{j=m-1 \\ j:r_j=0}}^{n-1} P(r + 2^{n-1})].
\end{aligned}$$

Los términos q y B representan valores de la cola de datos, en concreto, q indica el tamaño de la cola, *buffer*, y el término B representa el número de posiciones ocupadas en el buffer. En caso de que todos los time-slots compartidos estén ocupados por datos, puede ocurrir que se libere alguno de estos time-slots y se ocupe por datos procedentes del buffer. En la situación en la que todos los recursos reservados estén ocupados y en los time-slots compartidos al menos hay una llamada de voz ocupando uno de estos recursos, estos estados pueden provenir de estados en los cuales el time-slot compartido por la llamada de voz estaba disponible anteriormente y existían time-slots ocupados con números de identificador de time-slots menores.

5.2.3. Medidas de rendimiento

Con objeto de evaluar el modelo expuesto y analizar el grado de servicio del mismo, es necesario resolver el conjunto de ecuaciones que definen el sistema. Una vez resuelto el sistema se determinan los parámetros que pueden repercutir en la eficiencia del sistema. Teniendo en cuenta que el elemento incorporado es el buffer, en consecuencia la probabilidad de pérdida, la probabilidad de bloqueo y la probabilidad de demora son los factores clave que deben evaluarse.

Así, determinamos la probabilidad de pérdida del sistema - probabilidad de que bien un bloque de datos o bien una llamada llegue al sistema y se pierda - cuando se fija el tráfico ofrecido de datos y se varía el tráfico ofrecido de voz para una trama de $n = 8$ time-slots y con un número recursos reservados entre 1 y 2 time-slots.

En primer lugar, se evalúan las probabilidades de pérdida y de bloqueo del sistema con un cuando el tráfico de voz es superior al tráfico ofrecido de voz ($TO_v \gg TO_d$). En este caso se puede observar en la Figura 5.5 que, haciendo variar la capacidad del buffer, se obtienen valores alrededor del 2%, independientemente de la variación del tráfico de voz ofrecido para un tamaño del buffer de $q = 4$, hasta aproximadamente el 40% de probabilidad de pérdida para un tamaño del buffer de $q = 1$ para los valores más altos del tráfico ofrecido de voz. Esto significa que se produce una pérdida importante de las llamadas entrantes al sistema si el tamaño del buffer es menor a una capacidad de $q \leq 2$ elementos.

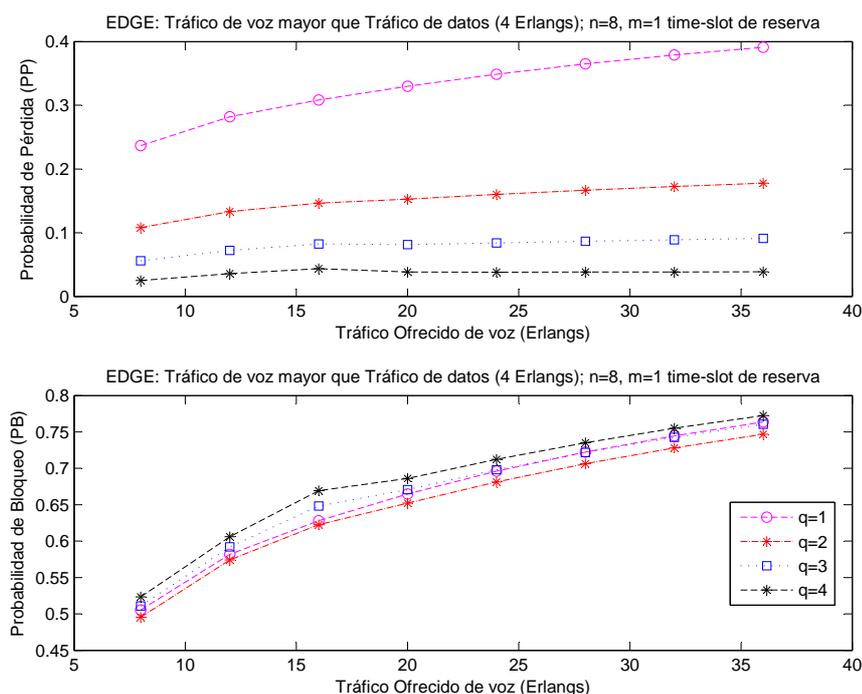


Figura 5.5: Probabilidad de Pérdida y Probabilidad de Bloqueo para un sistema en el que el tráfico de voz es mayor que el tráfico de datos, $n=8$, $m=1$

De la misma forma, en esta misma Figura 5.5 se puede observar que la probabilidad de bloqueo varía desde aproximadamente el 50 % hasta el 75 % con muy poca diferencia en función de la variación de la capacidad del buffer. Esto implica que el sistema en cualquier caso se mantiene bloqueado un porcentaje de tiempo muy alto, independientemente de la variación del tráfico ofrecido de voz y el tamaño de la cola. Es de destacar que la probabilidad de bloqueo, al contrario que sucede con la probabilidad de pérdida, es mayor cuanto mayor sea la capacidad de la cola puesto que durante más tiempo pueden estar sirviéndose las llamadas de voz o datos en los recursos sin que se produzcan pérdidas.

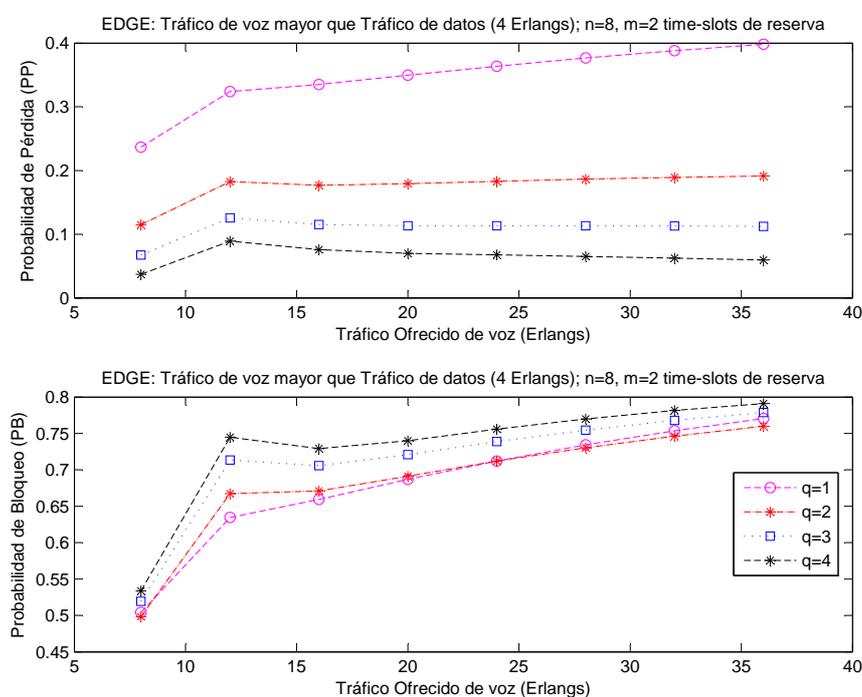


Figura 5.6: Probabilidad de Pérdida y Probabilidad de Bloqueo para un sistema en el que el tráfico de voz es mayor que el tráfico de datos, $n=8$, $m=2$

En la Figura 5.6 se muestra la incidencia que tiene el aumento del número de time-slots reservados $m = 2$. Para esta situación, manteniendo los mismos parámetros del caso anterior, observamos que la probabilidad de pérdida aumenta entre el 3 % y el 8 % con $q = 4$, sin embargo se produce un salto

considerable cuando el tráfico ofrecido de voz es de 12 Erlangs. Con este valor de tráfico de voz la probabilidad de bloqueo sube hasta un 63 % con $q = 1$ y hasta un 75 % con $q = 4$. Esto representa un aumento de la probabilidad de bloqueo del sistema de más de un 10 % con respecto a solo utilizar un time-slot de reserva. En cuanto a la probabilidad de pérdida los valores obtenidos son igualmente lineales con el aumento del tráfico ofrecido de voz y ligeramente superiores con respecto a los obtenidos para $m = 1$ time-slot reservado para el tráfico de voz.

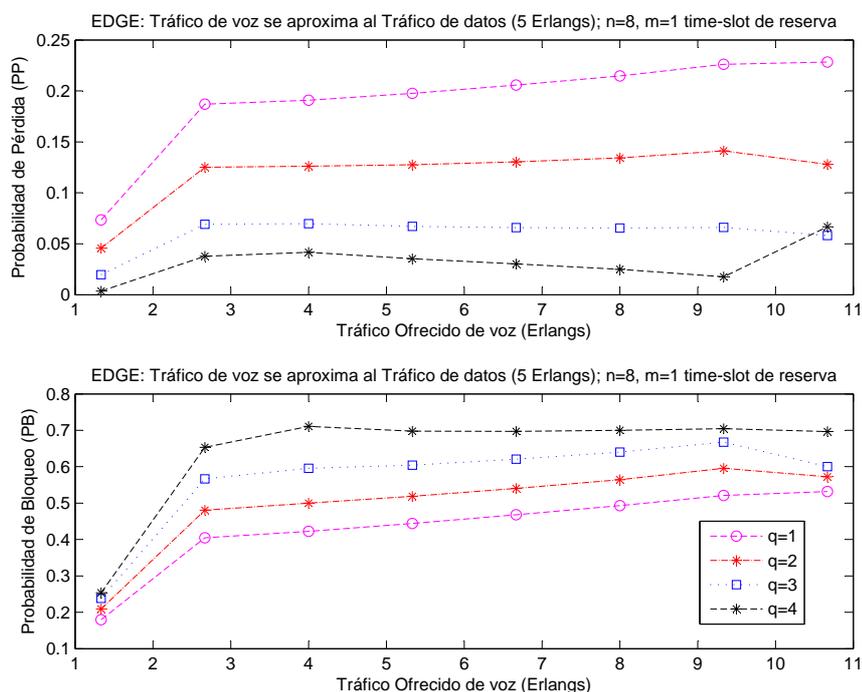


Figura 5.7: Probabilidad de Pérdida y Probabilidad de Bloqueo para un sistema en el que el tráfico de voz se aproxima al tráfico de datos, $n=8$, $m=1$

En segundo lugar, se evalúan las probabilidades de pérdida y de bloqueo del sistema cuando el tráfico de voz se aproxima al tráfico ofrecido de voz ($TO_v \approx TO_d$). Igualmente se puede observar en la Figura 5.7 que, haciendo variar la capacidad del buffer, se obtienen valores desde prácticamente el 0 % al 7 % de probabilidad de pérdida para un tamaño del buffer que va desde $q = 4$ hasta $q = 1$, para los valores menores del tráfico ofrecido de voz. Así como probabilidades de pérdida desde el 6 % hasta el 23 % para valores

más altos del tráfico de voz. De la misma forma, en este misma Figura 5.7 se puede observar que la probabilidad de bloqueo nuevamente se invierte con respecto a la probabilidad de pérdida en función del valor q , y varía desde el 18 %, valor mínimo obtenido con $q = 1$, hasta el 70 % para $q = 4$.

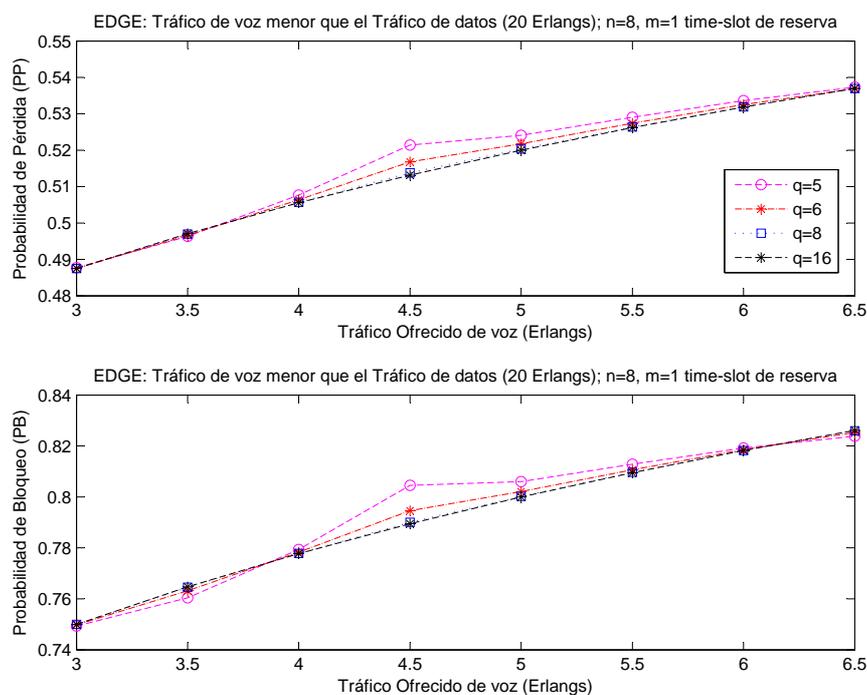


Figura 5.8: Probabilidad de Pérdida y Probabilidad de Bloqueo para un sistema en el que el tráfico de voz es menor que el tráfico de datos, $n=8$, $m=1$

En tercer lugar, si se trata la relación entre ambos tipos de tráfico de tal forma que el tráfico de voz sea bastante inferior al tráfico de datos ($TO_v \ll TO_d$), se puede observar en la Figura 5.9 que la probabilidad de pérdida, en general, es considerablemente mayor frente a los casos anteriores, donde el tráfico de voz era mayor o similar al de datos. Así, con valores de tamaño de cola superiores a los expuestos en los casos anteriores, con $q = 16$ y $q = 8$, puesto que se aumenta considerablemente el tráfico de datos ($TO_{datos} = 20$ Erlangs), las probabilidades de pérdidas aumentan y para el mismo tráfico ofrecido de voz los valores obtenidos son idénticos. En concreto para un tráfico de voz de 3 Erlangs, la probabilidad de pérdida ronda el 49 % hasta llegar cerca del 54 % para un tráfico de voz de 6,5 Erlangs. Si se disminuye el valor

del tamaño del buffer, $q = 5$ y $q = 6$, se obtienen probabilidades de pérdida cuyos valores aumentan ligeramente para un tráfico de voz entre 4,5 y 5 Erlangs.

Con respecto a la probabilidad de bloqueo (Figura 5.8), nuevamente para tamaños de buffer más altos, 8 y 16, los valores obtenidos son idénticos entre sí, llegando desde el 75 % al 82,5 %. También para valores de 4,5 y 5 Erlangs la probabilidad de bloqueo aumenta ligeramente hasta un 80 % con tamaños de buffer $q = 5$ y $q = 6$.

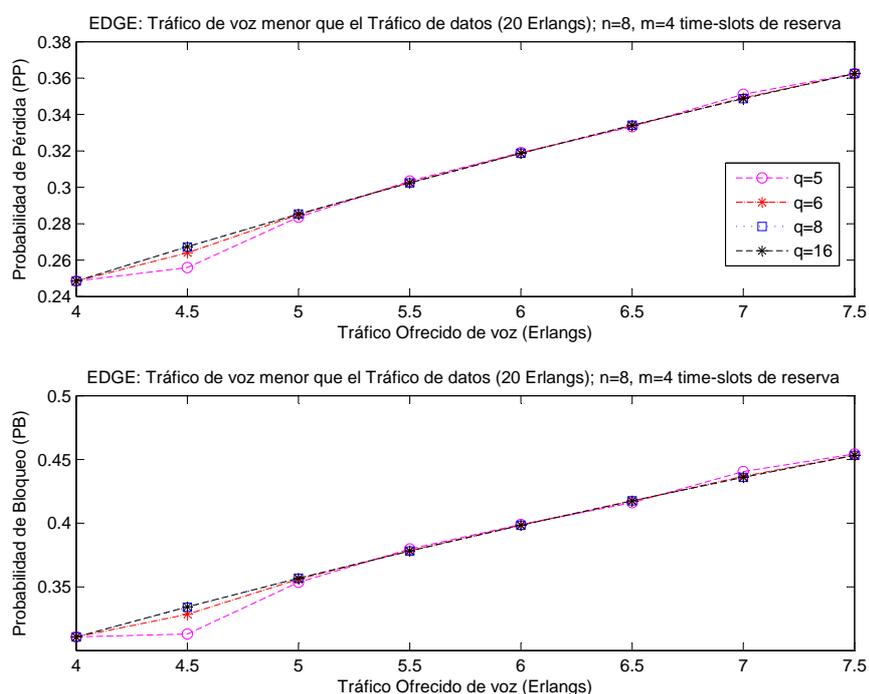


Figura 5.9: Probabilidad de Pérdida y Probabilidad de Bloqueo para un sistema en el que el tráfico de voz es menor que el tráfico de datos, $n=8$, $m=4$

Con el objetivo de disminuir tanto la probabilidad de pérdida como la probabilidad de bloqueo del sistema, en el caso en que el tráfico de voz sea bastante inferior al tráfico de datos ($TO_v \ll TO_d$), se puede observar en la Figura 5.9 que se ha aumentado a un valor de 4 el número de time-slots reservados para el tráfico de voz ($m = 4$). Con este nuevo escenario podemos observar que nuevamente los valores tanto de la probabilidad de pérdida como de la probabilidad de bloqueo son prácticamente similares independien-

temente de la variación en el tamaño de la cola ($q = 5..16$). No obstante las probabilidades de pérdida han disminuido desde el 50,5 % con $m = 1$ (Figura 5.8) hasta cerca del 25 % con $m = 4$ (Figura 5.9), siendo el tráfico ofrecido de voz de 4 Erlangs. La disminución de la probabilidad de bloqueo va del 78 % con $m = 1$ (Figura 5.8) al 31 % con $m = 4$ (Figura 5.9) para un tráfico ofrecido de voz de 6,5 Erlangs.

En general, en este tercer caso en el que el tráfico ofrecido de voz es significativamente menor que el tráfico ofrecido de datos, los valores obtenidos para la probabilidad de pérdida nos indican que son poco aceptables puesto que son considerablemente superiores a los obtenidos cuando el tráfico de voz es menor o similar al tráfico ofrecido de datos. Sin embargo, la información obtenida para la probabilidad de bloqueo en este último caso en el que $m = 4$ time-slots reservados para el tráfico de voz, el sistema obtiene unas probabilidades de bloqueo que rondan los valores obtenidos para el caso donde se aproxima el tráfico ofrecido de voz al tráfico ofrecido de datos.

5.3. Conclusiones

En este capítulo hemos introducido un nuevo mecanismo para la asignación de los recursos que, además de utilizar la reserva de time-slots, se incorpora el buffer de longitud limitada.

A partir del esquema presentado se analiza el modelo correspondiente al esquema de asignación compartida incluyendo la reserva para el tráfico de voz y el *buffering* para el tráfico de datos. Enumerando el número posibles de estados del sistema y representándolos vectorialmente.

En la resolución del modelo se incluyen diagramas de transición que muestran ejemplos de las correspondientes ecuaciones de Markov y, con expresiones analíticas cerradas, definen para cada uno de los estados las ecuaciones genéricas - independientemente del número de time-slots de la trama, el número de recursos reservados para el tráfico de voz y el número de elementos de la cola - que resuelven el sistema.

Por último, la resolución del sistema por medio de las ecuaciones de transición de los estados, nos permiten concluir distintos criterios que permiten al operador de red disponer de parámetros de QoS como son el análisis de la probabilidad de bloqueo del sistema y la probabilidad de pérdida del mismo. De esta forma, como conclusión general se obtiene que si el tráfico ofrecido de voz es significativamente superior al tráfico ofrecido de datos, la probabilidad

de pérdida del sistema recoge valores aceptables, si bien se demuestra que los valores para los cuales se pueden obtener pérdidas razonables es aproximando el valor del tráfico de voz al tráfico de datos. En ningún caso las pérdidas son aceptables cuando el tráfico ofrecido de voz es significativamente inferior al tráfico ofrecido de datos. Las probabilidades de bloqueo igualmente son aceptables cuando los dos tipos de tráfico, voz y datos, son aproximadamente similares entre sí, si bien aumentando sustancialmente el número de time-slots reservados para el tráfico de voz y aumentando el tamaño del buffer cuando el tráfico ofrecido de voz es considerablemente inferior al tráfico ofrecido de datos, se obtienen probabilidades de bloqueo que disminuyen con respecto a los casos estudiados.

Capítulo 6

Conclusiones y líneas futuras

6.1. Conclusiones

El objetivo general de esta tesis es la de proponer mecanismos para el desarrollo del sistema móvil celular EDGE. En particular, aportar contribuciones para la mejora de la eficiencia en cuanto a la asignación de los recursos radio en esta red móvil para el tráfico de voz y de radio paquetes.

Una de las propuestas generales es la de aportar nuevos modelos de forma analítica que permitan estudiar el rendimiento de la asignación de los recursos a los canales de los usuarios mejorando las tasas de pico. En concreto se desarrolla un modelo basado en cadenas de Markov para el estudio de la asignación de recursos en un entorno FDMA/TDMA.

En un entorno de referencia como es el sistema EDGE, la gestión de los recursos radio por parte de la red se centrará en la ordenación de estos recursos y la localización de cada uno de ellos para utilizar distintos criterios que permitan maximizar el uso de estos recursos ofreciendo la calidad de servicio (QoS) requerida por los usuarios.

Se presenta un primer modelo con distribuciones estadísticas tanto para voz como para datos correspondientes a llegadas markovianas y que se caracterizan por el uso ordenado de la asignación del tráfico de voz. Este primer esquema se presenta con la intención de maximizar el número de time-slots consecutivos libres (ráfagas) para la posterior asignación del tráfico de datos a los usuarios que desean transmitir mensajes, imágenes y vídeo. Empleando un método recursivo se ha enumerado y contabilizado todas las posibles secuencias de ráfagas para un valor dado de time-slots libres consecutivos.

Los resultados obtenidos con este primer modelo concluyen en que el mayor porcentaje de tiempo del sistema corresponde a estados con ráfagas cortas, en consecuencia en un posible esquema de planificación (scheduling) se atenderá de forma prioritaria a los mensajes más cortos con objeto de reducir el retardo global del sistema. Por el contrario con el método *repacking* se han obtenido resultados que revelan que las probabilidades de obtener ráfagas más largas son más altas y, por tanto, con un mayor beneficio para las aplicaciones que requieran una tasas de pico mayores.

Por un lado, en un segundo modelo se presenta la reserva de recursos radio para el tráfico de voz. En este esquema en el que la reserva de los time-slots está destinada al tráfico de voz y el resto de recursos se comparte por voz y datos, por medio de expresiones analíticas se concluye que las probabilidades de encontrar ráfagas largas aumenta a medida que la probabilidad de bloqueo del sistema disminuye. Si se examinan los escenarios con reserva de recursos para el tráfico de voz existe una característica común a todos ellos y es que la ráfaga mayor se concentra en la ráfaga de longitud 1 y la próxima ráfaga más larga corresponde al valor de ráfaga mayor posible, teniendo en cuenta los time-slots reservados para el tráfico de voz. Además se garantiza que las llamadas de voz sigan disponiendo de recursos incluso cuando el tráfico de datos aumente y también en el caso de traspasos procedentes de otras celdas.

Por otro lado, en el esquema de asignación de recursos utilizando la reserva de time-slots para el tráfico de datos, se puede concluir que se incrementan las tasas de pico del tráfico de datos de los usuarios disminuyendo el número de recursos reservados, independientemente de la probabilidad de bloqueo se incrementa la probabilidad de encontrar ráfagas para longitudes intermedias, se garantiza que las aplicaciones de datos sigan disponiendo de recursos incluso cuando el tráfico de voz sea alto y también en el caso de los traspasos.

En un tercer modelo el análisis de la utilización de memoria temporal para el tráfico de datos permite disminuir las pérdidas del sistema y aumentar la eficiencia del mismo. No obstante es necesario plantear que el tráfico de datos puede disponer de sistemas con buffering que, en concreto, correspondan a modelos de espera y pérdida con colas limitadas para minimizar la pérdida de paquetes y disminuir el bloqueo del sistema. En resumen, se obtiene que cuando el tráfico ofrecido de voz es significativamente superior al tráfico ofrecido de datos las pérdidas en el sistema disminuyen, si bien se demuestra que los valores para los cuales se pueden obtener pérdidas razonables es cuando los valores del tráfico de voz es aproximadamente similar al tráfico de datos. En ningún caso las pérdidas son aceptables cuando el tráfico ofrecido de voz

es significativamente inferior al tráfico ofrecido de datos. El bloqueo del sistema también es inferior en el caso en el que los tráficos de voz y de datos son aproximadamente similares entre sí, si bien aumentando sustancialmente el número de time-slots reservados para el tráfico de voz y aumentando el tamaño del buffer cuando el tráfico ofrecido de voz es considerablemente inferior al tráfico ofrecido de datos, se ratifica que la probabilidad de bloqueo del sistema disminuye.

Es de destacar que este trabajo sintetiza de forma genérica, los modelos propuestos por medio de expresiones analíticas cerradas que permiten obtener las probabilidades de encontrar un grupo de recursos radio consecutivos en el tiempo. De esta forma se simplifican los mecanismos de resolución empleados y planteados actualmente en la literatura científica. Este hecho permite que puedan ser asignados los recursos a uno o varios usuarios para satisfacer los requisitos de los servicios del tráfico de voz y de datos que gestionan los operadores. Así, independientemente del número de time-slots que utilice la trama la validez de los modelos es general y, por tanto, de aplicación en otras tecnologías actuales y futuras.

En resumen las contribuciones de esta tesis se basan en líneas de ingeniería de tráfico que permiten desarrollar por medio de expresiones cerradas genéricas la distribución de los recursos radio de una red de comunicaciones en un entorno TDMA. Es necesario para ello que la asignación de los recursos sea contiguas para minimizar el retardo y maximizar el *throughput*.

6.2. Líneas futuras de investigación

A partir de este trabajo se abre un conjunto de posibilidades sobre las que puede resultar de interés profundizar y que a continuación se enumeran:

- Estudiar parámetros como el retardo y la variación del retardo dentro del entorno EDGE haciendo uso de los esquemas de asignación de los recursos utilizados en este trabajo.
- Estudiar y analizar diversas políticas de planificación (scheduling), a partir de los esquemas presentados en este trabajo, con objeto de estimar en qué medida puede afectar la gestión de la asignación de los recursos a las distintas aplicaciones en cuanto a los retardos de cada uno de los servicios que se pretenden ofrecer. Incrementando las tasas de transmisión de forma selectiva para las aplicaciones de datos sin que

ello signifique la degradación del resto de las aplicaciones de datos y el servicio de voz.

- Analizar el orden de asignación de los recursos en el que se ha de distribuir el tráfico, tanto de datos como de voz. En concreto estudiar mecanismos que, en el momento de asignar los recursos a los usuarios, se estimen las posibilidades de ocupar aquellos recursos siempre con el mínimo o siempre con el máximo número de identificador de time-slot disponible independientemente del tipo de tráfico. De esta forma la asignación de recursos no siempre es contraria al sentido de asignación para los datos, por un lado, y, por otro lado, la asignación de las llamadas de voz.
- Estudiar y analizar los tiempos de permanencia en cada uno de los estados, usando los esquemas de asignación de los recursos compartida para voz y datos junto con el mecanismo de reserva y el de *buffering*, con objeto de conocer cuánto tiempo se puede mantener las tasas de datos o bien disminuir las tasas para permitir a otros usuarios compartir los recursos.
- Analizar y estudiar haciendo uso de los esquemas estudiados, los modelos de distribución de llegadas combinadas con las llegadas del tráfico de voz Markoviano, el tráfico de datos que se pueda ajustar al actual tráfico IP como son los procesos MMAP o procesos NIRP (múltiples fuentes ON/OFF).
- Estudiar y evaluar nuevos entornos de trabajo, como por ejemplo los sistemas de transmisión mediante luz visible, usando lámparas como transceptores (tecnología conocida como *Visible Light Communications*), según propone el estándar que actualmente desarrolla el IEEE como ampliación de la familia Bluetooth (IEEE 802.15.7). En este caso se trataría de estudiar la asignación de los recursos en una red de sensores. En concreto, se verificarían técnicas de asignación de canal cuando el coordinador no tiene recursos (time-slots) que asignar a un nuevo usuario, teniendo en cuenta las limitaciones que el medio físico supone en este caso, diferenciándola del canal radio. Así se pueden mapear los parámetros de QoS y como asignar en función del dispositivo inalámbrico múltiple de time-slots para reducir los retardos de transmisión/recepción.

- En las aplicaciones en las comunicaciones espaciales igualmente se trata de un esquema FDMA/TDMA con un repositorio de portadoras y time-slots en el que se requiere mejorar los parámetros de QoS para altas frecuencias. Se trata también de estudiar la asignación de los recursos pero incorporando en vez de una asignación aleatoria de los recursos que sea ordenada y en función de la capacidad requerida por los usuarios.
- Mapear para LTE (UMTS Long Term Evolution) la asignación ordenada de time-slots. En el estándar LTE los recursos tiempo-frecuencia se subdividen de acuerdo a una estructura de trama de 10ms, que a su vez se subdivide en diez subtramas de 1ms, cada una de las cuales se divide en dos time-slots de 0,5ms. Cada time-slot comprende siete o seis símbolos OFDM, según el caso, realizar el mapeado de grupos de elementos de recursos usando esquemas como los presentados en este trabajo.

Bibliografía

- [3GPP GERAN (2011)] 3GPP, “Technical Specification Group GSM/EDGE Radio Access Network. Multiplexing and multiple access on the radio path (Release 10),” 3GPP TS 45.002 v.10.1.0 (May 2011)
- [3GPP GERAN (2004)] 3GPP, “Technical Specification Group GSM/EDGE Radio Access Network. Overall description - stage 2(Release 5),” 3GPP TS 43.051 (2003-2004)
- [3GPP HSPA (2007)] 3GPP TS 25.999 High Speed Packet Access (HSPA) evolution (Release 7)” 3GPP 2007
- [3GPP LTE (2008)] 3GPP, “3GPP TS 32.521 Long Term Evolution (Release 8),” 3GPP (2008)
- [3GPP Release’99] 3GPP, “Quality of Service (QoS), concept and architecture.(Release 1999),” 3GPP TS 23.107 (1999)
- [ETSI GPRS (2000)] ETSI “Technical Specification GSM 03.64, Overall description of the GPRS radio interface – stage 2,” ETSI V8.4.0, April 2000.
- [ETSI GPRS (1997)] ETSI, GSM 03.64, “Overall Description of the GPRS Radio Interface”, version 5.1 .0, 1997- I 1.
- [HSCSD ETSI (1997)] ETSI, GSM 03.34, ”High Speed Circuit Switched Data (HSCSD)” , version 5.0.1, 1997-04.
- [UMTS ETSI (2000)] ETSI, UMTS 21.01, ”Requirements for the UMTS Terrestrial Radio Access System (UTRA)”, version 3.0.1, 1997-10. Recommendation ITU-R M. 1225, ”Guidelines for Evaluation of Radio Transmission Technologies for IMT-2000”.
- [UMTS ETSI (1997)] ETSI, ”Selection procedures for the choice of radio transmission technologies of the Universal Mobile Telecommunications System (UMTS),” ETSI UMTS 04-02 UTRA, June 1997.
- [Ajib et alt. (1999)] W. Ajib and P. Godlewski, “Acknowledgment operations in the RLC layer of GPRS,” In IEEE International Mobile Multimedia Communications, 1999. (MoMuC ’99), pp. 311-317, Nov. 1999.
- [Ameigeiras et alt. (2000)] P.J. Ameigeiras Gutiérrez, J. Wigard, P.N. Andersen, H.C. Damgaard and P. Mogensen, “Performance of link adaptation in GPRS networks,” In Vehicular Technology Conference, 2000. IEEE VTS-Fall VTC 2000. 52nd, vol. 2, pp. 492-499, Sept. 2000.

- [Al Agha et al. (2000)] K. Al Agha and K. Boussetta, "Dynamic slot allocation for multicasting in GPRS systems," In IEEE 51st Vehicular Technology Conference. 2000 VTC, vol. 3, pp. 2355-2359, Spring 2000.
- [Bada et al. (2002)] J. Bada, and F. Casadevall, "Performance of service disciplines in GPRS systems with heterogeneous traffic," Electronics Letters, vol. 38, pp. 247-248, Feb. 2002
- [Balachandran et al. (2001)] K. Balachandran, K. Chang, W. Luo and S. Nanda, "System level interference mitigation schemes in EGPRS: Mode-0 and scheduling," IEEE Vehicular Technology Conference, VTC2001, vol. pp. 2489-2493, 2001
- [Ball et al. (2004)] C.F. Ball, K. Ivanov, L. Bugi and P. Stöckl, "Analysis and optimization of the (E)GPRS RLC protocol by simulations and measurements," In the 15th IEEE International Symposium on Personal, Indoor and Mobile Radio Communications, PIMRC 2004, vol. 1, pp. 115-119, Sept. 2004.
- [Barreto et al. (2007)] A.N. Barreto, L.G.U. García and E. Souza, "GERAN Evolution for increased speech capacity," In Vehicular Technology Conference, 2007, VTC2007-Spring. IEEE 65th, pp. 22-25, April 2007.
- [Bestak (2008)] R. Bestak, "Evolution of Mobile Networks" IEEE The 15th International Conference on Systems, Signals and Image Processing, IWSSIP 2008, pp.29 – 32, 2008
- [Bettstetter (1999)] C. Bettstetter, "Global Wireless Internet Access with GPRS," In Proc. EUNICE Open European Summer School Barcelona, Spain, pp. 151-162, Sept 1999.
- [Bilgic et al. (1999)] M. Bilgic, K. Essigmann, T. Holmstrom, M. Lord and M. Renschler, "Quality of Service in General Packet Radio Service," IEEE International Workshop on Mobile Multimedia Communications, pp. 226-231, Nov. 1999.
- [Brasche and Walke (1997)] G. Brasche and B. Walke, "Concepts, services, and protocols of the new GSM phase 2+ General Packet Radio Service," IEEE Communications Magazine, vol. 35, pp. 94-104, Aug. 1997.
- [Cai and Goodman (1997)] J. Cai and D. Goodman, "General Packet Radio Service in GSM," IEEE Communications Magazine, vol. 35, pp. 122-131, Oct. 1997.
- [Cai et al. (2000)] J. Cai, L.F. Chang, K. Chawla and X. Qiu, "Providing differentiated services in EGPRS through packet scheduling," IEEE Global Telecommunications Conference 2000, GLOBECOM'00 vol. 3, pp. 1515-1521, Nov.-Dec. 2000
- [Calin et al. (1999)] D. Calin, S. Malik and D. Zeglache, "Traffic scheduling and fairness for GPRS air interface," IEEE Vehicular Technology Conference, VTC1999, vol. pp. 834-838, 1999
- [Cavalcanti et al. (2002)] F.R.P. Cavalcanti, W.M. Sousa Jr., Y.C.B. Silva and T.F. Maciel, "Combined performance of packet scheduling and smart antennas for data transmission in EGPRS," IEEE 55th Vehicular Technology Conference, 2002. VTC Spring 2002, vol. 2, pp. 797 - 801, May 2002

- [Chang et al. (2001)] L.F. Chang, X. Qiu, K. Chawla and J. Cai, "Providing differentiated services in EGPRS through radio resource management," IEEE International Conference on Communications, ICC 2001, vol. 7, pp. 2296-2301, June 2001
- [Chen et al. (1996)] L. Chen, S. Yoshida, and H. Murata, "A dynamic channel assignment algorithm for voice and data integrated TDMA Mobile Radio," Proceedings of 1997 International Conference on Information, Communications and Signal Processing, ICICS 1997, vol. 1, pp. 215-219, 1997.
- [Cheng et al. (2008)] H. Cheng, X. Wang, M. Huang, S. Yang, "A Review of Personal Communications Services", The 9th International Conference for Young Computer Scientists, pp. 616-621 IEEE Computer Society, 2008
- [Demetrescu (1997)] C. Demetrescu, "LLC-MAC Analysis of General Packet Radio Service in GSM," In Bell Labs Technical Journal, Lucent Technologies, pp. 37-50, July-Sept. 1999.
- [Featherstone et al. (2001)] W. Featherstone and D. Molkdar, "Performance analysis of varying the link adaptation BLER switching point in EGPRS," In 12th IEEE International Personal, Indoor and Mobile Radio Communications, vol. 2, pp. F-116 - F120. Sep-Oct. 2001
- [Feng et al. (2005)] H. Feng and Y. Tsai, "Using priority, buffering, threshold control, and reservation techniques to improve channel-allocation schemes for the GPRS system," IEEE Transactions on Vehicular Technology, vol. 54, no. 1, pp. 286-306, January 2005.
- [Furuskär et al. (1999)] A. Furuskär, D. Bladsjö, S. Eriksson, S. M. Frodigh, S. Jäverbring and H. Olofsson, "System performance of the EDGE concept for enhanced data rates in GSM and TDMA/136" Wireless Communications and Networking Conference, 1999. WCNC. 1999 IEEE vol.2 pp. 752-756, 1999
- [Furuskär et al. (1999)] A. Furuskär, M. Frodigh, H. Olofsson and J. Skold, "EDGE: Enhanced Data Rates for GSM and TDMA/136 Evolution", in IEEE Personal Communications, pp. 56-66, June 1999.
- [Gozalvez et al. (2004)] J. Gozalvez, M. Lopez-Benitez and O. Lazaro, "Guaranteeing Quality of Service in mobile radio networks by means of link adaptation algorithms," Symposium on Wireless Communications Systems, vol. 7, pp. 5230-5234, Sept. 2004
- [Gozalvez et al. (2000)] J. Gozalvez and J. Dunlop, "GPRS link adaptation switching thresholds and intervals," Electronics Letters vol. 36, pp. 1311-1312. July 2000
- [Hälönen et al. (2002)] T. Hälönen, J. Romero and J. Meleró, "GSM, GPRS and EDGE Performance - Evolution towards 3G/UMTS", John Wiley and Sons, Chichester, 2002. Citado en pag.6
- [Hlavacs et al. (2002)] H. Hlavacs, G. Haring, A. Kamra and M. Bansal, "Modeling Resource Management for Multi-Class Traffic in Mobile Cellular Networks," Proceedings of the 35th Hawaii International Conference on System Sciences, Page(s): 1539 - 1548, 2002.

- [Hoff et al. (1998)] S. Hoff, M. Meyer and J. Sachs, "Analysis of the General Packet Radio Service (GPRS) of GSM as Access to the Internet," In IEEE International Conference on Universal Personal Communications 1998, ICUPC'98, vol. 1, pp. 415-419, Oct. 1998.
- [Hong et al. (1986)] D. Hong and S. Rappaport, "Traffic model and performance analysis for cellular mobile radio telephone systems with prioritized and nonprioritized handoff procedures," IEEE Trans. on Vehicular Technology, Page(s): 77 - 92 vol.3, August 1986.
- [Hoymann et al. (2002)] C. Hoymann, P. Stuckmann, "Traffic Engineering for the evolution of GPRS/EDGE networks," The 54th IEEE Vehicular Technology Conference, vol. 2, pp. 558-561, Fall 2001.
- [Irnich et al. (2002)] T. Irnich and P. Stuckmann, "Analytical performance evaluation of Internet access over GPRS and its comparison with simulation results" The 13th IEEE International Symposium on Personal, Indoor and Mobile Radio Communications, 2002 Page(s): 2127 - 2131 vol.5, September 2002.
- [Itkonen et al. (2002)] J. Itkonen, V. Salomaa and J. Lempiäinen, "Air interface capacity for GPRS/EDGE over GSM traffic load," In IEEE 56th Vehicular Technology Conference. 2002 VTC, vol. 1, pp. 396-400, Fall 2002.
- [Ivanov et al. (2008)] K. Ivanov, C.F. Ball, R. Müllner and H. Winkler, "Paving the path for high data rates by GERAN evolution EDGE2 with dual-carrier," IEEE 29th International Symposium on Personal, Indoor and Mobile Radio Communications, 2008. PIMRC2008, pp. 1-6, Sept. 2008.
- [Ivanovich et al. (2002)] M. Ivanovich, M. Zukerman, P. Fitzpatrick and M. Gitlis, "Performance between circuit allocation schemes for half- and full-rate connections in GSM," Proceedings of the IEEE 56th Vehicular Technology VTC 2002-Fall, Volume 2, 24-28 Sept. 2002 Page(s):1279 - 1283 vol.2, September 2002.
- [Jaupitre et al. (2002)] J. Itkonen, V. Salomaa and J. Lempiäinen, "Air interface capacity for GPRS/EDGE over GSM traffic load," In IEEE 56th Vehicular Technology Conference. 2002 VTC, vol. 1, pp. 396-400, Fall 2002.
- [Johanson et al. (1998)] C. Johanson, L. Verdier and F. Khan, "Performance of different scheduling strategies in a packet radio system," IEEE International Conference on Universal Personal Communications vol.1, pp. 267-271, 1998
- [Kalden et al. (2000)] R. Kalden, I. Meirick and M. Meyer, "Wireless Internet Access based on GPRS," In IEEE Personal Communications vol. 4, pp. 8-18, 2000.
- [Kennedy et al. (1999)] K. Kennedy and R. Litjens, "Performance evaluation of a hybrid radio resource allocation algorithm in a GSM/GPRS network," In the 10th IEEE International Symposium on Personal, Indoor and Mobile Radio Communications, PIMRC 1999, vol. 1, pp. 21-26, Sept. 1999.
- [Kochem et al. (2003)] A.C.B. Kochem and E.L. Bodanese, "A quality of service management scheme over GPRS," In Proceedings IEEE SoutheastCon, pp. 74-80, April 2003.

- [Kodikara et al. (2004)a] C. Kodikara, S. Worrall, A.M. Kondo and S.N. Fabri, ‘‘Link adaptation for streaming video in EGPRS mobile networks,’’ IEEE Vehicular Technology Conference, VTC2004, vol. 5, pp. 2763-2767, May 2004
- [Kodikara et al. (2004)b] C. Kodikara, S.N. Fabri, A.M. Kondo, ‘Link adaptation for real-time video communications networks,’’ IEE Proceedings-Communications, vol. 151, pp. 438- 444, Oct. 2004
- [Lee et al. (2002)] D-S. Lee and C-M. Chen, ‘‘Weighted Fair Queueing and compensation techniques for wireless packet switched networks,’’ The 5th International Wireless Personal Multimedia Communications 2002, vol. 3, pp. 887-891, Oct. 2002
- [Leung et al. (2001)] K.K. Leung, P.F. Driessen, K. Chawla and X. Qiu, ‘‘Link adaptation and power control for streaming services in EGPRS wireless networks,’’ IEEE Journal on selected areas in communications, vol. 19, pp. 2029-2039, Oct. 2001
- [Ludwig et al. (1997)] R. Ludwig and D. Turina, ‘‘Link Layer Analysis of the General Packet Radio Service for GSM,’’ In IEEE International Conference on Universal Personal Communications 1997, ICUPC’97, vol. 2, pp. 525-530, Oct. 1997.
- [Mihovska et al. (2009)] A. Mihovska, E. Tragos, E. Mino, J. Luo, C. Mensing, R. Fracchia, S. Horrich, L. Hui, A. Klockar and S. Kyriazakos, ‘‘Requirements and Algorithms for Cooperation of Heterogeneous Radio Access Networks,’’ In Wireless Personal Communications: An International Journal, Kluwer Academic Publishers Hingham, MA, USA. vol. 50, issue 2, pp. 207-245 , Jul. 2009.
- [Molkdar et al. (2001)] D. Molkdar and W. Featherstone, ‘‘Impact of imperfect link adaptation in EGPRS,’’ Second International Conference on 3G Mobile Communication Technologies, pp. 277-281. 2001
- [Molkdar et al. (2002)] Molkdar, D. Featherstone and W. Larnbotharan S, ‘‘An overview of EGPRS: the packet data component of EDGE,’’ in Electronics and Communication Engineering Journal, vol. 14, no. 1, pp. 21-38, Feb. 2002.
- [Müllner et al. (2004)] R. Müllner, C.F. Ball, K. Ivanov and H. Winkler, ‘‘Advanced Quality of Service strategies for GERAN mobile radio networks,’’ In the 15th IEEE International Symposium on Personal, Indoor and Mobile Radio Communications, PIMRC 2004, vol. 3, pp. 2161-2165, Sept. 2004.
- [Ni et al. (1999)] S. Ni and S-G. Häggman, ‘‘GPRS performance estimation in GSM circuit switched services and GPRS shared resource systems,’’ In IEEE International Symposium on Wireless Communications and Networking Conference, WCNC 1999, vol. 3, pp. 1417-1421 , Sept. 1999.
- [Pang et al. (1999)] Q. Pang, A. Bigloo, V. Leung, C. Scholefield, ‘‘Service scheduling for General Packet Radio Service,’’ IEEE Wireless Communications and Networking Conference vol.3, pp.1229-1233, 1999
- [Park (2002)] J.H. Park, ‘‘Wireless Internet Access for Mobile Subscribers based on the GPRS/UMTS network,’’ In IEEE Communications Magazine, pp. 38-49, April 2002.

- [Prasad et al. (2009)] R. Prasad and A. Mihovska "New Horizons in mobile and wireless communications. Volume 1. Radio Interfaces," Mobile Communications Series. Artech House. 2009.
- [Qiu et al. (2000)] X. Qiu, K. Chawla, L.F. Chang, J. Chuang, N. Sollenberger and J. Whitehead, "An enhanced RLC/MAC design for supporting integrated services over EGPRS," In Wireless Communications and Networking Conference, WCNC. 2000 IEEE, vol. 2, pp. 907-912, Sept. 2000.
- [Queseth et al. (1999)] O. Queseth, F. Gessler and M. Frodigh, "Algorithms for link adaptation in GPRS," In IEEE 49th Vehicular Technology Conference, vol. 2, pp. 943-947, May 1999.
- [Ramirez et al. (2008)] C. Ramírez, S.P. Mansilla and J. Paradells, "Improved Channel Allocation for Voice and Data Traffic with Resource Reservation for Voice Traffic in EDGE System," Journal of Communications, JCM, ISSN 1796-2021 Copyright 2006-2011 by ACADEMY PUBLISHER vol 3, No 2 pp. 34-42, Apr 2008.
- [Ramirez et al. (2007)] C. Ramírez, J. Paradells, S.P. Mansilla, "Improving EDGE for rural environments through channel allocation for voice and data traffic," 2nd IEEE International Symposium on Wireless Pervasive Computing, February 2007.
- [Ramirez et al. (2006)a] C. Ramírez, J. Paradells, S.P. Mansilla, "Analysis of consecutive idle time slots in mobile networks," 17th Annual IEEE International Symposium on Personal, Indoor and Mobile Radio Communication, pp. 1-5, September 2006.
- [Ramirez et al. (2006)b] C. Ramírez, J. Paradells, S.P. Mansilla, "Radio resource allocation strategies to guarantee data traffic in cellular networks," 3rd IEEE International Symposium on Wireless Communication Systems, pp. 597 - 600, September 2006.
- [Ramirez et al. (2002)a] C. Ramírez, R. Vidal and J. Paradells, "An integrated QoS and mobility management scheme for cellular packet-data networks," In Proceedings LONIIS Workshop on Telecommunication Networks and Teletraffic Theory, pp. , Jan-Feb. 2002
- [Ramirez et al. (2002)b] C. Ramírez and J. Paradells, "Analysis of scheduling schemes to provide quality of service in cellular packet-data networks," International Conference Telecommunications and Electronics, pp. July 2002
- [Ramírez et al. (2000)] C. Ramírez and J. Paradells, "GPRS and approach of quality of service end to end using a protocol based on wireless networks ," The 6th International Conference on Information Systems Analysis and Synthesis, pp. - ,July 2000
- [Saily et al. (2008)] M. Säily, J. Hukkonen and E. Zacarías, "EGPRS2 uplink performance for GERAN evolution," In Vehicular Technology Conference, 2008, VTC2008-Spring. IEEE, pp. 2572-2576, May 2008.
- [Sau et al. (1998)] J. Sau and C. Scholefield, "Scheduling and Quality of Service in the General Packet Radio Service (GPRS)," IEEE International Conference on Universal Personal Communications vol. 2, pp. 1607-1071 , 1998

- [Simonsson et al. (2000)] A. Simonsson, M. Almgren and M. Thurfjell, "A power control and scheduling concept for EGPRS," IEEE Vehicular Technology Conference, VTC 2000, vol. 5, pp. 2219-2224, Sept. 2000
- [Skehill et al. (2007)] R. Skehill, M. Barry, W. Kent, M. O'Callaghan, N. Gawley and S. Mcgrath, "The common RRM approach to admission control for converged heterogeneous wireless networks," In IEEE Wireless Personal Communications, vol. 14, issue 2, pp. 48-56, April 2007.
- [Stuckmann et al. (2001)] P. Stuckmann and O. Paul, "Dimensioning GSM/GPRS networks for circuit and packet switched services," In Proc. of the 10th Symposium on Wireless Personal Multimedia Communications, pp. 597 - 602, Sept. 2001.
- [Stuckmann et al. (2002)] P. Stuckmann, N. Ehlers, and B. Wouters, "GPRS traffic performance measurements," IEEE 56th Vehicular Technology Conference, vol.2, pp. 1289 - 1293, 2002.
- [Stuckmann et al. (2003)] P. Stuckmann and O. Möller, "Advanced scheduling and admission control techniques for cellular packet radio networks," The 5th European Personal Mobile Communications Conference, 2003 pp. 214-219, April 2003
- [Tan et al. (1995)] P.T. Tan, E. Gunawan, C.B. Soh and B.H. Soong, "Dynamic model of prioritized handoff procedures for cellular mobile networks" IEEE Global Telecommunications Conference, 1995. GLOBECOM '95., Page(s): 519 - 523 vol.1, 1995.
- [Tang et al. (2004)] S.Y. Tang, S. Thilakawardana and R. Tafazolli, "Dynamic radio resource management in GSM/GPRS using scalable resource allocation technique," In IEEE International Personal, Indoor and Mobile Radio Communications, PIMRC'04 vol.1 pp. 545-550, 2004
- [Timus et al. (2011)] B. Timus, P. Soldati, K. Dongwoo Kim and J. Zander, "Cross-Layer Resource Allocation Model for Cellular-Relaying Network Performance Evaluation," IEEE Transactions on Vehicular Technology, vol.60, pp. 2765 - 2776, 2011.
- [Tsao (2000)a] S-L. Tsao "Extending Earliest-Due-Date scheduling algorithms for wireless networks with location-dependent errors," IEEE Vehicular Technology Conference, VTC 2000, vol. 1 pp. 223-228, 2000
- [Tsao (2000)b] S. Tsao "Quality of Service over GPRS data network," Vehicular Technology Conference, IEEE VTC-Fall VTC 2000, 52nd, vol. 3, pp. 1001-1007, Sept. 2000.
- [Turina et al. (1996)] D. Turina, P. Beming, E. Schoster and A. Andersson, "A proposal for multi-slot MAC layer operation for packet data channel in GSM," 5th IEEE International Conference on Universal Personal Communications, ICUPC 1996, vol.2, pp. 572-576, 1996.
- [Vieira et al. (2006)] R.D. Vieira, L. Aguayo and G.W.O. Costa, "EQOS Call Admission and Scheduler Parameter Evaluation for Geran," IEEE 17th International Symposium on Personal, Indoor and Mobile Radio Communications, 2006, pp. 1-5, Sept. 2006
- [Vornfeld (2002)] U. Vornfeld, "Analytical Concepts for GPRS Network Dimensioning," IEEE International Conference on Communications, vol. 5, pp. 3156 - 3160, 2002.

-
- [Vornefeld (2003)] U. Vornefeld, "Analytical Models and Methods for E-GPRS Performance Evaluation," International Conference on Communication Technology Proceedings, vol. 2, pp. 1363 - 136, 2003.
- [Vriendt et al. (2002)] J. De Vriendt, P. Lainé, C. Lerouge and X. Xu, "Mobile Network Evolution: A revolution on the move" IEEE Communications Magazine, pp.104-111, April 2002
- [Yang et al. (2001)] J-S. Yang, C-C Tseng and R-G. Cheng, "Dynamic scheduling framework on RLC/MAC layer for General Packet Radio Service," International Conference Distributed Computing Systems Workshop , pp. 441-447, April 2001
- [Zander et al. (2001)] J. Zander and S-L Kim, "Radio resource management for wireless networks ," Artech House mobile communications series, ISBN 1-58053-146-6, 2001.
- [Zhang et al. (2000)] M. Zhang and A. MacDonald, "Performance simulation for shared voice/data EDGE channel," In IEEE 54th Vehicular Technology Conference. 2001 VTC, vol. 3, pp. 1394-1401, Fall 2001.
- [Zangi et al.(1998)] K. C. Zangi, A. Furuskar and M. Hook, "EDGE: Enhanced Data for Global Evolution of GSM and IS-136", Proc. Multi Dimensional Mobile Communications, MDMC98, 1998.