

UNIVERSIDAD POLITÉCNICA DE VALENCIA
DEPARTAMENTO DE COMUNICACIONES



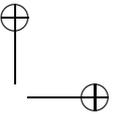
TESIS DOCTORAL

«CONTRIBUCIÓN AL CONTROL DE ADMISIÓN EN REDES MÓVILES
CELULARES MULTISERVICIO»

Autor: David García Roger
Ing. de Telecomunicación

Directores: Jorge Martínez Bauset
Vicent Pla Boscà
Dres. Ingros. de Telecomunicación

VALENCIA
FEBRERO 2007



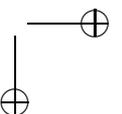
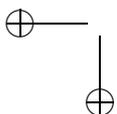
Clouds are not spheres, mountains are not cones, coastlines are not circles, and bark is not smooth, nor does lightning travel in a straight line.

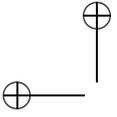
Benoît B. Mandelbrot

All models are wrong, but some are useful.

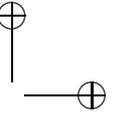


George Edward Pelham Box





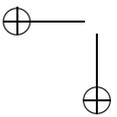
|



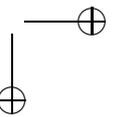
—

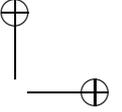
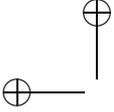
—

*Pro M. Dominica dies
Dulcis et mea mater.*



|





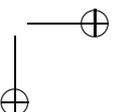
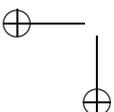
Agradecimientos

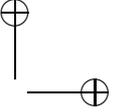
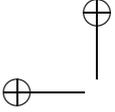
Aunque sólo fuera con unas sencillas palabras, no deseaba dejar pasar la oportunidad de mostrar mi gratitud al conjunto de individuos, colectivos e instituciones sin cuya colaboración esta tesis no hubiera tenido sentido, a saber:

A los miembros del Grupo de Interconexión de Redes de Banda Ancha, esto es: a Vicente por la incesante experiencia investigadora que siempre ha aportado; a Jorge y Vicent por su entusiasmo, asistencia y continuos desvelos en la dirección de mi tesis; a Pablo y José Ramón por el interés mostrado en mi labor y sus esfuerzos por colaborar; a Luis por haberme metido en todo esto; y a María José, José Manuel, Ángel, Nacho y otros tantos compañeros de laboratorio... por haber estado siempre ahí.

A los miembros del Departamento de Comunicaciones de la Universidad Politécnica de Valencia, en especial a Pablo, Alberto, Ángel, Paco y Óscar por compartir sus destrezas docentes con uno que acaba de empezar en esto de la enseñanza universitaria.

Finalmente, agradecer el apoyo recibido de la Unión Europea y su Sexto Programa Marco, a través de la red de excelencia europea *Design and Engineering of the Next Generation Internet* (EuroNGI); del Gobierno Español a través del Ministerio de Ciencia y Tecnología (proyectos TIC2000-1041-C03-02 y TIC2003-08272), y a través del Ministerio de Educación y Ciencia (proyecto TSI2005-07520-C03-03); y de la *Generalitat Valenciana* (beca de formación de personal investigador de carácter predoctoral CTB/PRB/2002/267).



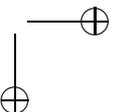
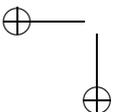


Abstract

Mobile cellular telephony has grown quickly from a ground-breaking initiative (dating back to the start of the 80's) to the way of life that is nowadays, with thousands of millions of user terminals worldwide. However, as with any wireless resource, the ranges of radio frequency spectrum assigned to that type of systems are limited. Therefore, this growth beyond all early forecasts makes the efficient management of the radio resource a major challenge for service providers. In addition, this management contributes to control the demands of a diversified traffic in continuous increase, as well as assures the viability and financial success of that kind of networks.

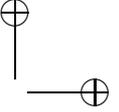
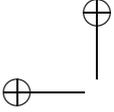
“Session Admission Control” (SAC) is a radio resource management mechanism useful for the design and operation of mobile cellular networks. The particular features of that sort of networks (scarcity and variability in the amount of resources and terminal mobility) are an added difficulty to the SAC. Specifically, mobility makes complex to probabilistically guarantee that a new session will have enough available resources in the visited cells during the session lifetime to meet the “Quality of Service” (QoS) requirements of that new session and all the ongoing sessions already present in the system.

Due to efficiency in the reuse of the radio frequency spectrum cellular fragmentation is increasing, and as a result, cell sizes are being reduced even more, producing a greater handover frequency. These, along with the disparity of QoS requirements, resource demands, and traffic profiles among the evolving next generation services, cause that in the absence of a suitable



management mechanisms the network yields a poor utilization of the radio resources.

The work included in this thesis focuses on contributing in the characterization, understanding and development of mechanisms for the proper resource management in mobile cellular networks. Explicitly, this effort focuses on the development of models, algorithms and methods for studying the admission control from a stationary perspective, and the development of schemes able to optimize the behavior of the admission control in an adaptive way with respect to the nonstationary conditions of the real systems.

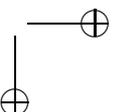
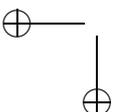


Resum

La telefonia mòbil cel·lular ha passat d'ésser una iniciativa pionera (que data de principis dels anys 80), a transformar-se en el nou estil de vida que és a l'actualitat, amb milers de milions de terminals repartits per tot el món. No obstant, com amb tot recurs radio, la porció de l'espectre assignada per a aquest tipus de sistemes resulta un bé escàs. El creixement més enllà de tota predicció anticipada ha convertit la gestió eficient d'aquest recurs en un desafiament essencial al que els operadors cel·lulars han d'enfrontar-se quotidianament. Aquesta gestió, asismisme, contribueix a administrar les demandes d'un tràfic diversificat i en continuat augment, a la vegada que assegura la viabilitat i l'èxit comercial d'aquest tipus de xarxes.

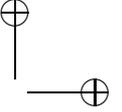
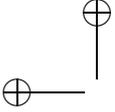
El "Control d'Admissió de Sessions" (CAS) constitueix un mecanisme de gestió dels recursos radio, d'importància en el disseny i operació de les xarxes mòbils cel·lulars. Les característiques pròpies d'aquest tipus de xarxes (escassetat i variabilitat en la quantitat de recursos i la mobilitat dels terminals) aporten una major complicació al CAS. En concret, la mobilitat fa complex el garantir de manera probabilística que una nova sessió tindrà suficients recursos disponibles en les cèl·lules visitades mentre la sessió està en curs com per a respectar els requisits de "Qualitat de Servei" (QoS) de la mencionada nova sessió i del rest de les ja presents en el sistema.

Per raons d'eficiència en la reutilització de l'espectre radioelèctric, la fragmentació cel·lular està experimentant un increment la conseqüència del qual són cèl·lules de tamany cada vegada més reduït, i per això, un augment en



la freqüència amb la que es tenen que produir els traspassos entre cèl·lules. Això, unit a la disparitat de requisits de QoS, demandes de recursos i perfils de tràfic ofert entre tots els serveis emergents de nova generació provoca que l'absència d'un mecanisme adequat de gestió desemboque en una utilització inadequada dels recursos radio.

El treball contingut en aquesta tesi tracta de contribuir a la caracterització, comprensió i desenvolupament de mecanismes per a la convenient gestió dels recursos en les xarxes mòbils cel·lulars. En concret, l'aportació que compren aquest esforç inclou el desenvolupament de models, algoritmes i mètodes per estudiar el control d'admissió des d'una perspectiva estacionària, i el desenvolupament d'esquemes que optimitzen el comportament del control d'admissió de manera adaptativa amb respecte a les situacions no estacionàries dels sistemes reals.

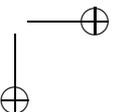
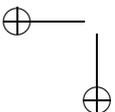


Resumen

La telefonía móvil celular ha pasado de ser una iniciativa pionera (que data de los albores de los años 80), a convertirse en el nuevo estilo de vida que es en la actualidad, con miles de millones de terminales repartidos por todo el mundo. Sin embargo, como todo recurso inalámbrico, la porción de espectro radioeléctrico asignada para este tipo de sistemas resulta un bien escaso. El crecimiento más allá de toda predicción anticipada ha convertido la gestión eficiente de este recurso en un desafío fundamental al que los operadores celulares deben enfrentarse cotidianamente. Dicha gestión, asimismo, contribuye a administrar las demandas de un tráfico diversificado y en continuo aumento, a la vez que asegura la viabilidad y el éxito comercial de este tipo de redes.

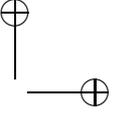
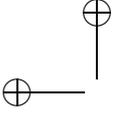
El «Control de Admisión de Sesiones» (CAS) constituye un mecanismo de gestión de los recursos radio, de importancia en el diseño y operación de redes móviles celulares. Las características propias de este tipo de redes (escasez y variabilidad en la cantidad de recursos y la movilidad de los terminales) aportan una mayor complicación al CAS. En concreto, la movilidad hace complejo garantizar probabilísticamente que una nueva sesión tendrá suficientes recursos disponibles en las células visitadas mientras la sesión esté en curso como para respetar los requisitos de «Calidad de Servicio» (QoS) de dicha nueva sesión y del resto ya presentes en el sistema.

Por razones de eficiencia en la reutilización del espectro radioeléctrico, la fragmentación celular está experimentando un aumento cuya consecuencia



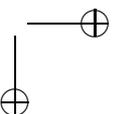
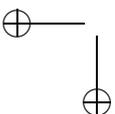
son células de tamaño cada vez más reducido, y por ende, una mayor frecuencia con la que se tienen que producir los traspasos entre células. Esto, unido a la disparidad de requisitos de QoS, demandas de recursos y perfiles de tráfico ofrecido entre todos los servicios de nueva generación emergentes provoca que la ausencia de un mecanismo adecuado de gestión redunde en una utilización inadecuada de los recursos radio.

El trabajo contenido en esta tesis trata de contribuir en la caracterización, comprensión y desarrollo de mecanismos para la apropiada gestión de los recursos en las redes móviles celulares. En concreto, la aportación que comprende este esfuerzo incluye el desarrollo de modelos, algoritmos y métodos para estudiar el control de admisión desde una perspectiva estacionaria, y el desarrollo de esquemas que optimicen el comportamiento del control de admisión de manera adaptativa con respecto a las situaciones no estacionarias de los sistemas reales.

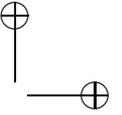
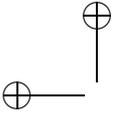


Índice general

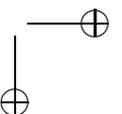
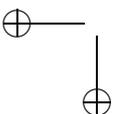
1	Introducción	1
2	El estudio del control de admisión en redes celulares	9
2.1	Modelo del sistema	10
2.1.1	Modelos de colas	11
2.1.2	Suposiciones e hipótesis del modelo	14
2.2	Control de admisión en redes celulares	18
2.2.1	Información a disposición de la política	19
2.2.2	Familias de políticas	19
2.2.3	Procedimientos para el diseño de políticas	24
3	Impacto de la movilidad en el dimensionado de redes	29
3.1	Antecedentes	30
3.2	Descripción del modelo	31
3.2.1	Sistema fijo	31
3.2.2	Sistema móvil	33
3.3	Análisis	34
3.3.1	Caracterización del sistema fijo	34



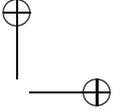
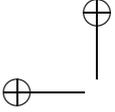
3.3.2	Modelos de tráfico	36
3.3.3	Caracterización del sistema móvil	38
3.3.4	Dimensionado del sistema fijo	40
3.3.5	Dimensionado del sistema móvil	42
3.4	Resultados numéricos	43
3.4.1	Sistema fijo	43
3.4.2	Sistema móvil	44
3.5	Conclusiones	45
4	Optimización del control de admisión	47
4.1	Antecedentes	48
4.2	Descripción del modelo	50
4.3	Políticas de control de admisión	52
4.3.1	Políticas con solución en forma de producto	53
4.3.2	Políticas de <i>trunk reservation</i>	55
4.3.3	Otras políticas	55
4.4	Diseño y Dimensionado	55
4.4.1	Análisis de políticas	57
4.4.2	Síntesis de políticas	62
4.5	Comparativa de políticas	70
4.5.1	Resultados numéricos	71
4.5.2	Sensibilidad	74
4.6	Algoritmo de escalada para la optimización de políticas	79
4.6.1	Descripción del modelo	80
4.6.2	Análisis del modelo	81
4.6.3	Fundamentos de funcionamiento	82



4.6.4	Descripción del algoritmo	89
4.6.5	Evaluación numérica de la complejidad computacional	94
4.7	Conclusiones	97
5	Esquema adaptativo de control de admisión	99
5.1	Antecedentes	101
5.2	Descripción del modelo	104
5.3	Esquema adaptativo	106
5.3.1	Fundamentos	110
5.3.2	Funcionamiento	112
5.4	Análisis	114
5.5	Esquemas adaptativos comparados	117
5.5.1	Esquema ZL	117
5.5.2	Esquema WZZZ	120
5.5.3	Nuestro esquema (caso monoservicio)	121
5.6	Resultados	123
5.6.1	Comparación de esquemas para monoservicio	125
5.6.2	Evaluación de prestaciones para multiservicio	134
5.7	Conclusiones	140
6	Conclusiones	147
	Apéndices	153
A	Notación, variables y parámetros más utilizados	153
B	Abreviaturas y acrónimos	155

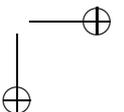
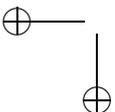


C Publicaciones	159
C.1 Relacionadas con la tesis	159
C.1.1 Capítulo de libro	159
C.1.2 Revista	159
C.1.3 Congreso	160
C.2 Otras publicaciones	162
C.2.1 Congreso	162
 Bibliografía	 163

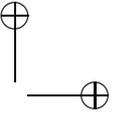
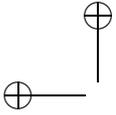


Índice de figuras

2.1	Alternativas consideradas en CA en sistemas celulares.	13
2.2	Familias de políticas de control de admisión.	24
3.1	Dos modelos del sistema	32
3.2	Diagrama de estados para los sistemas BPP	37
3.3	Diagrama de estados de un terminal libre	38
4.1	Dos procedimientos de trabajo con políticas de control de admisión.	56
4.2	Sistema de contraejemplo empleando política GM.	57
4.3	Representación del espacio de estados del sistema de contraejemplo.	58
4.4	Una generación del algoritmo genético, paso a paso.	66
4.5	Detalle del procedimiento de cruzamiento uniforme.	68
4.6	Detalle de los procedimientos de mutación.	69
4.7	Sensibilidad de la tasa máxima a variaciones en los parámetros de configuración de la política.	75
4.8	Razón entre las probabilidades de bloqueo experimentadas por las peticiones de traspaso y sus respectivos objetivos.	77



4.9	Incremento relativo de las probabilidades de bloqueo para tras- pasos.	78
4.10	Algoritmo de escalada para determinar la configuración ópti- ma de la política MFGC.	84
4.11	Espacios de soluciones, política IL.	85
4.12	Espacios de soluciones, política GM.	86
4.13	Espacios de soluciones, política MGC.	87
5.1	Dependencia de la probabilidad de bloqueo con los parámetros de configuración de la política MGC.	109
5.2	Operación del esquema adaptativo de CAS.	112
5.3	Operación del esquema adaptativo de CAS (detalle).	113
5.4	Algoritmos de ajuste de la política de CAS.	115
5.5	Diagrama de estados de la CTMC en un escenario con dos clases.116	
5.6	Ejemplo de la probabilidad de bloqueo de traspasos para dos instancias de ejecución distintas del esquema ZL con carga es- tacionaria.	119
5.7	Funcionamiento del esquema adaptativo para peticiones nuevas.122	
5.8	Funcionamiento del esquema adaptativo para peticiones de traspaso.	123
5.9	Algoritmos de ajuste ejecutados por el esquema para traspasos. 124	
5.10	Probabilidad de bloqueo con carga estacionaria.	126
5.11	Otros parámetros con respecto a la carga estacionaria.	127
5.12	Comportamiento transitorio de los esquemas adaptativos en presencia de un incremento repentino en el tráfico ofrecido. . .	131
5.13	Dependencia con k del intervalo de convergencia para P^h	133
5.14	Dependencia con k del intervalo de convergencia para P^n	134



|

Índice de figuras

5.15 Dependencia con Δl^n del intervalo de convergencia para P^h . . . 135

5.16 Dependencia con Δl^n del intervalo de convergencia para P^n . . . 136

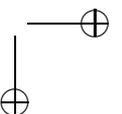
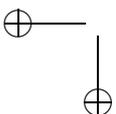
5.17 Variación de P_i con la carga en condiciones estacionarias. 143

5.18 Comparativa de factores de utilización de recursos. 144

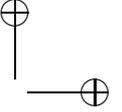
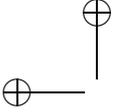
5.19 Comportamiento transitorio de las probabilidades de bloqueo. 145

—

—

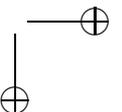
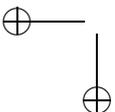


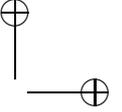
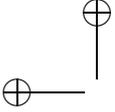
|



Índice de cuadros

3.1	Dimensionado del sistema fijo	44
3.2	Dimensionado del sistema móvil	45
3.3	Impacto de la movilidad sobre el dimensionado del sistema fijo	45
4.1	Parámetros de las configuraciones	72
4.2	Capacidad (λ_{max} en sesiones/s)	72
4.3	Configuración de políticas optimizadas con búsqueda exhaustiva	73
4.4	Comparación del algoritmo HCO y PMC con el algoritmo pro- puesto (<i>our algorithm</i>), en Mflops.	97
5.1	Notación usada para describir el modelo.	107
5.2	Configuración de un escenario monoservicio	108
5.3	Definición de los escenarios estudiados	137
5.4	P_i/B_i del esquema adaptativo para MGC y carga estacionaria igual a la capacidad.	138
5.5	P_i/B_i del esquema adaptativo para MFGC y carga estacionaria igual a la capacidad.	139





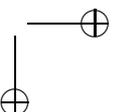
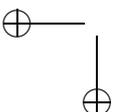
Capítulo 1

Introducción

En el último cuarto de siglo la telefonía móvil celular ha experimentado una progresión formidable. Sólo el dato de los varios miles de millones de terminales móviles que se calcula que existen en la actualidad en todo el mundo, hace palidecer las predicciones pioneras más optimistas de principios de la década de los años 80, cuando se creía que la cifra de un millón de terminales era la cantidad verosímil más grande que se podía anticipar que existiera a finales del siglo XX [RK03]. Esta progresión resulta notable, más aún si se considera que la capacidad que está disponible en el medio de transmisión de la interfaz radio proviene del espectro radioeléctrico, y éste, por sus características propias, es un recurso escaso. De hecho, la actividad de los operadores de este tipo de redes está regulada por los gobiernos mediante autorizaciones que permiten hacer uso de unas porciones específicas del espectro en un área geográfica concreta, y que conllevan asimismo la obligación de no transmitir en ninguna otra banda de frecuencias.

La restricción impuesta por el espectro radioeléctrico es significativa, por ello una cantidad considerable de avances tecnológicos se ha orientado a superar esta limitación. A grandes rasgos, las alternativas son cuatro:

- Incrementar el recurso: tecnologías para el aprovechamiento de las por-



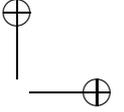
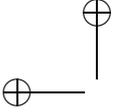
ciones del espectro todavía sin utilizar.

- Incrementar el rendimiento espectral: modulaciones más eficientes, mecanismos más sofisticados de acceso al medio, etc.
- Incrementar el rendimiento geográfico: mecanismos de reutilización del espectro (fragmentación celular, agregación de sectores, etc.), aumento de potencia en las estaciones base, instalación de repetidores, etc.
- Gestionar adecuadamente el recurso: empleo de técnicas de control de admisión de sesiones, planificación de paquetes, control de los traspasos, etc.

Si bien el uso de las tres primeras alternativas hasta los límites impuestos por la tecnología (o por la regulación existente) aumenta la capacidad efectiva de la redes móviles, no es menos cierto que el crecimiento actual de usuarios y el aumento sostenido en las demandas de ancho de banda de unos servicios diversificados justifican la importancia de la gestión eficiente de los recursos, es decir: el uso apropiado de los mismos para la maximización de un objetivo concreto.

El «Control de Admisión de Sesiones» (CAS) constituye un mecanismo substancial de gestión de los recursos radio, a tener en cuenta tanto en el diseño previo («planificación») como durante el funcionamiento propiamente dicho («operación») de las redes móviles celulares. Conceptualmente, este objeto de estudio tiene sus orígenes en la «Red Digital de Servicios de Banda Ancha» (RDSI-BA) y tecnologías correspondientes, tales como *Asynchronous Transfer Mode* (ATM) [Sat94]. Ciertos elementos de estas investigaciones precursoras pueden tener aplicación actual, pero es indudable que características propias de las redes móviles (no contempladas en la RDSI-BA) como la escasez y variabilidad de la cantidad de recursos (consecuencia del comportamiento aleatorio del interfaz radio) y la movilidad de los terminales, aportan una mayor complicación al CAS.

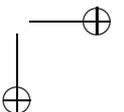
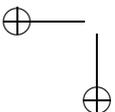
El presente trabajo presta especial atención al impacto que tiene la movilidad en el CAS. De hecho, antes de que el CAS admita una nueva sesión



debe asegurarse que dicha decisión no irá en perjuicio de la propia sesión aceptada ni del resto de sesiones todavía presentes en el sistema. Esto se logra asociando una garantía (probabilística) de que la red tendrá suficientes recursos disponibles en todas las células visitadas por el terminal que inicia la sesión como para respetar tanto los requisitos de «Calidad de Servicio» (QoS) de dicha sesión, como del resto del conjunto de sesiones que estaban ya en curso.

Como en toda economía de escala, es el propio éxito de las redes móviles el que atrae a nuevos clientes a las mismas. Una consecuencia directa del aumento continuo que experimenta la demanda de este tipo de servicios es que los sistemas celulares actuales se diseñan para poder crecer con el tiempo. En la mayoría de ocasiones esto conlleva añadir más estaciones base y por consiguiente disminuir el tamaño de las células [Lee91]. Refinar la fragmentación celular —bien mediante *cell splitting* (particionar aquellas células que cubren la zona donde se localiza el incremento de la demanda) o mediante *overlaid cells* (células que comparten la misma estación base en el centro, pero cubren un área menor)— incrementa la frecuencia con la que se tienen que producir traspasos entre células y asimismo la cantidad de traspasos que suceden durante una sesión dada. El aumento de los traspasos, junto a la variedad de demandas de ancho de banda y de requisitos de QoS exigidos por los servicios emergentes de nueva generación provoca que la ausencia de un mecanismo adecuado de gestión cause una utilización inadecuada de los recursos radio [VC00]. Esta es la principal razón de que se mantenga la atención por esta materia, impresión respaldada además, por el notable volumen de contribuciones que a día de hoy ven la luz en publicaciones especializadas.

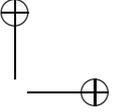
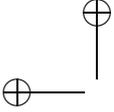
El trabajo contenido en esta tesis trata de aportar en la caracterización, comprensión y desarrollo de mecanismos para la apropiada gestión de los recursos en las redes móviles celulares. En concreto, la primera parte de este esfuerzo comprende el desarrollo de modelos, algoritmos y métodos para estudiar el control de admisión desde una perspectiva estacionaria basada en la optimización. Este es el enfoque tradicional: el que limita el CAS a la fase de planificación de la red, considerando un escenario estacionario (de



acuerdo con las demandas de tráfico anticipadas por el operador) y que da como resultado una política de CAS estática. Este tipo de resultados suele ser de utilidad para intuir tendencias de comportamiento, obtener cotas de prestaciones y mejorar la comprensión y perspectiva del sistema con visos a la posterior simulación o construcción de un prototipo (alternativas a veces prohibitivas). En contraposición al enfoque clásico, durante la fase de operación de la red, el escenario de funcionamiento es «no estacionario» —desafortunadamente— y una política de CAS fija no suele ser la apropiada. La otra motivación de este trabajo persigue proporcionar una alternativa al problema anterior mediante el diseño de esquemas adaptativos de CAS que respondan adecuadamente a la situación específica de la red en un determinado instante de tiempo. La segunda parte de la contribución se centra pues, en el desarrollo de estos esquemas, diseñados para modificar el comportamiento del control de admisión de manera adaptativa en respuesta a las situaciones no estacionarias que se plantean en los sistemas reales.

Estructura de la tesis

La parte final de este capítulo de introducción presenta un breve repaso del progreso de las comunicaciones móviles celulares. En el capítulo 2 se describe el estado del arte en materia de CAS en redes móviles celulares por medio de la bibliografía más relevante, con especial atención al enfoque analítico. En el capítulo 3 se evalúa el efecto que tiene la introducción de la movilidad sobre el dimensionado de las redes móviles y se describe la extensión a entornos móviles multiservicio de una técnica exacta de análisis propuesta inicialmente para sistemas fijos en [Ive87]. En el capítulo 4 se realiza un análisis comparativo de diferentes políticas de CAS en sistemas celulares multiservicio (principalmente en términos de la capacidad de la célula, y de la sensibilidad de la capacidad a las sobrecargas de tráfico y a la precisión con que se estiman los parámetros del sistema). La obtención de resultados se emprende concibiendo el diseño del control de admisión como un problema de optimización. También se desarrolla un algoritmo para optimi-



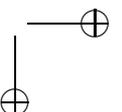
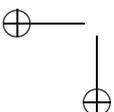
zar ciertas políticas de control de admisión que reduce en varios órdenes de magnitud el coste computacional asociado a la determinación de los valores óptimos de los parámetros con respecto a otros trabajos publicados anteriormente [HUCPOG03c, HUCPOG03a, PMCG05]. En el capítulo 5 se aborda el diseño y evaluación de políticas CAS que adapten sus parámetros de configuración de manera adecuada, en función de las condiciones de tráfico —que se suponen cambiantes con el tiempo—. Aspectos de la propuesta que se valoran comparativamente con soluciones anteriores de la literatura incluyen: velocidad de adaptación a las nuevas condiciones de tráfico, el soporte del multiservicio, y la eficacia con que se respetan los objetivos de QoS exigidos por el operador. Por último, el capítulo 6 recapitula los esfuerzos efectuados, aporta conclusiones, y apunta líneas de trabajo futuro.

Evolución histórica de las comunicaciones móviles celulares

Tanto el concepto celular, como la conciencia de que emplear células para reutilizar el espectro podría incrementar significativamente la capacidad, fueron dos ideas expresadas en 1947 por Don H. Ring de *Bell Laboratories* en trabajos que no llegaron a publicarse.¹ El concepto fue redescubierto en los propios *Bell Telephone Laboratories* en 1966, donde un memorando de Richard Frenkiel y Philip Porter puso de manifiesto que la capacidad del servicio podía crecer enormemente mediante la reutilización de radiocanales de voz. Su documento explicaba que el área a servir se podía dividir en «células», y la interferencia se podía gestionar asegurándose que los usuarios del mismo canal estaban lo suficientemente alejados los unos de los otros. Cuanto más agresivamente se utilizaran los canales, mayor capacidad se podía extraer de las mismas células dado un cierto ancho de banda; la contrapartida era que esquemas más agresivos favorecerían que las llamadas percibieran más interferencias.

En 1979 la operadora japonesa NTT desplegó en la ciudad de Tokio el

¹Bell System Technical Journal, enero (1979); también en *Wireless*, Bell Labs Technology Trends & Development, 4 (2000), no. 1, 1-16.



primer sistema de telefonía celular del mundo [Wes02]. Durante los años siguientes se introdujeron en países del primer mundo diversos sistemas celulares analógicos —la llamada «primera generación» (1G) de sistemas de telefonía celular—, de tal manera que, por ejemplo, hacia finales de 1983, hubo servicios comerciales en Japón, Países Escandinavos y los Estados Unidos. La etiqueta de 1G acabó aplicándose de manera retroactiva a todas aquellas redes celulares con un interfaz aire analógico, principalmente *Advanced Mobile Service* (AMPS) propietario de AT&T. El sistema nórdico de telefonía móvil (desarrollado por Ericsson y Nokia) se desplegó en dos versiones en Escandinavia y en una versión en Francia. Sistemas analógicos, bien similares, bien basados en estas tecnologías, se desplegaron en otras naciones, incluyendo la mencionada Japón, además de Inglaterra y Sudáfrica. Todos los sistemas de primera generación empleaban *frequency division multiple access* (FDMA) porque el resto de tecnologías de acceso múltiple requerían una transmisión digital. Un problema común a todos estos sistemas fue su mútua incompatibilidad. A pesar de ello, contribuyeron a perfeccionar los fundamentos del diseño eficaz de células.

El siguiente paso significativo lo constituyó la aplicación de la tecnología digital a la transmisión telefónica de la voz —redes celulares de «segunda generación» (2G)—. En Europa se creó en 1982 el *Groupe Spécial Mobile* (GSM) con la intención de desarrollar un sistema paneuropeo de telefonía celular. En 1989 este esfuerzo dio lugar a la publicación del estándar GSM (*Global System for Mobility*) por parte del *European Telecommunications Standards Institute* (ETSI), empleando *Time Division Multiple Access* (TDMA), con la esperanza de que se convirtiera, algún día, en un estándar mundial. En Norteamérica, los servicios denominados de *Personal Communications Services* (PCS) se difundieron primero a nivel regional y luego a nivel nacional, proporcionándose el servicio a través de una mezcla competitiva de tecnologías TDMA —*Digital AMPS*, (D-AMPS) y *Digital Cellular System*, (DCS-1900) una variedad de GSM para la banda de 1900 MHz— y sobre todo *Code Division Multiple Access* (CDMA), en concreto, *cdmaOne* (estándar IS-95 de ANSI). Los países que adoptaron redes celulares digitales optaron por seguir bien

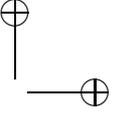
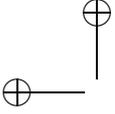
el modelo norteamericano de CDMA o el europeo de GSM, en virtud de las áreas de influencia de cada uno de los estándares. Japón aparte de CDMA, adicionalmente tuvo su propio estándar, *Personal Digital Cellular* (PDC).

GSM se lanzó comercialmente entre 1991 y 1992 [VLLX02] y es el ejemplo más exitoso de todos los sistemas 2G que crecieron y se consolidaron aceleradamente a lo largo de la década de los noventa. Su servicio portador de datos original mediante conmutación de circuitos *Circuit Switched Data*, (CSD), fue con posterioridad complementado con servicios de datos adicionales introducidos en la interfaz radio, mientras se esperaba la llegada de la siguiente generación de telefonía celular. Estos estándares fueron *High-Speed Circuit-Switched Data* (HSCSD) también en conmutación de circuitos o *General Packet Radio Service* (GPRS) ya en conmutación de paquetes, a parte de las mejoras en la tasa binaria introducidas en la capa física con *Enhanced Data rates for Global Evolution* (EDGE) [VLLX02, Wes02].

Hacia 2002 comenzaron los primeros despliegues a gran escala de la «tercera generación» de telefonía celular (3G). La iniciativa 3G y sus estándares nacieron durante la década de los noventa [MK00, ZJB00] con dos objetivos: añadir la capacidad necesaria para transportar datos de usuario a velocidades elevadas a través de redes celulares —sin menoscabo del tráfico de tiempo real—; y unificar el mundo bajo un único estándar que permitiera la movilidad global [Wes02, VLLX02]. Esta última meta, sin embargo, no se consiguió porque la *International Union of Telecommunications* (ITU) carecía de la influencia necesaria para persuadir a los gobiernos nacionales a implantar un estándar común. El resultado fue la aprobación de cinco estándares diferentes que cumplían con los requisitos generales de las comunicaciones de 3G. El conjunto de estándares se denominó IMT-2000 (de *International Mobile Telecommunications*). El sucesor de GSM es *wideband CDMA* (W-CDMA), —estándares IMT-DS (*Direct Spread*) también conocido como *W-CDMA Frequency Division Duplex* (FDD), e IMT-TC (*Time Code*) también conocido como *W-CDMA Time Division Duplex* (TDD)—. El sucesor de cdmaOne es cdma2000, una iniciativa del *CDMA Development Group* (CDG) que consiguió su aprobación como estándar conforme a IMT-2000 por parte de la ITU. Junto

a ellos, *Time Division Synchronized CDMA* (TD-SCDMA) —que forma parte del propio estándar IMT-TC— necesita mención, puesto que es el estándar desplegado en China.

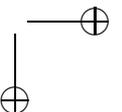
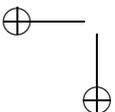
La iniciativa de «cuarta generación» (4G) surgió a principios de la década de los noventa, mucho antes de que las mismas tecnologías de 3G fueran estandarizadas [RK03]. En la actualidad, la 4G pretende ser un conjunto plural de redes de acceso inalámbricas con acceso global e integrado a Internet. Los estándares han ido cambiando, y del empleo inicial de *Asynchronous Transfer Mode* (ATM), el gran éxito de Internet ha justificado que los esfuerzos se centren sobre *Internet Protocol version 6* (IPv6). El concepto de 4G incluye ciertas posibilidades motivadas por nuevas tecnologías de red de área local con acceso inalámbrico (WLAN), como estaciones base en emplazamientos estratégicos [Sal04] con prestaciones en el rango de los centenares de Mbps (quizás Gbps), o montadas sobre trenes o camiones móviles. Todo ello para conseguir el objetivo de que el usuario final de Internet acceda en cualquier momento a esta red con la tecnología (inalámbrica o no) más adecuada para el caso, (paradigma de *Always Best Connected*) [GJ03].



Capítulo 2

El estudio del control de admisión en redes celulares

Las aportaciones previas en materia de gestión de recursos en redes celulares son tan cuantiosas que un compendio completo de la bibliografía sería una tarea inabordable (no obstante véase en este sentido los trabajos recopilatorios de [Ahm05, GB06]). En su lugar, este capítulo procura realizar un repaso de aquellas contribuciones más relevantes, con especial atención a las que constituyen el actual estado del arte en el campo del Control de Admisión (CA). La selección de referencias se ha realizado dando preferencia a aquellos estudios que optan por una caracterización formal del modelo empleado y su posterior resolución analítica. Dos partes conforman este capítulo, una en la que se revisan las aportaciones con respecto a los modelos empleados y sus suposiciones, y otra en la que se detallan los diversos esquemas propuestos para llevar a cabo el CA. Esta división obedece a criterios puramente prácticos —dar una mayor cobertura a los contenidos más vinculados al objeto de la presente tesis— y no pretende ser una clasificación ideal ni óptima.



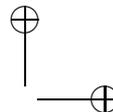
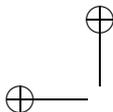
2.1 Modelo del sistema

El sistema que se aborda en este estudio consta de un conjunto de particularidades esenciales que son el fundamento común de todo modelo que se ha propuesto para el mismo:

1. La región geográfica servida por la red está dividida en áreas de transmisión elementales denominadas «células» donde todos los terminales reciben servicio de una estación base vía radio. La existencia de múltiples células permite la reutilización de las frecuencias y resulta en una mayor capacidad en comparación con una única zona de servicio cubierta con los mismos recursos radio.
2. Los terminales experimentan lapsos de actividad y reposo. A un intervalo de actividad se le denomina «sesión» y se identifica con el período de tiempo en el que se mantienen ocupados una cierta cantidad de recursos de la interfaz radio.¹ Para las redes de «conmutación de circuitos», una sesión se corresponde con una conexión y por tanto con la asignación y reserva de una cantidad fija y exclusiva de recursos durante todo el tiempo que dure la misma. En las redes de «conmutación de paquetes» una sesión se corresponde, en términos imprecisos, con el lapso de tiempo durante el que se produce un envío prolongado de paquetes entre dos terminales en comunicación. En este tipo de redes, sin embargo, los terminales sólo reciben recursos cuando disponen de datos a enviar; de hecho, la transmisión de datos de manera intermitente o a ráfagas requiere que los recursos empleados por parte de la sesión se interpreten desde un punto de vista estadístico, de manera semejante a conceptos como el «ancho de banda efectivo»² [GAN91, EM93, EE99].

¹En esta tesis se considera exclusivamente tráfico compuesto por sesiones de *streaming* que requieren una mínima tasa de transferencia (unidades de recurso) para funcionar adecuadamente así como requisitos temporales como retardo acotado y variación del retardo (*jitter*) acotado.

²El «ancho de banda efectivo» es una noción que trata de precisar atributos propios de las fuentes de tráfico, incluida la caracterización de ráfaga —*burstiness*— y los requisitos de servicio.



3. Los terminales son móviles; debido a esto, cualquier terminal —con sesión en curso o no— puede en un determinado momento cambiar de una célula a otra. Al objeto de evitar la interrupción de la comunicación en curso existe un solapamiento entre las zonas de cobertura de células adyacentes. Asimismo se asume un mecanismo de «traspaso» (*handover*, *handoff*) de la sesión en progreso, de una célula, sector o portadora a otra.
4. La cantidad de recursos radio en cada célula es una cantidad constante *Fixed Channel Allocation* (FCA). Aunque se han propuesto muchos algoritmos de asignación dinámica de recursos, *Dynamic Channel Allocation* (DCA) que mejoran la gestión eficiente de los recursos³, por razones prácticas —la mayor complejidad asociada a las estrategias DCA— en la práctica se emplea mayoritariamente FCA.

2.1.1 Modelos de colas

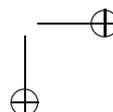
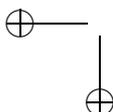
Habitualmente, cuando se le presenta una petición a la célula, el CA debe determinar si la solicitud se aceptará o bloqueará. La clase de sistemas estudiados da lugar a una variedad de modelos de colas en función del tratamiento (ver figura 2.1) que se aplique a las peticiones que se aceptan y se bloquean⁴ y el comportamiento subsiguiente de las mismas.

A las peticiones aceptadas se les presentan dos alternativas: que el sistema disponga de suficientes recursos como para satisfacer las demandas

Este concepto, además, está estrechamente conectado al control de admisión y a los requisitos de servicio de cada clase de tráfico. En [EM93] se demuestra que el ancho de banda efectivo de una fuente está acotado entre dos términos: la tasa de pico y la tasa media —dos criterios extremos en los que se basaron muchos diseños de la «red digital de servicios integrados de banda ancha» (*broadband ISDN*)—. La utilidad principal del «ancho de banda efectivo» es que permite trasladar metodologías de diseño de las redes de conmutación de circuitos para su aplicación a las de conmutación de paquetes.

³Una introducción al tema se puede encontrar en [TJ91]. Una minuciosa recopilación y comparación de esquemas estáticos y dinámicos de asignación de canales figura en [KN96].

⁴Son peticiones «bloqueadas» todas aquellas que no se aceptan de inmediato tras su llegada.



en términos de recursos de la petición —en cuyo caso se acepta la petición normalmente— o lo contrario —en cuyo caso, dado que se ha aceptado la petición, se requiere acomodarla de algún modo dentro del sistema—. Existen dos maneras de realizarlo: una es liberar los recursos suficientes, bien optando por desalojar sesiones en curso —quizás de menor prioridad— o bien degradando las demandas de aquellas sesiones en progreso que así lo admitan —porque se trate de tráfico de tiempo real pero de tasa adaptativa, o bien de tráfico elástico de datos—; la otra manera es degradar las demandas de la propia petición a aceptar, si es que existe la posibilidad.

Con respecto a las peticiones bloqueadas tienen dos alternativas: o la de ser rechazadas directamente o la de pasar a una cola en la que esperen hasta que queden libres los recursos necesarios para su admisión. En el primer caso —sin cola de espera—, la petición rechazada puede bien producir uno o varios reintentos tras un intervalo de tiempo, o bien anular su intención de acceder al sistema y finalmente desistir. En el segundo caso —con cola de espera—, puede verificarse que bien esta espera se dilate durante tanto tiempo como sea necesario hasta que finalmente se admita la misma (cliente con paciencia infinita), o bien que el tiempo de espera que esté dispuesto a pasar el cliente no sea indefinido con lo que finalmente —si no se admite la petición con la suficiente prontitud— el cliente abandonará la cola y el sistema (cliente con paciencia finita).

Todas las alternativas enumeradas y sus combinaciones se han estudiado en la literatura concerniente al CA en sistemas celulares. Con diferencia, el enfoque más practicado —y manejable— es optar, ante un bloqueo, por el rechazo de la petición y posterior desistimiento de la misma. Este enfoque es común hasta tal punto que el grueso de contribuciones en la materia lo incluyen en una u otra manera (ver los compendios de [Ahm05, GB06]). Además, es el escenario inicialmente supuesto para el desarrollo de esta tesis.

La alternativa de emplear una cola para las peticiones bloqueadas también ha recibido abundante atención, principalmente como mecanismo para dar prioridad a las peticiones de traspaso frente a las peticiones de sesiones nue-

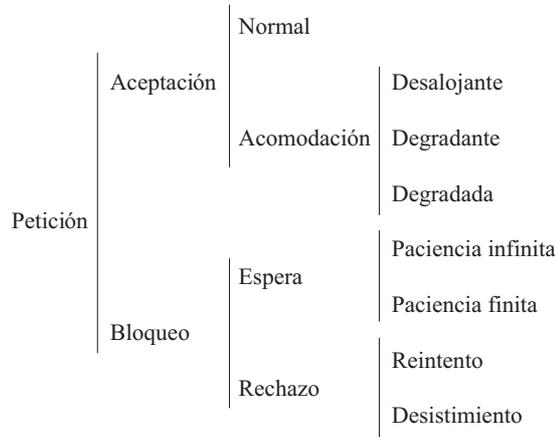


Figura 2.1: Alternativas consideradas en CA en sistemas celulares.

vas (ver [PG85, HR86, TJ91, TJ92, KN96, TRV98, PCG02, LJCP03, XT04] y las referencias que citan). En cuanto al comportamiento de los clientes en la cola, algunos trabajos [TJ92, ET99, DRFG99a, DRFG99b, Fan00, LJCP03, XT04] vinculan la paciencia finita de los traspasos bloqueados y puestos en cola con el tiempo de permanencia del móvil en el área de solape, y en virtud de ello tratan de servir antes a aquellos clientes que por su posición están próximos a abandonar este área. Asimismo, el encolado de peticiones de establecimiento de nuevas sesiones se ha planteado como solución para aumentar el tráfico cursado total [Gué88, DJ92]. Otras contribuciones estudian escenarios donde se encolan ambos tipos de peticiones [CSC94, McM95, LJCP03]. Finalmente, en modelos de sistemas con dos clases de tráfico (voz y datos, identificables respectivamente con tráfico con y sin requisitos de tiempo real), se proponen colas distintas según la clase de tráfico a la que pertenezcan [ZA02, WZA03]; también se sopesa la admisión forzosa de una sesión de voz traspasada, bien mediante el desalojo (*preemption*) de una sesión de datos, o bien mediante la toma de una parte del ancho de banda asignado al tráfico de datos [SW04], es decir: los requisitos de ancho de banda de las sesiones de datos son elásticos.

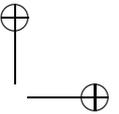
Ejemplos de modelado del fenómeno de los reintentos de las peticiones

de sesiones nuevas son [TGM97, AL02]. El estudio de un sistema con reintentos de peticiones de sesiones nuevas y traspasos se aborda en [MDCL⁺01, DBGGMBCG05], donde la complejidad del modelo construido obliga a una aproximación posterior para poder llevar a cabo el análisis.

2.1.2 Suposiciones e hipótesis del modelo

Consideraciones del entramado celular

Dos premisas que se asumen frecuentemente en la investigación del CA en sistemas celulares son: 1) todas las células se encuentran en equilibrio estadístico y su comportamiento es independiente entre sí; 2) la cantidad de recursos de que dispone cada célula y el tráfico ofrecido a las mismas es homogéneo en toda la región geográfica servida, por lo que las tasas de generación de sesiones nuevas y de traspasos son equivalentes en todas las células [HR86]. La ventaja substancial de estas hipótesis es que permiten analizar el sistema a partir del comportamiento de una única célula aislada, con lo que los modelos matemáticos planteados son más tratables. No obstante, esta aproximación puede ser inadecuada cuando existen, por ejemplo, regiones donde se concentra el tráfico (denominados *hot-spots*, literalmente, «puntos calientes»), para los cuales, un modelo de múltiples células sería más conveniente [MS01]. Como el interés de suponer más de una célula se ve contrarrestado por la complejidad de guardar los detalles acerca del estado de cada célula considerada, se han planteado soluciones a medio camino entre el modelo de célula aislada y los modelos de múltiples células. Una alternativa frecuente consiste en relajar la segunda premisa, es decir: suponer un escenario de células heterogéneas (en términos de tráfico y recursos), pero en el que el proceso de llegada de traspasos a una célula sólo depende del caudal medio de traspasos procedente de las células adyacentes a la objeto de estudio —y no del estado específico en que puedan encontrarse esas células en un determinado momento—. Es práctica usual hallar estas tasas de traspaso mediante un método de búsqueda local basado en iteraciones de punto fijo [McM91, MS01].



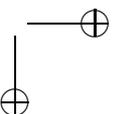
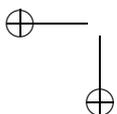
Sin embargo es discutible la precisión que estos enfoques híbridos aportan en aquellas situaciones en las que el flujo medio de traspasos no da la suficiente información acerca del comportamiento de los terminales móviles —como por ejemplo, cuando movilidad de los propios terminales es elevada [MS01]—.

Teletráfico y variables aleatorias

Por su naturaleza estocástica, ciertos parámetros de las redes celulares se suelen representar mediante variables aleatorias en los modelos que se formulan para su estudio. Estos parámetros incluyen entre otros: el intervalo de tiempo entre las peticiones de conexión de un tipo y servicio, duración de la sesión, duración restante de una sesión tras un traspaso, tiempo de residencia en una célula, tiempo de ocupación de recursos en una célula por sesión. . . En los estudios con un enfoque analítico resulta importante que las distribuciones escogidas para caracterizar estas variables aleatorias produzcan modelos abordables matemáticamente. De ahí que por sus atributos simplificadores, lo habitual sea suponer una distribución exponencial para la mayoría de estas magnitudes.⁵ Sin embargo —como con cualquier aproximación— la verosimilitud de dicha elección depende principalmente de lo factible y ajustada que sea la distribución respecto a la realidad. Aunque mediciones sobre redes reales parecen invalidar las hipótesis simplificadoras [BJ00, JL96], en torno a esta cuestión no existen en la literatura conclusiones definitivas (ver [RT01] y referencias incluidas en él), por lo que la alternativa es conocer los argumentos emitidos a favor y en contra de cada aproximación.

Llegada de peticiones de nuevas sesiones. La práctica habitual consiste en emplear una aproximación de población infinita que permita prescindir del número de terminales presentes en la célula en un determinado momento.

⁵Otra razón es la influencia ejercida por enfoques clásicos como los de [HR86, McM91], pioneros en emplear procesos de llegada de Poisson y tiempos de ocupación de recursos distribuidos exponencialmente.

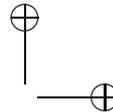
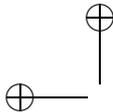


El proceso de llegada así caracterizado es de Poisson (tiempo entre llegadas exponencial). Pocos estudios critican la validez de esta aproximación en sistemas móviles celulares: Barceló y Sánchez en [BS99] afirman que el proceso conjunto de peticiones de sesiones nuevas y traspasos concuerda con una distribución más suave⁶ que Poisson; en el modelo de Alfa y Li en [AL02] el proceso de llegadas es un *Markovian Arrival Process* (MAP), considerablemente más general que un proceso de Poisson y cuya complejidad de cálculo explica la ausencia de resultados numéricos en el artículo.

Llegada de peticiones de traspaso. La hipótesis de usar un proceso de Poisson para modelar este parámetro ha sido más discutida. Esto es debido a que existe una relación de dependencia entre el proceso de llegada de los traspasos y las suposiciones empleadas, respectivamente, para el proceso de llegadas de las peticiones de sesiones nuevas y para el tiempo de residencia en una célula. No obstante, los resultados de la literatura no son plenamente coincidentes entre todos los estudios. Chlebus y Ludwin [CL95] pusieron de manifiesto que aunque la llegada de peticiones de sesiones nuevas siga un proceso de Poisson, el bloqueo que experimentan estas peticiones distorsiona el proceso concurrente de generación de traspasos a las células vecinas, convirtiéndolo en más suave que Poisson. A pesar de ello, constatan que la hipótesis de aproximar mediante de un proceso de Poisson es sensata para sistemas en baja o media carga. Sidi y Starobinski [SS96, SS97] justifican el uso de un proceso de Poisson con entramados celulares extensos y tráfico ofrecido homogéneo, y lo refutan en redes de pocas células con tráfico heterogéneo. Rajaratnam y Takawira [RT00, RT01] y Doorn y Ta [vDT03] coinciden con [CL95] en la suavidad del tráfico de traspasos y añaden la movilidad alta a la lista de situaciones en las que considerar un proceso de Poisson podría ser inadecuado. En contraposición a [CL95], Orlik y Rappaport [OR01] analizan tres escenarios de complejidad decreciente (mosaico T_1 ⁷ con traspasos

⁶Es decir, $CV < 1$, donde CV es el coeficiente de variación de una variable aleatoria X , definido como $CV_X = \sigma_X / E[X]$.

⁷El mosaico T_1 consta de una célula hexagonal central y un anillo de otras seis células a su alrededor. El mosaico de tipo T se forma añadiendo anillos concéntricos de células en torno a

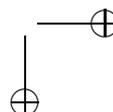
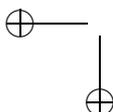


desde la periferia según un proceso de Poisson, célula aislada con traspasos según un *Markov-Modulated Poisson Process* (MMPP) equivalente, y célula aislada con traspasos según un proceso de Poisson); y concluyen que no hay discrepancias manifiestas que invaliden la aproximación, especialmente en alta carga. Del trabajo de Zeng y Chlamtac [ZC99] se desprende que un modelo verosímil de la distribución del tiempo de residencia en una célula (distribución gamma con alta varianza) desautoriza la aproximación de Poisson para el proceso de llegada de traspasos. Finalmente, en [DTL03] se analizan conjuntamente las consecuencias de modelar el tiempo entre llegada de traspasos con una distribución Erlang, hiperexponencial de varianza elevada y exponencial (tráfico de Poisson) de tasa comparable, y se concluye que la distribución Erlang no arroja diferencias significativas respecto a la distribución exponencial, —al contrario que con la distribución hiperexponencial—.

Tiempo de ocupación de los recursos por la sesión. La distribución exponencial es una aproximación apoyada por Hong y Rappaport [HR86] y Guérin [Gué87] para células circulares. El estudio de [ZD97] también confirma este enfoque cuando la duración de la sesión está distribuida exponencialmente. Esta práctica contrasta con la preferencia en estudios basados en medidas [JL96, BJ00] por otros tipos de distribuciones (log-normales y combinaciones de las mismas) para modelar el tiempo de ocupación de los recursos. El escrutinio experimental de Hidaka et al [HSSK01, HSSK02], constató que este parámetro posee rasgos de autosemejanza. Diversos trabajos posteriores han revisado el impacto de la hipótesis exponencial en las prestaciones de los modelos con respecto a las otras opciones, pero sin unanimidad: [KZ97, XT03] concluyen adecuada la hipótesis, mientras que [LC97, HSSK02] señalan circunstancias en las que se verifica una discrepancia manifiesta.

Tiempo de residencia en una célula. Los estudios de Hidaka et al. [HSSK01, HSSK02] también manifiestan el carácter autosemejante de esta magnitud.

una célula inicial. El subíndice indica el número de anillos alrededor del mosaico de partida. Por convención, T_0 está formado por una célula.

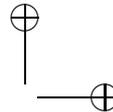
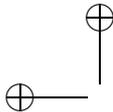


Schweigel [Sch03] halla una distribución para este parámetro considerando células rectangulares y la compara con la aproximación exponencial. Las simulaciones más genéricas de Zonoozi y Dassanayake [ZD97] muestran el ajuste del tiempo de residencia en una célula a una distribución gamma generalizada.

2.2 Control de admisión en redes celulares

Una política de Control de Admisión (CA) es un conjunto de directrices que gobiernan la decisión de aceptar una petición o no dentro de un sistema de telecomunicaciones. La mayoría de contribuciones de carácter analítico realizadas en este terreno emplean un enfoque para las políticas de CA orientado a la carga, es decir: la admisión se basa bien en la cantidad de sesiones en curso o bien en el factor de utilización de los recursos, «*number based admission control*» [Ahm05]. Esta preferencia obedece principalmente a la superior tratabilidad matemática que se deriva de los modelos y a que su posterior implementación práctica resulta más simple. No obstante, este enfoque, al que se recurre habitualmente en sistemas donde la capacidad está limitada por bloqueo, también se ha considerado una buena aproximación para aquellos sistemas en los que la capacidad está limitada por interferencia (ver por ejemplo [KFZK04] y las referencias que contiene). Esta sección, en concreto, se encarga de destacar, dentro de este enfoque, aquellas políticas de Control de Admisión de Sesiones (CAS) en redes móviles celulares más significativas de la literatura con respecto a tres criterios:

- La información con la que las políticas toman sus decisiones.
- El principio de funcionamiento propio de cada familia de políticas.
- Los procedimientos aplicados para obtener la política más adecuada.



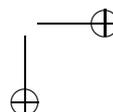
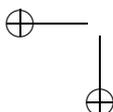
2.2.1 Información a disposición de la política

Casi sin excepción, las políticas de CAS planteadas en la literatura hacen uso del estado de la célula en la que se ejecutan para formular sus decisiones de admisión o rechazo en un momento concreto. La información de estado habitualmente procede del número de sesiones en progreso o de recursos ocupados en la célula objeto de interés, pormenorizado por servicio y tipo de petición o resumido en un valor agregado. Parece razonable que si se aumenta la cantidad de información a disposición del CAS se pueden mejorar las prestaciones de la política —a costa de incrementar la complejidad de la implementación—. En esta línea, el grueso de iniciativas se centran en obtener los suficientes indicios como para anticipar la llegada de las peticiones más prioritarias —los traspasos— antes de que se lleven a cabo. Estos elementos de juicio pueden provenir del estado del vecindario celular: en [NS96] y [WWL02] se predice así la probabilidad de bloqueo a una cantidad de tiempo vista; o del itinerario histórico de los terminales y el trazado urbanístico: en [LAN97, CB00, YL02, CS02, SK04] se predice de esta manera el instante temporal y la célula destino sobre la que se producirá un traspaso.

2.2.2 Familias de políticas

La diferencia más substancial entre las políticas de CAS que se proponen en la literatura es el empleo que se hace de la información que tiene a su disposición el proceso de decisión.

La política de CAS más simple es la ausencia de la misma, *Complete Sharing* (CS), consistente en admitir una petición siempre que hayan suficientes recursos disponibles en el sistema. En [HR86] esta política se analiza para un solo servicio y tiene el inconveniente de que trata de la misma forma a las sesiones nuevas y a los traspasos cuando, en general, resulta más inapropiado que se interrumpa una sesión en curso. Esta desventaja es ya significativa en sistemas multiservicio donde una política de CS puede implicar: 1) que ciertas clases de peticiones monopolicen los recursos; 2) un trato similar para

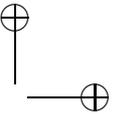
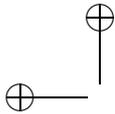


los servicios de valor añadido y los de menor importancia —lo que acarrea un deficiente rendimiento económico de los recursos y el correspondiente impacto en los beneficios obtenidos por el operador—. Los inconvenientes de CS justifican la abundancia de políticas de CAS en la literatura. Estas contribuciones se pueden clasificar en base al significado y propiedades del criterio de aceptación empleado.

Políticas de *trunk reservation*

La familia de políticas reunidas colectivamente bajo el nombre de *trunk reservation* (TR), *guard channel* (GC), «reserva de canales» o «reserva de recursos» constituyen el grupo más empleado en la literatura. La característica común a todas ellas es que la decisión de aceptar una petición está supeditada a que el número de recursos libres sea mayor que cierta cuota definida previamente —que actúa a modo de reserva explícita (pero no exclusiva)—. La política GC —también conocida como *Static Guard Channel* o *Conventional Guard Channel*— se introdujo en [PG85, HR86] para redes móviles celulares y limitaba la reserva a múltiplos enteros de la unidad de recursos elemental (el canal). Posteriormente, la política *Fractional Guard Channel* (FGC) [RTN97] —también conocida como *New Call Thinning Scheme* o *Special Guard Channel*— extendió el GC a fracciones de canal para aumentar las prestaciones. Ejemplos de contribuciones al GC y FGC en sistemas monoservicio son [McM95, Sch90, CH96, CG01, KA99, PC01, DS02, DS04]. Generalizaciones de estas políticas a sistemas multiservicio se plantean en [CC97, LLC98] para GC —denominada *Multiple Guard Channel* (MGC)— y en [HUCPOG03a] para FGC —conocida como *Multiple Fractional Guard Channel* (MFGC)—.

Respecto a las prestaciones de estas políticas, para un escenario monoservicio, se demuestra en [RTN97, Bar01] que GC y FGC son óptimas para funciones comunes de la QoS objetivo. Un escenario multiservicio se estudia bajo ciertos criterios en [AJK01], donde usando un modelo fluido la política de admisión óptima que se halla también es de la clase *trunk reservation*; sin embargo sus resultados no son generalizables. Estas conclusiones implican



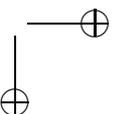
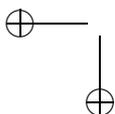
que la capacidad que puede lograrse en las redes celulares con el uso de esta familia de políticas es superior a la que se consigue con la mayoría del resto de familias (exceptuando las políticas estacionarias aleatorizadas de la sección 2.2.2). No obstante, la superioridad en el tráfico cursado se obtiene para valores estacionarios de los parámetros de teletráfico y a costa de un mayor coste computacional en el proceso de resolución de estas políticas. Precisamente este último inconveniente ha suscitado la proposición de algoritmos aproximados que disminuyen el coste del cálculo necesario para trabajar con estas políticas [BS97, BM98, CPVAOG04].

Políticas con solución en forma de producto

La particularidad propia de esta familia de políticas es que producen una solución para las probabilidades de estado del sistema completo que es el producto normalizado de las probabilidades de estado de cada clase de petición implicada, considerada aisladamente —*product form solution* (PFS)— [Nel95]. Un subconjunto grande e importante de esta familia es la de las políticas «convexas en coordenadas» —*coordinate convex* (CC)—. En [Aei78] se muestra que la solución en forma de producto es válida para todas las políticas convexas en coordenadas. El recíproco no es cierto: por ejemplo en [Kau81] se logra obtener una solución en forma de producto para un sistema con una distribución para el tiempo de ocupación de recursos con transformada de Laplace racional y el resultado es válido para cualquier política arbitraria (y no solamente para las convexas en coordenadas).

Una política convexa en coordenadas se define por el «conjunto convexo en coordenadas»⁸ de estados posibles a los que está limitada. En este tipo de políticas se acepta una petición si y sólo si después de aceptar la petición el sistema permanece dentro del conjunto permitido de estados. Para las políticas de la familia *trunk reservation* esta condición no se cumple porque

⁸Sea $S := \{(x_1, \dots, x_N) : x_i \in \mathbb{N}, i = 1, \dots, N\}$ se dice que $\Omega \subseteq S$ ($\Omega \neq \emptyset$) es un conjunto convexo en coordenadas si cumple la propiedad siguiente: si $(x_1, \dots, x_i, \dots, x_N) \in \Omega$ y $x_i > 0$, entonces $(x_1, \dots, x_i - 1, \dots, x_N) \in \Omega$ [Ros95].

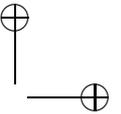
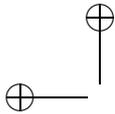


ciertas transiciones entre estados permitidos están prohibidas. Así, a excepción del caso monoservicio, las políticas del tipo *trunk reservation* carecen de una solución en forma de producto.

El beneficio de la solución en forma de producto se aprecia si se tiene en cuenta que para obtener las probabilidades de estado en equilibrio de un determinado proceso de Markov bajo una política concreta es necesario resolver las ecuaciones de balance global. En general se recurre a procedimientos de resolución aproximada de ecuaciones lineales (por ejemplo iteraciones de Gauss-Seidel) de escasa escalabilidad incluso a sistemas con un número modesto de servicios o recursos [CC97]. En contraste, la presencia de una solución en forma de producto permite aplicar métodos analíticos eficaces que disminuyen decisivamente la complejidad de estudio característica de estos sistemas (ver [Ros95] para un análisis comparativo de los mismos). A todo esto hay que añadir la ventaja de que, en presencia de políticas con solución en forma de producto, los modelos considerados son insensibles al tipo de distribución considerada para la variable aleatoria que modela el tiempo de utilización de los recursos, siendo solamente de importancia su valor medio [Ive87].

Dentro de las políticas convexas en coordenadas existen dos grandes subfamilias con puntos en común: 1) las de «reparto» —*partitioning policies* (PP)— en las que se emplea la adjudicación de porciones de recurso a las diferentes clases de peticiones; 2) las de «umbral» —*threshold policies* (TP)— en las que se hace uso de una limitación en la cantidad de sesiones que cada clase de petición puede tener en curso en el sistema. Jordan y Varaiya en [JV94] identifican la política convexa en coordenadas óptima.

Ejemplos de políticas *threshold* son: *Integer Limit* (IL) [Ive87] —también llamada *Upper Limit* (UL) [CLW95]—, que maneja cotas superiores para el número de sesiones en curso admisibles para cada clase de petición y para cuya resolución eficiente existe un «algoritmo de convolución» [Ive87, Ive02]; *Fractional Limit* (FL), analizada en [LA95] donde se discuten técnicas para definir políticas similares a IL, extendiéndolas mediante el uso de cotas fracciona-

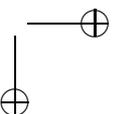
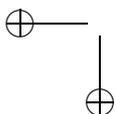


rias. Ejemplos de políticas *partitioning* son: *Complete Partitioning* (CP) [RT89] que descompone los recursos en subconjuntos dedicados a cada clase de petición; *Guaranteed Minimum* (GM), estudiada en [CLW95] reserva porciones de los recursos en exclusiva para clase y adicionalmente destina el resto de recursos para uso común y compartido para todas aquellas clases que hayan agotado su reserva.

Las prestaciones de la familia de políticas con solución en forma de producto respecto a la familia de políticas del tipo *trunk reservation* se ha estudiado en [BS97] y [GMP04]. Ambos trabajos coinciden en que existe un compromiso entre la reducción de complejidad computacional que ofrecen las políticas con solución en forma de producto y las prestaciones en términos de capacidad (inferiores a las que ofrecen las políticas de la familia *trunk reservation*). En [GMP04] la disminución de capacidad ronda el 5–15%. En [BS97] se conciben escenarios diseñados específicamente para que las ventajas de emplear políticas *trunk reservation* sean considerablemente superiores (clases de peticiones con una importante asimetría mutua en términos tanto de recursos demandados por sesión, tasa de llegadas y probabilidades de bloqueo objetivo). En situaciones de sobrecarga substancial, el trabajo de [BF01] realiza un análisis comparativo de diversas políticas concluyendo que la política convexa en coordenadas *Upper Limit* es la más deseable.

Políticas estacionarias aleatorizadas

Aunque las políticas de la familia *trunk reservation* consiguen unas prestaciones significativas en la mayoría de contextos, en general la política estacionaria óptima en el caso multiservicio [BC02, PCG04] no es del tipo *trunk reservation*. En estos escenarios la política óptima pertenece a una superfamilia de políticas denominadas «estacionarias aleatorizadas» —*randomized stationary* (RS)—, donde la decisión de admitir un tipo de petición depende del estado actual del sistema —pormenorizado por tipo de petición en lugar de considerar la cantidad total de recursos en uso como en la políticas *trunk reservation*— y (quizás) de un elemento aleatorio [Ros70].



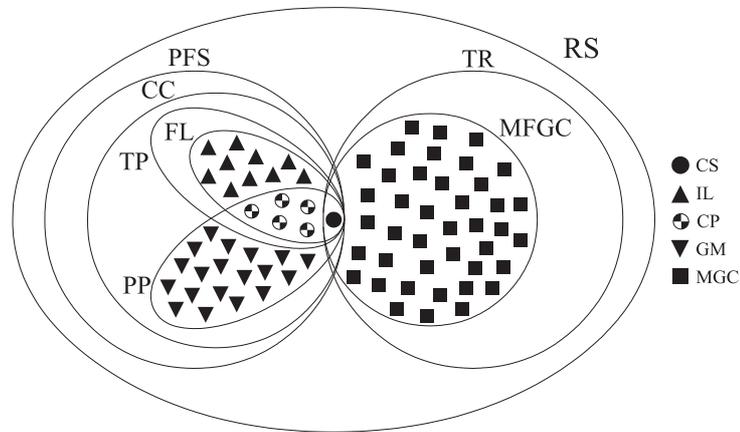
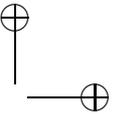
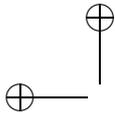


Figura 2.2: Familias de políticas de control de admisión.

A modo de resumen, la figura 2.2 muestra las familias de políticas de control de admisión mencionadas por medio de un diagrama de Euler-Venn.

2.2.3 Procedimientos para el diseño de políticas

En las redes celulares la selección de la política de CAS más adecuada depende de la labor de intermediación que se exija a la política —igualar el acceso a recursos que tienen todas las clases, aumentar la utilización global de los recursos, o asignar parte de estos recursos a las clases de mayor prioridad—. En cualquier caso debe precisarse una metodología que respalde la utilidad de elegir una política de CAS frente a las otras alternativas posibles. La perspectiva usual suele pasar por plantear una política de CAS novedosa y confirmar que sus prestaciones son superiores a soluciones previamente establecidas en la literatura. Una propuesta más profunda e infrecuente consiste en formular el proceso de decisión del CAS como un problema de optimización, es decir: hallar la decisión óptima (aceptación o rechazo, quizás dependiente de un elemento aleatorio) que el CAS debe tomar para cada posible situación del sistema (estado actual del mismo y tal vez la historia anterior) con el pro-



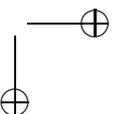
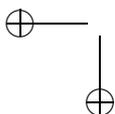
pósito de maximizar una función objetivo (equidad, utilización, caudal...) que representa a los parámetros de prestaciones de interés en el problema, todo ello sujeto al conjunto de restricciones que definen el propio sistema (cantidad de recursos, etc.).

Una herramienta frecuente en la optimización de políticas de CAS es el «proceso de decisión de Markov» —*Markov Decision Process* (MDP) [Ros70]—. Aunque este enfoque cuenta con un marco teórico relativamente bien establecido, existen funciones a optimizar y políticas para las que los MDPs no son aplicables; en dichas situaciones se debe acudir a otras alternativas habituales de la literatura de optimización que aporten mayor flexibilidad a la resolución del problema.

La optimización de la política de CAS respecto a diferentes criterios se ha empleado en escenarios monoservicio en [RTN97, Bar01, YR97], en este último en vista de la complejidad del problema recurre a un algoritmo genético para hallar una política subóptima. En escenarios multiservicio se ha recurrido a *linear programming* para solucionar el MDP [BC02, PCG04, XCW01]. En estos escenarios, la penalización más significativa la entraña la «dimensionalidad» (*curse of dimensionality*⁹). Métodos más eficientes que *linear programming* se han aplicado para manejar esta complejidad: utilizar un algoritmo genético [XCW00], hacer uso de un algoritmo basado en ascenso (*hill climbing*) [GMP04, GMP05], o bien como en [EAYH01a, EAYH01c, EAYH01b, EAYH01d, PGGMCG04] emplear un método de aprendizaje automático como *reinforcement learning* [SB98].

Finalmente, otra perspectiva que ha de valorarse cuando se diseña un sistema de CAS es el despliegue de políticas estáticas o adaptativas. Las políticas estáticas consideran los parámetros del sistema —las tasas de llegada de cada clase de peticiones, etc.— como estacionarios y buscan el diseño de una política de CAS para el escenario más pesimista (al caso peor). Es razonable pensar

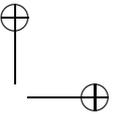
⁹«*The curse of dimensionality*» es una expresión acuñada por Richard E. Bellman a propósito del problema causado por el rápido incremento que experimenta el dominio espacial cuando se añaden dimensiones extra a un espacio matemático. R.E. Bellman, *Adaptive Control Processes*. Princeton University Press, 1961.



que las políticas así concebidas son relativamente efectivas cuando el perfil de tráfico se comporta de manera consistente con las previsiones formuladas durante el diseño de las mismas. Una alternativa es desarrollar políticas que ofrezcan alguna protección cuando el tráfico de una clase aumenta por encima de lo anticipado: por ejemplo, *virtual partitioning* (VP) [BM98, YMW⁺04] extiende las políticas de *trunk reservation* asignando un doble *guard channel* para cada clase, uno para las situaciones de carga normal y otro (más restrictivo) indicado para las situaciones en las que una clase de peticiones excede la capacidad nominal asignada.

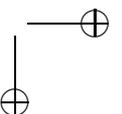
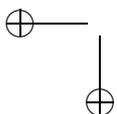
Frente a los enfoques estáticos, las políticas adaptativas consideran los parámetros del sistema como no estacionarios y bien emplean información histórica de los patrones de tráfico o bien los estiman periódicamente para seguir las condiciones de la red con mayor precisión. Por lo común, las estrategias adaptativas diseñadas para operar en escenarios de tráfico no estacionario o bien hacen uso de dilatadas ventanas de medida para estimar los parámetros del sistema —y en consecuencia periodos de convergencia demasiado largos como para hacer frente a las condiciones de operación real— como en [ZL01, WZZZ03, QLdM04], o no proporcionan una indicación explícita de cómo debe configurarse la ventana temporal como en [JJ01, WLR⁺04, HKK04, MHY05]. Además, como se subraya en [GRDBMBP05], los estudios dedicados al comportamiento circunstancial del esquema frente a la no estacionariedad del sistema (como el de [NHA05]) son excepción y la mayoría de las contribuciones en materia de esquemas adaptativos sólo consideran el régimen estacionario y no proporcionan ninguna evidencia acerca del comportamiento de los mismos en régimen transitorio. En consecuencia, da la impresión de que una característica fundamental de los esquemas adaptativos, como es su velocidad de convergencia a las nuevas condiciones de operación, no se ha explorado lo suficiente en la literatura.

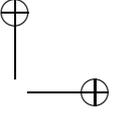
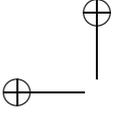
Los mecanismos adaptativos de CAS se han estudiado en otros contextos. Por ejemplo, en ciertas propuestas adaptativas el ajuste de los parámetros de configuración de la política se basa en estimaciones de los patrones de movilidad y de las tasas de llegadas de traspasos derivadas del número actual de



Capítulo 2. El estudio del control de admisión en redes celulares

sesiones en curso en las células vecinas, como las políticas de *guard channel* adaptativo de [YL97, RSAK99, YK02, SK04]. Otro ejemplo de esquemas adaptativos se propone en [HKK04], donde se consideran aplicaciones multimedia de tasa adaptativa.





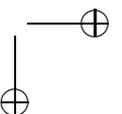
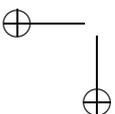
Capítulo 3

Impacto de la movilidad en el dimensionado de redes

—

Durante el estudio analítico de los sistemas multiservicio fijos en presencia de políticas de control de admisión, una de las principales complicaciones es el tamaño del sistema de ecuaciones que se necesita resolver para obtener las prestaciones del sistema multiservicio. El número de operaciones, desafortunadamente, no guarda una relación lineal con las dimensiones del propio sistema a modelar —tanto en términos del número de servicios que se contemplan como de la cantidad de recursos a disposición de los servicios— por lo que, incluso la resolución directa de sistemas multiservicio fijos relativamente modestos puede llegar a necesitar un esfuerzo computacional excesivo. No es pues de extrañar que muchas de las contribuciones en este campo se hayan dedicado a facilitar el procedimiento de resolución exacta de las ecuaciones que describen estos sistemas, sea tratando de buscar soluciones en forma de producto, en forma de funciones generatrices basadas en transformadas o poniendo en práctica observaciones astutas que conduzcan a soluciones elegantes, (aunque es de anticipar que desafortunadamente no todas las políticas producen dichas simplificaciones) [Ros95].

La necesidad de que los modelos matemáticos sean tratables aumenta de



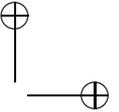
importancia, aun más si cabe, cuando el foco de atención pasa de los sistemas fijos a los sistemas móviles, debido a que a estos últimos hay que añadir la complejidad inherente a todo el modelado de los trasposos entre células. Para tratar de dar una solución satisfactoria a este problema, en este capítulo se detalla la extensión a entornos móviles celulares de una técnica exacta de análisis y dimensionado desarrollada originalmente para el estudio de sistemas multiservicio fijos. Se asume que los servicios considerados se pueden describir mediante modelos de tráfico Binomial-Poisson-Pascal (BPP)¹ que solicitan acceso a un conjunto de recursos limitados donde las sesiones que no encuentran suficientes recursos libres se pierden (*lost-call-cleared system*). Las políticas de control de admisión estudiadas son *Complete Sharing* (CS) e *Integer Limit* (IL) [Ive87] —también llamada *Upper Limit* (UL) [CLW95]—.

El resto del capítulo está estructurado de la siguiente manera: en la Sección 3.1 se introduce la materia, en la Sección 3.2 se detallan los parámetros del modelo empleado para describir tanto el sistema fijo multiservicio, como el sistema móvil multiservicio, en la Sección 3.3 se procede a la caracterización analítica de los modelos fijo y móvil para distintos modelos de tráfico y se describe el procedimiento de dimensionado empleado, en la Sección 3.4 se presenta un ejemplo numérico que confirma la utilidad de la herramienta analítica extendida a escenarios móviles y justifica la necesidad de modelar adecuadamente el tráfico de trasposos y finalmente la Sección 3.5 se resume este capítulo y se detallan las conclusiones obtenidas.

3.1 Antecedentes

La herramienta analítica que se emplea durante este estudio está basada en el trabajo de Iversen en [Ive87] para el análisis de sistemas multiservicio fijos. La técnica de [Ive87] se fundamenta en un cálculo eficiente y estable numéricamente —mediante el empleo de convoluciones— de la solución en forma

¹Los BPP son una clase de procesos de llegada con tasas linealmente dependientes del estado.



de producto de un sistema que emplea una política de control de admisión que pertenece a esta familia (como por ejemplo, CS).

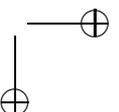
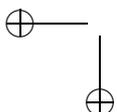
Durante la labor de introducir la movilidad en el sistema multiservicio se requiere la determinación de la tasa de llegadas de las peticiones de traspaso. Durante esta tarea se opta por obtener la misma a partir de un procedimiento iterativo que equilibra los flujos de traspasos entrantes y salientes sobre la célula objeto de estudio, a similitud de [AMDCL⁺99]. Sin embargo, también es práctica usual considerar que la tasa de llegada de peticiones de traspaso es una fracción conocida de la tasa de llegada de peticiones de sesiones nuevas [Jab96].

3.2 Descripción del modelo

3.2.1 Sistema fijo

Se considera el sistema compuesto por un conjunto de R servicios diferentes que compiten por una cantidad limitada de C unidades de recurso. Cada servicio tiene una caracterización de fuente que representa el proceso de generación de peticiones de conexión (sólo peticiones de sesiones nuevas en el caso del sistema fijo) asociado a dicho servicio. El servicio r ($r = 1, \dots, R$) se caracteriza por los siguientes parámetros:

- c_r
El número de unidades de recurso necesarias para cursar una petición.
- n_r
El número máximo de peticiones simultáneas en curso. Se debe cumplir que $n_r c_r \leq C$. Este parámetro permite definir políticas de control de admisión del tipo IL. Cuando se emplea la política CS este valor es el máximo posible, es decir: $n_r = \lfloor C/c_r \rfloor$, donde « $\lfloor x \rfloor$ » denota al entero inmediatamente inferior a x .



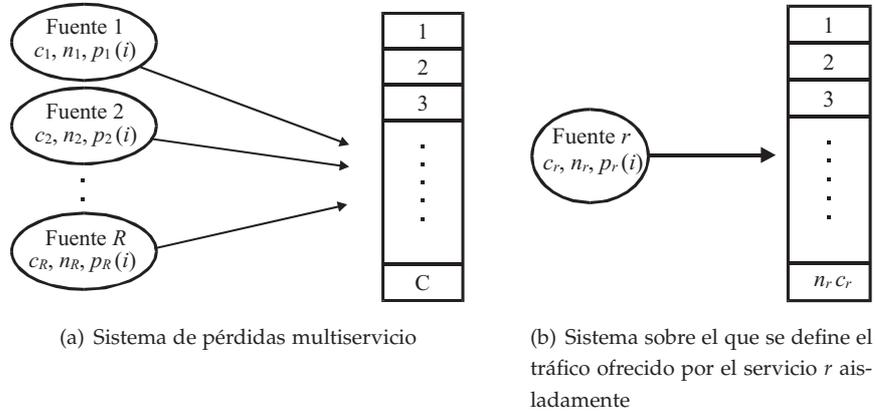
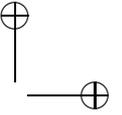
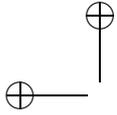


Figura 3.1: Dos modelos del sistema

- m_r
Tiempo medio de servicio de las sesiones del servicio r . Aunque se asume que la variable aleatoria que modela este parámetro está distribuida exponencialmente con tasa $\mu_r = 1/m_r$, los modelos considerados son insensibles al tipo de distribución, siendo sólo relevante su valor medio [Ive87].
- a_r
Tráfico ofrecido por terminal del servicio r .
- $\{\lambda_r(i)\}; i = 0, \dots, n_r c_r$
La tasa de generación de peticiones de conexión definida sobre el sistema de la Fig. 3.1(a). En general será función del estado del sistema. El vector $\{\lambda_r(i)\}$ tiene $n_r c_r + 1$ elementos, pero sólo están definidos los elementos que son múltiplos de c_r , el resto son cero.²

²Los valores de $\{\lambda_r(i)\}$ dependen del modelo de tráfico considerado. Por ejemplo, si considera un modelo de población finita, (caso Binomial) entonces $\lambda_r(kc_r) = (M_r - k)\lambda'_r$ donde M_r es el número de individuos/terminales asociados al servicio r y λ'_r es la intensidad de tráfico por terminal libre. El valor de λ'_r puede obtenerse a partir de la expresión que lo relaciona con el tráfico ofrecido por terminal y el tiempo medio de servicio $a_r = \lambda'_r / (\lambda'_r + \mu_r)$. En la Sección 3.3 se trata en mayor profundidad este asunto.



- f_r
Relación entre el número de terminales/usuarios de cada servicio r y el número total de los mismos (denominado generalmente «vector de penetración del servicio»).
- $\{p_r(i)\}; i = 0, \dots, n_r c_r$
Las probabilidades de estado en régimen permanente definidas sobre el sistema de la Fig. 3.1(b). En dicho sistema hipotético sólo se considera la existencia del servicio r . El vector $\{p_r(i)\}$ tiene $n_r c_r + 1$ elementos, pero sólo están definidos los elementos que son múltiplos de c_r , el resto son cero. Las $\{p_r(i)\}$ se determinan mediante la siguiente expresión:

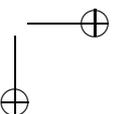
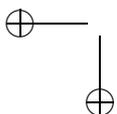
$$p_r(kc_r) = p_r(0) \frac{\prod_{j=0}^{k-1} \lambda_r(jc_r)}{k! \mu_r^k}; \quad k = 0, \dots, n_r \quad (3.1)$$

Sólo los valores relativos de $\{p_r(i)\}$ son relevantes para obtener los parámetros de tráfico que definen las prestaciones del sistema, por lo que se puede fijar $p_r(0) = 1$. No obstante, es conveniente normalizar $\{p_r(i)\}$ de forma que $\sum_{i=0}^{n_r c_r} p_r(i) = 1$.

3.2.2 Sistema móvil

Se analiza el sistema a partir del comportamiento de una única célula aislada (ver Sección 2.1.2). Para introducir la movilidad se modifica la caracterización del servicio r ($r = 1, \dots, R$) de la siguiente manera:

- $\{\lambda_r(i)\}$ pasa a ser $\{\lambda_r^n(i)\}; i = 0, \dots, n_r c_r$
La tasa de generación de peticiones de establecimiento de nuevas sesiones del servicio r (función del estado del sistema).
- Se introduce $\{\lambda_r^h\}$
La tasa de generación de peticiones de traspaso del servicio r procedentes de células adyacentes. El tiempo entre llegadas de trasposos es una variable aleatoria que sigue un proceso de Poisson.



- Se introduce m_r^h
 Tiempo medio entre dos peticiones de traspaso consecutivas para la misma sesión del servicio r (también denominado «tiempo medio de residencia en la célula» o *dwell time*). Se asume que la variable aleatoria que modela este parámetro está distribuida exponencialmente con tasa $\mu_r^h = 1/m_r^h$.

3.3 Análisis

3.3.1 Caracterización del sistema fijo

Dados dos vectores de probabilidades de estado $x = \{x(0), \dots, x(w_x)\}$ e $y = \{y(0), \dots, y(w_y)\}$, se define el operador convolución «*» como aquel que aplicado a x e y da como resultado el vector $z = x * y$ con los siguientes elementos:

$$z(i) = \sum_{j=u(i)}^{v(i)} x(i-j)y(j); \quad i = 0, \dots, w_z \quad (3.2)$$

donde los elementos de los vectores u y v están definidos de la siguiente manera:

$$u(i) = \text{máx}(0, i - w_x) \equiv \begin{cases} 0 & 0 \leq i < w_x \\ i - w_x & w_x \leq i \leq w_z \end{cases} \quad (3.3)$$

$$v(i) = \text{mín}(i, w_y) \equiv \begin{cases} i & 0 \leq i < w_y \\ w_y & w_y \leq i \leq w_z \end{cases} \quad (3.4)$$

Para determinar el valor de los diferentes parámetros de tráfico que definen las prestaciones del sistema se sigue el siguiente procedimiento:

1. Por sucesivas convoluciones de los vectores probabilidad de estado normalizados $\{p_r(i)\}$ de todas las fuentes, en cualquier orden, se obtiene

el vector $q_N = \{q_N(0, \dots, q_N(C))\}$, que representa las probabilidades de estado del sistema de la Fig. 3.1(a).

2. Para cada una de las fuentes, por ejemplo la r -ésima, se «deconvoluciona» q_N en dos vectores $q_{N/r}$ y p_r , de manera que $q_N = q_{N/r} * p_r$. La complejidad computacional asociada al cálculo de las diferentes convoluciones puede reducirse utilizando el algoritmo propuesto en [IS97].
3. La «probabilidad de bloqueo» (PB) o *time congestion* para el servicio r se determina mediante:

$$PB_r = \frac{1}{N_z} \left[p_r(n_r c_r) \sum_{i=0}^{C-n_r c_r} q_{N/r}(i) + \sum_{j=C-c_r+1}^C \sum_{i=0}^{(n_r-1)c_r} q_{N/r}(j-1) p_r(i) \right] \quad (3.5)$$

donde $N_z = \sum_{i=0}^C q_N(i)$. En (3.5) están contenidos los términos que definen la PB del servicio r en un determinado instante del tiempo: bien que el número de sesiones en curso sea igual a n_r o bien que el número de recursos libres no sea suficiente como para cursar una petición de inicio de nueva sesión.

4. La «probabilidad de pérdidas» (PP) o *call congestion* para el servicio r se determina mediante la siguiente expresión:

$$PP_r = \frac{p_r(n_r c_r) \lambda_r(n_r c_r) \sum_{i=0}^{C-n_r c_r} q_{N/r}(i) + \sum_{j=C-c_r+1}^C \sum_{i=0}^{(n_r-1)c_r} q_{N/r}(j-i) p_r(i) \lambda_r(i)}{\sum_{j=0}^C \sum_{i=0}^j q_{N/r}(j-i) p_r(i) \lambda_r(i)} \quad (3.6)$$

La PP del servicio r representa la fracción de peticiones no cursadas del servicio r sobre el total de peticiones ofrecidas al sistema por el servicio r .

5. El tráfico ofrecido (TO) por el servicio r , se define como el tráfico que sería capaz de cursar de dicho servicio un sistema con infinitos recursos.

$$TO_r = \sum_{i=0}^{n_r c_r} i p_r(i) \quad (3.7)$$

donde $\{p_r(i)\}$ está normalizado.

6. El tráfico cursado (TC) del servicio r por parte del sistema se obtiene mediante la siguiente expresión:

$$TC_r = \frac{1}{N_z} \left[\sum_{j=0}^C \sum_{i=0}^j i q_{N/r}(j-i) p_r(i) \right] \quad (3.8)$$

7. La «congestión de tráfico» (CT) o *traffic congestion* para el servicio r se obtiene mediante la siguiente fórmula:

$$CT_r = \frac{TO_r - TC_r}{TO_r} \quad (3.9)$$

3.3.2 Modelos de tráfico

La técnica que se acaba de describir se puede aplicar a un conjunto muy extenso de modelos de tráfico: sólo es necesario concretar para cada uno de ellos el valor que toma $\{p_r(i)\}$ (es decir, las probabilidades de estado en régimen permanente definidas sobre el sistema ficticio de la Fig. 3.1(b)). En concreto se tratarán en este trabajo dos casos concretos del modelo de tráfico BPP: a) el caso Binominal (modelo de Engset/Engset-B); b) el caso Poisson (modelo de Erlang/Erlang-B).

Las fuentes BPP se modelan mediante procesos de nacimiento y muerte —un caso especial de proceso de Markov de tiempo continuo, por lo que el tiempo entre transiciones sigue una distribución exponencial—. Su diagrama de estados se muestra en la Fig. 3.2. Dependiendo de la definición de λ_n y γ_n se originan los diferentes modelos de tráfico.

Distribución de Poisson

Para un servicio r se asume que los tiempos de servicio están distribuidos exponencialmente con valor medio $1/\mu_r$ y que el proceso de llegadas es un proceso de Poisson —equivalente a un número infinito de terminales con una

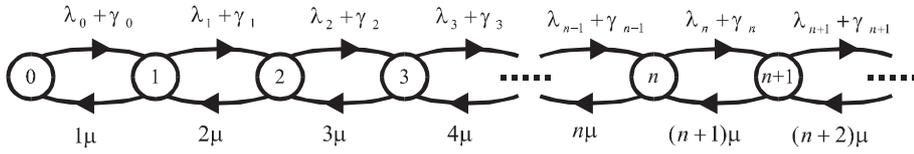


Figura 3.2: Diagrama de estados para los sistemas BPP

tasa total limitada a λ_r (modelo de «población infinita»)—. Como resultado, el proceso de tráfico se convierte en un proceso de nacimiento y muerte puro ($\lambda_n = \lambda_r, \gamma_n = 0$), un proceso de Markov simple, fácil de manejar matemáticamente, y para el cual existen gran cantidad de resultados deducidos previamente, (ver por ejemplo [Ive02]). Las probabilidades de estado del servicio r definidas en la fórmula 3.1 se pueden expresar de la siguiente forma:

$$p_r(kc_r) = p_r(0) \frac{1}{k!} \left(\frac{\lambda_r}{\mu_r} \right)^k \quad (3.10)$$

Distribución Binominal

Se asume que existe un número limitado de terminales M_r por servicio r (modelo de «población finita»). Cada terminal (ver Fig. 3.3) produce sesiones con una intensidad λ'_r cuando está libre (en reposo). Cuando tiene una sesión en curso, la intensidad es cero. Así pues, el proceso de llegadas depende del estado ($\lambda_n = M_r \lambda'_r, \gamma_n = -n \lambda'_r$), es decir: si n terminales tienen sesiones en curso, entonces la tasa agregada de llegadas al sistema equivale a $(M_r - n) \lambda'_r$. Las probabilidades de estado en régimen permanente del servicio r para el sistema de la Fig. 3.3 son $p(0) = \mu_r / (\lambda'_r + \mu_r)$ y $p(1) = \lambda'_r / (\lambda'_r + \mu_r)$. Si se denota como $a_r = p(1)$ y $b_r = p(0) = 1 - a_r$, entonces a_r es el tráfico ofrecido por terminal. De esta manera las probabilidades de estado del servicio r definidas en la fórmula 3.1 se pueden expresar de la siguiente

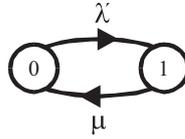


Figura 3.3: Diagrama de estados de un terminal libre

forma:

$$\begin{aligned}
 p_r(kc_r) &= p_r(0) \frac{\prod_{j=0}^{k-1} \lambda_r(jc_r)}{k! \mu_r^k} = p_r(0) \frac{\prod_{j=0}^{k-1} \lambda'_r(M_r - j)}{k! \mu_r^k} = p_r(0) \left(\frac{\lambda'_r}{\mu_r} \right)^k \frac{\prod_{j=0}^{k-1} (M_r - j)}{k!} = \\
 &= p_r(0) \left(\frac{a_r}{1 - a_r} \right)^k \frac{\prod_{j=0}^{k-1} (M_r - j)}{k!}; \quad k = 0, \dots, n_r \quad (3.11)
 \end{aligned}$$

Como se puede observar, las probabilidades de estado son función sólo del tráfico ofrecido por individuo a_r y no de λ'_r ni de μ_r . De la misma manera, los parámetros que definen las prestaciones del sistema, (ecuaciones 3.5–3.9) tampoco lo son.

3.3.3 Caracterización del sistema móvil

La determinación de λ_r^h se realiza mediante un procedimiento iterativo, basado en el balance de los flujos de traspaso entrante y saliente a la célula estudiada. Se puede inicializar la iteración con $\lambda_r^{h,(0)} = 0$ o con una conjetura para el valor que se juzgue próximo al verdadero. Por ejemplo, si el sistema se encuentra en equilibrio estadístico, las tasas de llegada de trasposos se relacionan con las tasas de llegada de sesiones nuevas y las probabilidades de bloqueo a través de la expresión [LMN94]

$$\lambda_r^h = \lambda_r^n \frac{1 - P_r^n}{\frac{\mu_r^h}{\mu_r} + P_r^h} \quad (3.12)$$

donde P_r^n y P_r^h son las probabilidades de bloqueo para sesiones nuevas y traspasos, respectivamente. Nótese que estos valores se podrían substituir por sus correspondientes objetivos de QoS, dado que se asume equilibrio estadístico.

El proceso iterativo comienza calculando el equilibrio de flujos para el servicio r definido por la expresión

$$\lambda_r^{h,(m)} = \mu_r^h \sum_{k=0}^{n_r} k \pi_r^{(m)}(kc_r) \quad (3.13)$$

donde $\lambda_r^{h,(m)}$ es la tasa de llegada de peticiones de traspaso del servicio r procedentes de las células adyacentes en la m -ésima iteración y el vector $\{\pi_r^{(m)}(i)\}$ es la distribución del número de sesiones en curso del servicio r en la m -ésima iteración. Esta distribución se obtiene mediante la expresión

$$\pi_r^{(m)}(kc_r) = \frac{1}{N_z^{(m)}} \sum_{i=0}^{C-kc_r} q_{N/r}^{(m)}(i) p_r^{(m)}(kc_r) \quad (3.14)$$

donde $N_z^{(m)} = \sum_{i=0}^C q_N^{(m)}(i)$ es la constante de normalización, $\{q_N^{(m)}(i)\}$ es el vector de convolución total y $\{q_{N/r}^{(m)}(i)\}$ es el vector de convolución parcial donde falta el servicio r . Como en otros casos, el vector $\{\pi_r^{(m)}(i)\}$ tiene $n_r c_r + 1$ elementos, pero sólo están definidos aquellos cuyo índice es múltiplo de c_r , el resto son cero.

Los elementos del vector $\{p_r^{(m)}(i)\}$ para el modelo de tráfico Binomial, por ejemplo, se obtienen a partir de la expresión:

$$p_r^{(m)}(kc_r) = p_r^{(m)}(0) \prod_{i=0}^{k-1} \frac{(M_r - i) \lambda_r' + \lambda_r^{h,(m-1)}}{(i+1)(\mu_r + \mu_r^h)}; \quad k = 0, \dots, n_r \quad (3.15)$$

El proceso iterativo se da por terminado cuando $(\lambda_r^{h,(m)} / \lambda_r^{h,(m-1)}) < \delta$ donde δ es una cierta tolerancia.

Tras lograr el equilibrio de flujos, las expresiones para los parámetros de prestaciones obtenidas para el modelo fijo siguen siendo válidas, es decir, las probabilidades de bloqueo se obtienen de (3.5), las probabilidades

de pérdidas para sesiones nuevas se obtienen de (3.6). Finalmente, substituyendo $\lambda_r^h(i)$ por λ_r^h en (3.6), se cancela λ_r^h en numerador y denominador y $PP_r^h = PB_r$, como era de esperar dado que la intensidad de trasposos es independiente del estado del sistema.³ A su vez, el tráfico cursado se obtiene a partir de la expresión:

$$TC_r = \sum_{i=0}^{n_r c_r} i \pi_r(i) \quad (3.16)$$

De nuevo se considera que el vector $\{\pi_r^{(m)}(i)\}$ tiene $n_r c_r + 1$ elementos, pero sólo se definen aquellos cuyo índice es múltiplo de c_r , el resto son cero.

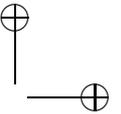
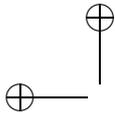
3.3.4 Dimensionado del sistema fijo

El procedimiento de dimensionado trata de determinar el número máximo de terminales/usuarios de cada servicio a los que puede servir el sistema sin comprometer los requisitos de Calidad de Servicio/Grado de Servicio (QoS/GoS) acordados previamente. Se manejan dos definiciones de GoS y en algunos casos, se asume que ambas restricciones se aplican simultáneamente:

- B_r^n , cotas superiores para la probabilidad de pérdidas/bloqueo de cada servicio por separado.
- \overline{B}_r^n , una cota superior para la media aritmética de las probabilidades de pérdidas/bloqueo.

El dimensionado de sistemas multiservicio reviste complejidad porque las probabilidades de pérdidas de los diferentes servicios no crecen monótonamente con el tráfico ofrecido. No obstante, esta dificultad puede reducirse notablemente si el operador es capaz de predecir el mencionado el vector de penetración $\{f_r\}$ de los servicios que componen el sistema. El algoritmo de dimensionado propuesto se basa en el algoritmo de la búsqueda binaria (*binary search* o *binary chop*) y varía según el modelo de tráfico que se considere.

³Esta es la propiedad PASTA (*Poisson Arrivals See Time Averages*).

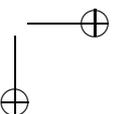
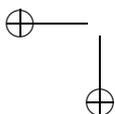


Distribución Binominal

El algoritmo comienza por definir los límites del intervalo sobre el que se realizará la búsqueda de la población de usuarios óptima. Un intervalo de partida adecuado podría ser $[0, M_T]$, con $M_T = C / (\text{mín}(a_r) \text{mín}(c_r))$. Es decir, se asume C como cota superior para el tráfico agregado que se puede ofrecer al sistema y que éste consta de un único servicio hipotético cuyos usuarios ofrecen un tráfico de $\text{mín}(a_r)$ y requieren $\text{mín}(c_r)$ recursos por sesión. El resultado es una cota superior razonable M_T para la población agregada de usuarios. A continuación se elige el punto medio del intervalo, $M_T/2$, se reparte la población entre los diferentes servicios según el vector penetración, se obtienen las PP_r y se verifica el cumplimiento de los objetivos de GoS. Si se incumplen, la población agregada es excesiva y se debe elegir el nuevo punto medio en el subintervalo inferior $[0, M_T/2]$. De lo contrario, la población es insuficiente y se debe escoger el nuevo punto medio en el subintervalo superior $[M_T/2, M_T]$. El algoritmo se repite hasta que se cumple exactamente el GoS impuesto (improbable), hasta que la diferencia entre el GoS deseado y el obtenido sea inferior a una tolerancia ϵ o —debido a que la cantidad de terminales es un número entero— cuando no se pueda refinar la solución más allá del límite impuesto por la necesidad de redondear el número de terminales al entero inmediatamente inferior. Finalmente el valor de población pormenorizado para cada servicio se obtiene a partir del vector penetración y el valor $M_{T(\text{máx})}$ de población total máxima obtenido, por medio de la expresión $M_r = M_{T(\text{máx})} f_r$.

Distribución de Poisson

El extremo superior del intervalo inicial de búsqueda para la tasa de llegada de peticiones de establecimiento de nuevas sesiones se propone, de manera similar, como la siguiente cota superior: $\lambda_T = (C \text{máx}(\mu_r)) / \text{mín}(c_r)$. El proceder es análogo al anterior; no obstante, no es necesario emplear mecanismos de redondeo, debido a que la tasa es una magnitud continua —a



diferencia del número de terminales por servicio—. De la misma manera, el valor de tasa pormenorizado para cada servicio se obtiene a partir del vector penetración por medio de la expresión: $\lambda_r = \lambda_{T(\text{máx})} f_r$.

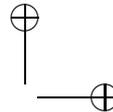
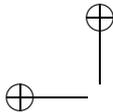
3.3.5 Dimensionado del sistema móvil

El algoritmo de búsqueda binaria explicado en la anterior sección puede aplicarse de manera similar al sistema con movilidad. Sin embargo ha de tenerse en cuenta que, a parte de mantener los objetivos de GoS en términos de la probabilidad de pérdidas para peticiones de nuevas sesiones, se deben definir objetivos de GoS adicionales que limiten la probabilidad de que una sesión en curso sufra una interrupción definitiva en la comunicación. Se denota como B_r^h a esta cota por servicio r . Las interrupciones se deben a una insuficiencia fortuita de recursos en la célula destino cuando se produce un traspaso y provocan que la sesión no se complete satisfactoriamente. La importancia de establecer estos objetivos es mayor en la medida en que, desde el punto de vista del usuario, resulta más indeseable la interrupción de una sesión ya en progreso que el bloqueo inicial de una petición de establecimiento de nueva sesión.

Suponiendo que, tanto las distribuciones del tiempo medio de duración de una sesión como del tiempo medio de residencia en una célula son exponenciales, la probabilidad de que una nueva sesión se complete satisfactoriamente —es decir, la probabilidad de que no fracase la comunicación, bien por rechazo de la petición de conexión o por fallo de traspaso— viene definida por la expresión

$$1 - PF_r = (1 - PP_r^n) PC_r \sum_{k=0}^{\infty} PH_r^k (1 - PP_r^h)^k = \frac{(1 - PP_r^n) PC_r}{1 - (1 - PP_r^h) PH_r} \quad (3.17)$$

donde PF_r es la probabilidad de fracaso de una sesión del servicio r , $PC_r = \mu_r / (\mu_r + \mu_r^h)$ es la probabilidad de que una sesión nueva del servicio r termine antes del siguiente traspaso y $PH_r = \mu_r^h / (\mu_r + \mu_r^h)$ es la probabilidad de que un terminal con sesión en curso del servicio r realice un traspaso. Nótese



que $\lim_{\mu_r^h \rightarrow 0} PF_r = PP_r^n$ (cuando la movilidad es nula, el fracaso es sólo función de la probabilidad de pérdidas de sesión nueva), y que $\lim_{\mu_r^h \rightarrow \infty} PF_r = 1$ (cuando la movilidad es elevada, la probabilidad de fracaso también lo es, debido a la mayor exposición de la sesión a un hipotético fallo de traspaso).

3.4 Resultados numéricos

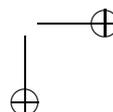
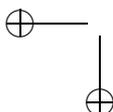
A continuación figuran algunos resultados que ilustran el impacto de la movilidad en el dimensionado de los sistemas estudiados. A lo largo de esta sección se supondrán modelos de tráfico Binomial (la población de terminales que forma parte del servicio es finita), sin embargo, las conclusiones obtenidas son similares para modelos de tráfico de Poisson.

3.4.1 Sistema fijo

Se considera una célula aislada con los siguientes parámetros: $C = 122$ unidades de recurso, $R = 3$ servicios, $c_r = \{1, 1, 2\}$ unidades de recurso demandadas por las peticiones de sesión nueva de cada servicio, $f_r = \{0.1, 0.6, 0.3\}$ vector de penetración por servicio, $a_r = 0.22$ Erlangs de tráfico ofrecido por cualquier terminal sin importar el servicio, $\overline{B}_r^n = 0.01$ como cota superior para el valor medio de la probabilidad de pérdidas y una política de control de admisión CS.⁴

Los resultados se pueden observar en la Tabla 3.1. La columna M_r indica el número máximo de terminales que puede soportar el sistema por cada servicio. Las columnas PP_r , PB_r y TC_r indican respectivamente las probabilidades de pérdidas, bloqueo y el tráfico cursado de cada servicio r .

⁴Nótese que debido a que los parámetros de tráfico que definen las prestaciones del sistema no son función de λ_r' ni de μ_r , sino sólo de a_r (ver Sección 3.3), no es necesario precisar el valor de dichos parámetros.



Cuadro 3.1: Dimensionado del sistema fijo

r	M_r	PB_r	PP_r	TC_r
1	36	0.007018	0.006748	7.878248
2	214	0.007018	0.006748	46.831810
3	107	0.015243	0.014070	46.561726
Total	357			101.27178
Media		0.009759	0.009189	

3.4.2 Sistema móvil

La comparación con el sistema móvil plantea las conclusiones más interesantes. Se asume el mismo escenario que en el apartado anterior, salvo que el GoS está expresado esta vez en términos de la probabilidad media de fracaso de la sesión, $\overline{PF}_r = 0.01$. Adicionalmente, se consideran los siguientes valores para los tiempos medios de las sesiones de cada servicio $\mu_r = \{2700, 160, 90\}$, y para los tiempos de residencia en la célula de un terminal con sesión en curso $\mu_r^h = \{120, 60, 60\}$. Para definir los parámetros del servicio 1 se ha pensado en un servicio de acceso a Internet, en el que el usuario permanece conectado durante un periodo de tiempo prolongado y tiene una movilidad más reducida. Para un sistema GSM/DCS que utilizan la banda de frecuencias de 1.8 GHz, con radios de célula típicos de 1 Km y suponiendo un modelo de movilidad lineal, los valores de μ_r^h definidos anteriormente se corresponden con unas velocidades medias de 30, 60 y 60 Km/h respectivamente.

Los resultados del dimensionado con movilidad se muestran en la tabla 3.4.2. Como se observa, la población máxima de terminales se ha reducido en un 10.36% y el tráfico cursado en un 8.25% respecto a los resultados del sistema fijo (tabla 3.1).

En la Tabla 3.4.2 se muestra el impacto que tiene la movilidad sobre el dimensionado del sistema fijo (que no tenía en cuenta este factor). Se observa la degradación del GoS, ya que $\overline{PF}_r = 5,95\%$ frente a la $\overline{PP}_r = 0,9\%$ del sistema original. Por otra parte, la movilidad afecta especialmente al servicio

Cuadro 3.2: Dimensionado del sistema móvil

r	M_r	PP_r^m	PP_r^h	PF_r	TC_r
1	32	0.000880	0.000935	0.021485	6.920475
2	192	0.000883	0.000935	0.003368	42.129208
3	96	0.001863	0.002099	0.004999	42.076259
Total	320				91.125942
Media		0.001208	0.001323	0.009951	

Cuadro 3.3: Impacto de la movilidad sobre el dimensionado del sistema fijo

r	M_r	PP_r^m	PP_r^h	PF_r	TC_r
1	36	0.005841	0.006089	0.125755	7.122127
2	214	0.005830	0.006089	0.021718	46.278606
3	107	0.012060	0.013156	0.031199	45.926348
Total	357				99.327082
Media		0.007913	0.008444	0.059557	

1, ya que $PF_1 = 12,57\%$, un valor considerablemente mayor a las demás. De los resultados anteriores se concluye que presumir un sistema fijo durante el dimensionado no es suficiente para asegurar los objetivos de QoS. De ahí que deba precisarse el impacto que ejerce la movilidad de los terminales sobre las prestaciones del sistema. Esta es la razón de que se haya extendido a sistemas móviles una técnica exacta diseñada, en principio, sólo para sistemas fijos.

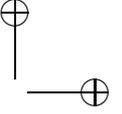
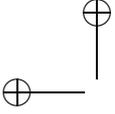
3.5 Conclusiones

El trabajo presentado parte de la técnica de análisis de Iversen [Ive87] para sistemas fijos —en presencia de la política de control de admisión con solución en forma de producto, *Integer Limit*—, y extiende dicha técnica para la evaluación de prestaciones de sistemas móviles celulares multiservicio, considerando tanto modelos de tráfico Binomial (población finita) como de

Poisson (población infinita).

La contribución principal reside en la introducción dentro de la técnica de análisis de la caracterización del sistema móvil, por medio de un procedimiento iterativo que equilibra los flujos de traspaso entrante a y saliente de la célula.

Los resultados numéricos obtenidos para el dimensionado de un sistema de ejemplo —al que se impone una cota para la probabilidad de pérdidas media— confirman la utilidad de la técnica. Asimismo muestran que un sistema móvil, en general, ve degradada su capacidad con respecto al mismo sistema considerado fijo, debido al incremento de tráfico procedente de los trasposos. Estos resultados indican la necesidad de que los operadores consideren durante la fase de diseño de la red, cuál es el impacto exacto que tendrá la movilidad sobre el número máximo de usuarios que podrá servir el sistema, y en consecuencia sobre la utilización que se hará de la red.



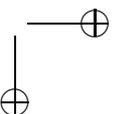
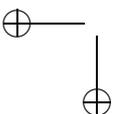
Capítulo 4

Optimización del control de admisión

En el capítulo 3 se ha analizado el impacto que tiene la movilidad sobre el dimensionado de los sistemas móviles celulares multiservicio, dejando patente que la degradación que sufren las prestaciones es, en general, diferente para cada servicio considerado. Esta disimilitud obedece, a su vez, a que la demanda de recursos que exige cada servicio es diferente —consecuencia directa de los parámetros propios que caracterizan cada servicio—. De ahí que los operadores consideren necesario introducir mecanismos para la gestión adecuada de los recursos, como por ejemplo, mecanismos de Control de Admisión (CA).

La política de *CA Integer Limit* que se introduce de manera implícita en el capítulo 3 es sólo una de las múltiples propuestas realizadas en este campo. A lo largo del presente capítulo se estudiarán otras diversas políticas también con solución en forma de producto, así como políticas para las que la técnica de análisis exacta del capítulo 3 deja de ser aplicable, como por ejemplo las de la familia de *trunk reservation*.

El resto del capítulo está estructurado de la siguiente manera: en la Sección 4.1 se introduce la materia, en la Sección 4.2 se detallan los parámetros del modelo empleado para describir el sistema móvil multiservicio, en la Sección 4.3 se definen las políticas de control de admisión estudiadas, en

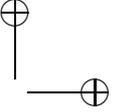
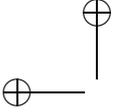


la Sección 4.4 se trata el diseño y dimensionado del sistema, explicando las herramientas desarrolladas para el análisis y síntesis de políticas, en la Sección 4.5 se presenta la comparativa entre las distintas políticas de control de admisión estudiadas, en términos de capacidad, sensibilidad a variaciones en los parámetros de configuración de la política y tolerancias a sobrecarga, en la Sección 4.6 se presenta un nuevo algoritmo de escalada para la optimización de políticas basado en el concepto novedoso del espacio de soluciones y finalmente en la Sección 4.7 se hace un resumen de este capítulo y se presentan las conclusiones.

4.1 Antecedentes

La gestión de los recursos radio continúa siendo un asunto de notable interés en la actualidad. El precursor principal de esta atención fue el desarrollo de la Red Digital de Servicios Integrados de Banda Ancha (RDSI-BA), que justificó el estudio de las redes fijas multiservicio [Ros95], aunque avances posteriores en las redes de acceso inalámbricas de banda ancha —*Local Multipoint Distribution Service* (LMDS), *Multichannel Multipoint Distribution Service* (MMDS), la familia de estándares IEEE 802.16 (WirelessMAN/WiMAX)— mantienen motivado el interés en esta materia. A esto hay que unir la atención fundamental que reciben actualmente los sistemas que integran movilidad y multiservicio —como la 3G de telefonía móvil celular— en parte debido al éxito que, por separado, han logrado tanto las redes móviles celulares (estándares GSM y cdmaOne) como la propia Internet.

El Control de Admisión de Sesiones (CAS) es un mecanismo de gestión de los recursos de importancia clave en el diseño y operación de las redes móviles multiservicio que desean proporcionar garantías de QoS. En este tema la literatura propone numerosas estrategias de CAS (ver a este respecto una descripción más detallada en la Sección 2.2) que difieren entre sí en la cantidad de información que tiene a su disposición el proceso que se encarga de decidir si una nueva petición de establecimiento o traspaso de debe ser

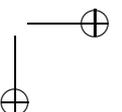
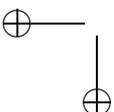


admitida o rechazada. Es razonable pensar que cuanto mayor información se proporcione a dicho proceso de decisión mayores serán las prestaciones de la política de control de admisión pero también su complejidad de implementación.

Sin embargo, la mayor o menor generalidad de una política es sólo uno de los aspectos que contribuyen a la complejidad del problema. Por un lado, tanto la cantidad de recursos del sistema como el número de servicios que tratan de acceder a los mismos afectan directamente a la dimensión del espacio de estados que se tiene que resolver numéricamente durante el análisis de una política. Por otro, el problema de dimensionar un sistema (sintetizar la política óptima que maximiza el tráfico que se puede ofrecer respetando los requisitos de QoS) requiere, usualmente, llevar a cabo varios de los procedimientos de análisis anteriormente mencionados.

En este capítulo se estudian miembros de las dos familias más significativas de políticas de control de admisión: 1) aquellas en las que la decisión de aceptación de una petición depende de la cantidad de recursos ocupados por las sesiones en curso de dicho tipo de petición —familia con solución en forma de producto—; 2) aquellas en las que esta decisión depende de la cantidad de recursos ocupados en el sistema¹ —familia de políticas de *trunk reservation*—. Como se explica, para obtener la configuración óptima de dichas políticas, los únicos parámetros que se requieren son aquellos que describen los servicios como procesos de Markov, y aquellos que fijan los objetivos de QoS para cada servicio, definidos en términos de las probabilidades de pérdidas de sesiones nuevas y traspasos (*handovers*) de sesiones ya en curso desde células vecinas. En un escenario inalámbrico esta distinción es necesaria, debido a que una sesión forzada a terminar a causa de un fallo de traspaso se considera más perjudicial que el rechazo de una petición de establecimiento de nueva sesión.

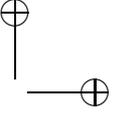
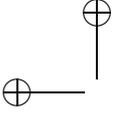
¹Se debe señalar que en los sistemas considerados, las unidades de recurso no se pueden distinguir de forma individualizada. Esto significa que la limitación de acceso ocasionada por el control de admisión es en términos de la cantidad máxima de recursos ocupados (totales o por tipo de sesión) y no en prohibir el acceso a ciertos recursos identificables de manera específica.



Las políticas con solución en forma de producto que se estudian son *Complete Sharing*, *Integer Limit* (IL) [Ive87] —también llamada *Upper Limit* (UL) [CLW95]—, *Fractional Limit* (FL) [LA95], una combinación de *Complete Partitioning* (CP) [RT89] y *Guaranteed Minimum* (GM) [CLW95] y finalmente una combinación de todas las políticas mencionadas hasta ahora. Las políticas de *trunk reservation* que se estudian son *Multiple Guard Channel* (MGC) [CC97, LLC98] y *Multiple Fractional Guard Channel* [HUCPOG03a]. Adicionalmente se trabaja con una política estacionaria aleatorizada [Ros70, BC02, PCG04], es decir, en la que la decisión del control de admisión depende únicamente del estado actual del sistema y de un componente aleatorio.

4.2 Descripción del modelo

Se asume un escenario homogéneo en el que todas las células son estadísticamente idénticas y cuyos comportamientos son independientes entre sí. En consecuencia las prestaciones del sistema se pueden analizar a partir del estudio de una única célula. Dicha célula dispone de una cantidad total de recursos C y atiende peticiones (sesiones) de distintos tipos de usuarios (que se consideran agupados en R servicios). Cada uno de estos servicios tiene unas características diferentes y unos requisitos de calidad de servicio distintos. Asimismo, para cada tipo de tráfico se distinguen peticiones de establecimiento de sesiones nuevas y peticiones causadas por un traspaso. Para facilitar la tratabilidad matemática del modelo, se asumen procesos de llegada de Poisson y que el tiempo de ocupación de los recursos por parte de una sesión está distribuido exponencialmente. A partir de estos supuestos, el sistema exhibirá la significativa propiedad de «memoria nula» (*memoryless*). En virtud de las suposiciones anteriores el estado del sistema se describe completamente a partir del vector $x = (x_1, \dots, x_R)$, donde x_r denota el número de sesiones en curso del servicio r , independientemente de si se iniciaron como peticiones de nueva sesión o como consecuencia de un traspaso. A pesar de esto, para simplificar la descripción de las políticas de control de admisión se defi-

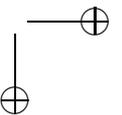
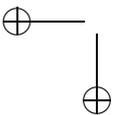


ne el vector $\mathbf{x} = (x_1^n, x_1^h, \dots, x_R^n, x_R^h)$, donde x_r^n, x_r^h es el número de sesiones en progreso (nuevas y traspasadas) del servicio r respectivamente. Obviamente, $x_r = x_r^n + x_r^h$. Adicionalmente, se denota como $c(\mathbf{x}) = \sum_{r=1}^R (x_r^n + x_r^h)c_r$ el número de recursos ocupados en el estado \mathbf{x} . Con estos condicionantes el proceso estocástico $\mathbf{x}(t)$, que describe el estado del sistema en el instante t , es una cadena de Markov de tiempo continuo, irreducible y con un número de estados finito, con un único vector de probabilidades de estado en régimen permanente π .

Cada servicio $r, r = 1, \dots, R$ está caracterizado por los siguientes parámetros:

- c_r
El número de unidades de recurso necesarias para cursar una petición del servicio r .
- $1/\mu_r^s$
La duración media de la sesión del servicio r suponiendo que los recursos del sistema son infinitos, (*unencumbered service time*).
- $1/\mu_r^d$
El tiempo medio de residencia del terminal en una célula, (*cell residence/dwell time*).
- $1/\mu_r$
El tiempo medio de ocupación de los recursos (*resource holding time*) por parte de una sesión.² Al suponer que las variables aleatorias para el tiempo de residencia en un célula y la duración de una sesión están distribuidos exponencialmente, $\mu_r = \mu_r^d + \mu_r^s$.
- λ_r^n
La tasa de llegada de peticiones de establecimiento de sesiones nuevas a una célula.

²Nótese que el tiempo de ocupación de los recursos no tiene por qué coincidir con la duración de la sesión ya que esta última comprende la utilización de recursos en una o varias células.



- λ_r^h
La tasa de llegadas de peticiones de traspaso a una célula.
- P_r^n
La probabilidad de que una sesión nueva no sea admitida (probabilidad de pérdidas para sesiones nuevas).
- P_r^h
La probabilidad de que una petición de traspaso no sea admitida (ídem de traspasos).

Aunque el valor de λ_r^h se podría determinar a partir de un proceso iterativo que equilibrase los flujos de traspasos entrantes y salientes, por razones de sencillez computacional se asume que la tasa de llegada de traspasos es una fracción conocida de la tasa de llegada de peticiones de nuevas sesiones, lo que suele ser una práctica común en la literatura [Jab96]. Por otro lado, los objetivos de QoS B_r^n, B_r^h están expresados como cotas para las probabilidades de pérdidas de sesiones nuevas y de traspasos, respectivamente.

4.3 Políticas de control de admisión

En esta sección se define el comportamiento de las políticas de control de admisión estudiadas. Todas sin excepción fundamentan la decisión de aceptación o rechazo de una petición en el estado en el que se encuentra el sistema en el instante en que llega la petición³ y de la propia clase de peticiones. Las clases de peticiones se definen a partir del servicio que las genera y de la naturaleza de las mismas (bien solicitud de nueva sesión o bien de traspaso). Así el número total de clases de peticiones ascenderá a $2R$.

³Recuérdese que debido a la memoria nula del modelo, no se tiene en cuenta la historia precedente del sistema.

4.3.1 Políticas con solución en forma de producto

CS (*Complete Sharing*)

En la política más simple de todas: una petición se acepta a condición de que existan suficientes recursos disponibles en el sistema. Así pues, el conjunto de recursos es de acceso común para todas peticiones y sólo se rechazan peticiones del servicio r si los recursos totales disponibles son inferiores a c_r unidades de recurso.

IL (*Integer Limit/Upper Limit*)

Se asocian dos parámetros, $l_r^n, l_r^h \in \mathbb{N}$ a cada servicio r . Se acepta una sesión nueva | traspasada del servicio r que llega en el estado x si $x_r^n < l_r^n$ | $x_r^h < l_r^h$, respectivamente, y se rechaza en caso contrario. Equivale a establecer una cota superior para el número de sesiones que cada clase de peticiones puede tener en curso simultáneamente en el sistema.

FL (*Fractional Limit*)

Se asocian dos parámetros, $q_r^n, q_r^h \in [0, 1]$ a cada servicio r . Se acepta una sesión nueva | traspasada del servicio r con probabilidad q_r^n | q_r^h , respectivamente. Equivale a rechazar un porcentaje fijo del tráfico entrante de cada clase de peticiones.

GM (*Guaranteed Minimum*)

Se asocian dos parámetros, $s_r^n, s_r^h \in \mathbb{N}$ a cada servicio r . El parámetro indica el número mínimo garantizado de sesiones que el sistema puede cursar a la vez de una cierta clase de peticiones (lo que implica la reserva en exclusiva de la correspondiente porción de recursos para cada clase). Si $\sum_{r=1}^R (s_r^n + s_r^h) c_r < C$ existe además $C - \sum_{r=1}^R (s_r^n + s_r^h) c_r$ unidades de recurso a disposición de todas las clases de peticiones para su uso compartido, (si $\sum_{r=1}^R (s_r^n + s_r^h) c_r = C$ los recursos se han repartido completamente entre las clases de peticiones y no existen recursos compartidos: es el caso de la política *Complete Partitioning*). Se acepta una sesión nueva | traspasada del servicio r que llega en el estado x

si $x_r^n < s_r^n \mid x_r^h < s_r^h$, respectivamente; si lo anterior no se cumple, se acepta la sesión si existen al menos c_r recursos libres en el conjunto compartido; de lo contrario se rechaza.

Adicionalmente se estudian todas las posibles combinaciones de las anteriores políticas.

FL+IL (*Fractional and Integer Limits*)

Se asocian cuatro parámetros $l_r^n, l_r^h \in \mathbb{N}$, $q_r^n, q_r^h \in [0, 1]$ a cada servicio r . Se acepta con probabilidad $q_r^n \mid q_r^h$ una sesión nueva | traspasada del servicio r que llega en el estado x si $x_r^n < l_r^n \mid x_r^h < l_r^h$, respectivamente, y en caso contrario se rechaza.

IL+GM (*Integer Limit with Guaranteed Minimum*)

Se asocian cuatro parámetros $s_r^n, s_r^h, l_r^n, l_r^h \in \mathbb{N}$ a cada servicio r . Se acepta una sesión nueva | traspasada del servicio r que llega en el estado x si $x_r^n < s_r^n \mid x_r^h < s_r^h$; si lo anterior no se cumple, se acepta la sesión si existen suficientes recursos disponibles en el conjunto compartido y $x_r^n < l_r^n \mid x_r^h < l_r^h$; de lo contrario, se rechaza.

FL+GM (*Fractional Limit with Guaranteed Minimum*)

Se asocian cuatro parámetros $s_r^n, s_r^h \in \mathbb{N}$, $q_r^n, q_r^h \in [0, 1]$ a cada servicio r . Se acepta una sesión nueva | traspasada del servicio r con probabilidad $q_r^n \mid q_r^h$ que llega en el estado x si $x_r^n < s_r^n \mid x_r^h < s_r^h$; si lo anterior no se cumple, se acepta la sesión con probabilidad $q_r^n \mid q_r^h$ si existen suficientes recursos disponibles en el conjunto compartido; de lo contrario, se rechaza.

FL+IL+GM (*Fractional and Integer Limits with Guaranteed Minimum*)

Se asocian seis parámetros $s_r^n, s_r^h, l_r^n, l_r^h \in \mathbb{N}$, $q_r^n, q_r^h \in [0, 1]$ a cada servicio r . Se acepta una sesión nueva | traspasada del servicio r con probabilidad $q_r^n \mid q_r^h$ que llega en el estado x si $x_r^n < s_r^n \mid x_r^h < s_r^h$; si lo anterior no se cumple, se acepta la sesión con probabilidad $q_r^n \mid q_r^h$ si existen suficientes recursos disponibles en el conjunto compartido y $x_r^n < l_r^n \mid x_r^h < l_r^h$; de lo contrario, se rechaza.

4.3.2 Políticas de *trunk reservation*

MGC (*Multiple Guard Channel*)

Se asocian dos parámetros $l_r^n, l_r^h \in \mathbb{N}$ a cada servicio r . Se acepta una sesión nueva | traspasada del servicio r que llega en el estado x si $c(x) + c_r \leq l_r^n$ | $c(x) + c_r \leq l_r^h$, respectivamente, y se rechaza en caso contrario. Equivale a establecer un umbral respecto a la ocupación del sistema, por encima del cual se rechazan las sesiones de dicha clase de peticiones.

MFGC (*Multiple Fractional Guard Channel*)

Se asocian dos parámetros $t_r^n, t_r^h \in \mathbb{R}$ a cada servicio r . Se acepta una sesión nueva | traspasada del servicio r que llega en el estado x si $c(x) + c_r \leq \lfloor t_r^n \rfloor$ | $c(x) + c_r \leq \lfloor t_r^h \rfloor$, respectivamente, donde « $\lfloor x \rfloor$ » denota al entero inmediatamente inferior a x ; si lo anterior no se cumple, se acepta con probabilidad $t_r^n - \lfloor t_r^n \rfloor$ | $t_r^h - \lfloor t_r^h \rfloor$ si $c(x) + c_r = \lfloor t_r^n \rfloor + 1$ | $c(x) + c_r = \lfloor t_r^h \rfloor + 1$; de lo contrario, se rechaza.

4.3.3 Otras políticas

RS (*Randomized Stationary*)

A cada estado x del sistema se asocian dos R -tuplas $q^n(x), q^h(x) \in [0, 1]^R$. Se acepta una sesión nueva | traspasada del servicio r que llega en el estado x con probabilidad $q_r^n(x)$ | $q_r^h(x)$. Nótese que mientras que las políticas anteriores son casos particulares de esta política, existen políticas RS que no pertenecen a ninguna de las familias anteriores (ver Sección 2.2.2 y Fig. 2.2).

4.4 Diseño y Dimensionado

Existen dos maneras de trabajar con políticas de control de admisión: llevar a cabo un «proceso de análisis» o realizar un «proceso de síntesis» (ver Fig. 4.1).

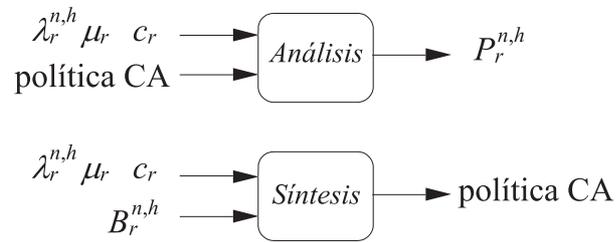


Figura 4.1: Dos procedimientos de trabajo con políticas de control de admisión.

Se denomina proceso de análisis a la función Φ que a partir de los parámetros que definen al sistema y el valor de los parámetros de configuración de la política de control de admisión, devuelve las probabilidades de pérdidas de cada clase de peticiones:

$$P_r^{n,h} = \Phi(\lambda_r^{n,h}, \mu_r, c_r, C, \text{política CA}) \quad (4.1)$$

El proceso de análisis es necesario para caracterizar el comportamiento de un sistema con una política de admisión concreta. Sin embargo, en el diseño o ajuste de políticas el problema es el inverso; a este respecto se denomina proceso de síntesis a la función Γ que a partir de los parámetros que definen el sistema y los objetivos de probabilidad de pérdidas, devuelve el valor de los parámetros de configuración de la política de control de admisión:

$$\text{política CA} = \Gamma(\lambda_r^{n,h}, \mu_r, c_r, C, B_r^{n,h}) \quad (4.2)$$

La aproximación común para afrontar el proceso de síntesis es mediante la ejecución iterativa de un proceso de análisis.

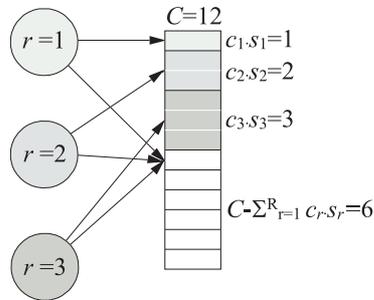


Figura 4.2: Sistema de contraejemplo empleando política GM.

4.4.1 Análisis de políticas

Políticas CS, IL

Para el análisis de las políticas CS e IL el enfoque aplicado es el detallado en la Sección 3.3 mediante el empleo del «algoritmo de la convolución» de [Ive87, Ive02].

Política GM

La existencia de una solución en forma de producto no asegura la aplicabilidad directa del algoritmo de la convolución. Este es el caso de la política GM que ilustra el siguiente contraejemplo. Sea el sistema fijo con parámetros: $C = 12$, $R = 3$, $c_r = \{1, 2, 3\}$, política GM con $s_r = \{1, 1, 1\}$ —se deduce que la cantidad de recursos dedicados a uso común asciende a $C - \sum_{r=1}^R s_r c_r = 6$, es decir: el número máximo de peticiones que se pueden cursar por servicio es $\{7, 4, 3\}$ respectivamente (ver Fig. 4.2)—.

La Fig. 4.3 muestra una representación tridimensional de todos los posibles valores que puede tomar el vector x que describe el estado del sistema.

Se pasa a calcular ahora las probabilidades de estado del sistema de dos maneras:

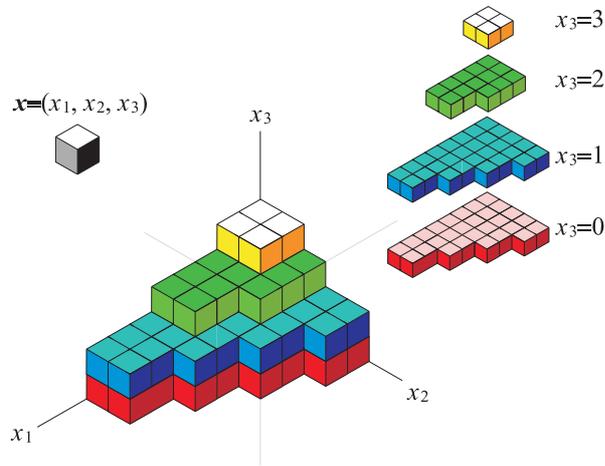
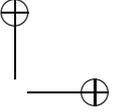
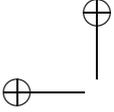


Figura 4.3: Representación del espacio de estados del sistema de contraejemplo.

- Mediante el algoritmo de la convolución.
- Mediante un estudio pormenorizado del espacio de estados.

El algoritmo de la convolución determina que por sucesivas convoluciones de los vectores probabilidad de estado normalizados $\{p_r(i)\}$ de cada clase de peticiones considerada aisladamente (en cualquier orden) se obtiene el vector $q_N = \{q_N(0), \dots, q_N(C)\}$ que representa las probabilidades de estado del sistema. Sin embargo, al comparar la convolución con el estudio detallado del espacio de estados, la convolución devuelve un valor erróneo (por exceso) para $q_N(11)$ y $q_N(12)$. Se subrayan a continuación los sumandos innecesarios:

$$\begin{aligned}
 q'_N(11) = & p_1(0)p_2(1)p_3(3) + \underline{p_1(2)p_2(0)p_3(3)} + p_1(1)p_2(2)p_3(2) + \\
 & + p_1(3)p_2(1)p_3(2) + \underline{p_1(5)p_2(0)p_3(2)} + p_1(0)p_2(4)p_3(1) + p_1(2)p_2(3)p_3(1) + \\
 & + p_1(4)p_2(2)p_3(1) + p_1(6)p_2(1)p_3(1) \quad (4.3)
 \end{aligned}$$



$$\begin{aligned}
 q'_N(12) = & p_1(1)p_2(1)p_3(3) + \underline{p_1(3)p_2(0)p_3(3)} + \underline{p_1(0)p_2(3)p_3(2)} + \\
 & + p_1(2)p_2(2)p_3(2) + p_1(4)p_2(1)p_3(2) + \underline{p_1(6)p_2(0)p_3(2)} + p_1(1)p_2(4)p_3(1) + \\
 & + p_1(3)p_2(3)p_3(1) + p_1(5)p_2(2)p_3(1) + p_1(7)p_2(1)p_3(1) \quad (4.4)
 \end{aligned}$$

Los términos sobrantes son absurdos —no se corresponden con un valor x perteneciente al espacio de estados— y son la consecuencia de que la convolución no tenga en cuenta el impacto de las reservas del resto de servicios sobre el número máximo de peticiones que puede cursar simultáneamente un servicio concreto. Estos términos sólo tienen sentido si no existiera una porción de recursos de uso común, es decir, si los recursos se hubieran repartido totalmente (política *Complete Partitioning*), lo cual no siempre es cierto para la política GM. La manera de resolver este inconveniente pasa por identificar en los elementos de la convolución todos aquellos términos que no pertenecen al espacio de estados para descartarlos posteriormente.

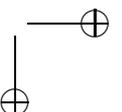
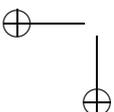
Política FL

En [LA95] se justifica la solución en forma de producto para una clase de políticas que incluye a FL. Sin pérdida de generalidad, para un sistema fijo, las probabilidades de estado resultan

$$p(\mathbf{x}) = \frac{1}{G} w(\mathbf{x}) \prod_{r=1}^R \frac{(A_r)^{x_r}}{x_r!} \quad (4.5)$$

donde G es una constante de normalización, $A_r = \lambda_r / \mu_r$ y $w(\mathbf{x})$ es un factor que debe cumplir unas ciertas premisas para que la probabilidad de estado tenga solución en forma de producto. Por ejemplo, en el caso de la política FL, (sistema fijo), dicho factor equivale a

$$w(\mathbf{x}) = \prod_{k=1}^R (q_k)^{x_k}. \quad (4.6)$$



donde se asume que en el caso extremo⁴ de que $q_k = 0$, $x_k = 0$ entonces $(q_k)^{x_k} = 1$.

Para emplear el algoritmo de la convolución se procede de la siguiente manera: se denota como $\{p_r(i)\}$ y $\{p_r^{\text{FL}}(i)\}$, $i = 0, \dots, C$, a los vectores probabilidad de estado normalizados en ausencia y en presencia de la política FL. Si se define ahora el vector auxiliar $\{p'_r(i)\}$, donde $p'_r(i) = p_r(i)(q_r)^i$; entonces $\{p_r^{\text{FL}}(i)\}$ equivale al vector $\{p'_r(i)\}$ normalizado. A partir de aquí se puede emplear $\{p_r^{\text{FL}}(i)\}$ en el papel de $\{p_r(i)\}$ y aplicar la metodología explicada en la Sección 3.3.

Políticas MGC, MFGC

Para el análisis de las políticas MGC y MFGC la mecánica empleada es la habitual. Fijados los parámetros del sistema y la configuración de la política (umbrales l_r^n, l_r^h ó t_r^n, t_r^h respectivamente), se plantean las ecuaciones de balance globales del proceso de Markov de donde se obtienen los valores para los parámetros de QoS. Al no existir soluciones en forma de producto, funciones generatrices basadas en transformadas u observaciones astutas que conduzcan a soluciones elegantes para esta familia de políticas, se debe recurrir a métodos numéricos [Coo81]. El tipo de ecuaciones a resolver resulta en un esquema de iteración relativamente sencillo para el que, aunque se conocen ciertos criterios de convergencia, estos son, a menudo, difíciles de aplicar en la práctica. El esquema se detalla a continuación:

Sea el sistema de ecuaciones lineales

$$Ax = b \tag{4.7}$$

⁴«Algunos libros de texto dejan sin definir la cantidad 0^0 , debido a que las funciones x^0 y 0^x tienen diferentes valores límite cuando $x \rightarrow 0$. Sin embargo, esto es un error. Se debe definir $x^0 = 1$ para todo x , si el teorema del binomio tiene que mantener su validez cuando $x = 0$, $y = 0$, y/o $x = -y$. ¡Este teorema es demasiado importante como para limitarlo de manera arbitraria! En contraste, la función 0^x carece de bastante importancia.» Ronald Graham, Donald Knuth, Oren Patashnik, *Concrete Mathematics*, 2nd Ed., Addison-Wesley, 1994.

donde A es una matriz cuadrada, b es un vector y x es un vector incógnita. Si se identifica (4.7) con el conjunto de ecuaciones de balance de estado, entonces x son las probabilidades de estado en equilibrio a determinar. Así (4.7) se puede reescribir como

$$(I - L - U)x = d \quad (4.8)$$

donde L, U son respectivamente, una matriz triangular inferior y superior con ceros a lo largo de la diagonal principal, e I es la matriz identidad, con unos a lo largo de la diagonal principal y ceros en cualquier otro lugar. Si se reescribe (4.8) como

$$x = Lx + Ux + d \quad (4.9)$$

lo que sugiere el siguiente esquema de iteración

$$x^{(n+1)} = Lx^{(n+1)} + Ux^{(n)} + d \quad (4.10)$$

donde el vector $x^{(n+1)}$ es la iteración n -ésima, $n = 0, 1, \dots$ (el x^0 inicial es un vector arbitrario con norma euclídea unidad, $\|x^0\| = 1$). La Ecuación 4.10 sugiere calcular cada componente de $x^{(n+1)}$ a partir de los últimos valores disponibles, que son todas las componentes de $x^{(n+1)}$ calculadas hasta el momento y las restantes componentes de $x^{(n)}$ calculadas en la anterior iteración. Al procedimiento de iteración de (4.10) se le han asociado los nombres de Seidel, Gauss, Liebmann, Nekrasov y probablemente varios más. La técnica general para resolver ecuaciones mediante iteración también se denomina «aproximación sucesiva» y «relajación». El nombre más común para este esquema es el de «iteración de Gauss-Seidel».

Se puede demostrar que una condición necesaria y suficiente para la convergencia del procedimiento de iteración para un x^0 inicial arbitrario es que todos los valores propios de la matriz de iteración M , $M = (I - L)^{-1}U$ residan dentro del círculo unitario.⁵ Este criterio de convergencia, en general es sólo de interés teórico, porque encontrar el radio espectral de la matriz de iteración es de ordinario, del mismo orden de dificultad que resolver el sistema

⁵Es decir: el procedimiento de iteración converge independientemente del vector inicial x^0 si y sólo si el radio espectral de la matriz de iteración es menor que la unidad.

de ecuaciones 4.7. Para ciertos casos especiales, se puede predecir el comportamiento del esquema de iteración a través de una inspección directa de la matriz A . En particular, una condición suficiente para la convergencia es que A sea «irreducible» y exhiba un comportamiento de «dominancia diagonal débil» [Coo81]. Esta condición, en general no se verifica para las ecuaciones de probabilidad de estados, pero sugiere por qué funciona el esquema de iteración: cuanto mayor sea la concentración de «masa» en la diagonal principal de la matriz A , el esquema convergerá con mayor velocidad y precisión. De hecho las ecuaciones de balance que se deben resolver para las políticas MGC y MFGC se caracterizan por un gran número de ceros lejos de la diagonal principal, facilitando la convergencia y convirtiendo en relativamente sencilla la programación de la correspondiente iteración de Gauss-Seidel.

4.4.2 Síntesis de políticas

Las técnicas de análisis mencionadas para las políticas IL, GM, FL, MGC y MFGC no permiten sintetizar de manera directa cuál es la configuración óptima de una política concreta. Sin embargo, se pueden emplear indirectamente para, mediante una «búsqueda inversa», tratar de obtener dicha configuración. No obstante, el carácter multidimensional incrementa la dificultad de un problema que por el empleo de métodos iterativos ya era costoso en términos de complejidad computacional. A causa de estos inconvenientes se considera además la familia de políticas RS, que incorpora tanto los beneficios de ser una política que contiene a las anteriores (con lo que en teoría se pueden obtener prestaciones superiores) como la particularidad de se puede resolver empleando técnicas que realizan la síntesis directa de la política óptima.

Política RS

Esta sección resume brevemente la contribución de Pla et al. [PCG03]. En [PCG03] se hace uso de la teoría de los «procesos de decisión de Markov» (MDP) [Ros70] junto a técnicas de programación lineal para realizar una síntesis eficiente de

políticas RS. Esta política supera a políticas previas y se considera cota superior de prestaciones. La formulación del problema como MDP tiene la ventaja de que puede obtenerse la configuración óptima de la política a partir de los valores de los parámetros del sistema y los objetivos de QoS, sin necesidad de una búsqueda complementaria que trate de ajustar las probabilidades de bloqueo a los objetivos.

El criterio de diseño empleado está comprendido por las funciones objetivo —diseñadas a partir de funciones de coste cuya media temporal iguala la correspondiente probabilidad de pérdidas— y un conjunto de restricciones. Tanto las funciones objetivo como las restricciones son lineales, por lo que el problema de síntesis se puede formular como un programa lineal. De hecho se puede usar el método «simplex» y otros algoritmos usuales. En consonancia con el resto del capítulo, el criterio de diseño seleccionado es el que trata de minimizar la diferencia entre la probabilidad de bloqueo de una clase de peticiones y una cota superior definida para las mismas (el objetivo de QoS de dicha clase). En la práctica, podría no existir una solución satisfactoria si el valor de la cantidad de recursos del sistema, C , no es lo suficientemente elevado. Así, un problema común de la fase de planificación es encontrar el valor mínimo de C tal que existe una solución válida, o su equivalente: encontrar la intensidad de tráfico ofrecido máxima que garantiza la existencia de una solución válida para un C concreto. Ambos problemas se pueden resolver satisfactoriamente mediante la aplicación del criterio de diseño mencionado.

Resto de políticas

Para todas las políticas restantes —a excepción de la política RS— el procedimiento de síntesis consta de varias tareas anidadas entre sí:

1. Buscar la configuración óptima de la política de control de admisión (de entre el conjunto de todas las posibles configuraciones).⁶ Esta búsqueda

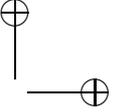
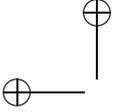
⁶Se recuerda que la configuración óptima es aquella capaz de cursar la intensidad de tráfico máxima que el sistema puede soportar en presencia de dicha política sin incumplir los objetivos de QoS establecidos, a dicha intensidad se la denomina «capacidad del sistema».

se realiza conforme a una determinada estrategia.

2. La búsqueda del punto 1 requiere conocer la cota superior para la intensidad de tráfico ofrecido que es capaz de soportar una configuración concreta de la política sin incumplir los objetivos de QoS. Averiguar esta cota superior necesita de un proceso iterativo similar al que se describe en la Sección 3.3.4.
3. Para conocer la cota superior del punto 2 es necesario averiguar si una determinada intensidad de tráfico ofrecido al sistema respeta (o no) los objetivos de QoS. Esto se consigue llevando a cabo un proceso de análisis (empleando el algoritmo de la convolución si procede o de lo contrario resolviendo el sistema de ecuaciones mediante un método iterativo).

Como se ha señalado, los sistemas multiservicio añaden al diseño de políticas el inconveniente de la dimensionalidad. A parte, las probabilidades de bloqueo experimentadas por los servicios no aumentan siempre de manera monótonica cuando se produce un incremento en la intensidad de tráfico ofrecido o un aumento en la rigurosidad de los requisitos de admisión establecidos por una política. No obstante, en cierto modo se puede reducir la complejidad si el operador es capaz de predecir con anticipación el «vector de actividad de cada clase de peticiones» f_r^n, f_r^h (concepto similar al vector de penetración de la Sección 3.2, desglosado ahora para las peticiones de sesiones nuevas y de traspaso) [HBBPGMIA00].

Cuando una política de control de admisión produce una solución en forma de producto y/o el número de recursos C del sistema es relativamente manejable, resulta factible realizar la búsqueda de la política óptima (punto 1) a partir de una búsqueda exhaustiva dentro del espacio de todas las configuraciones posibles de dicha política. Esta búsqueda exhaustiva se puede aplicar directamente a políticas cuyos parámetros de configuración son discretos (IL, GM, IL+GM, MGC). Para políticas con parámetros continuos (FL y sus combinaciones y MFGC), se puede emplear indirectamente esta



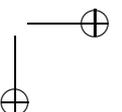
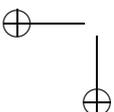
búsqueda —previa aplicación de los adecuados intervalos de discretización a cada uno de los parámetros—.

Debido a la dimensionalidad del problema, el coste computacional de su resolución no se escala beneficiosamente con el tamaño del mismo (a destacar la influencia perjudicial que tiene el aumento de la cantidad de recursos C y/o el número de servicios R sobre el coste). Cuando bien la multidimensionalidad, el dominio continuo de los parámetros de configuración de las políticas o la cantidad de recursos del sistema impide abordar de manera exhaustiva la búsqueda de la política óptima, se hace necesario hacer uso de algún tipo de algoritmo heurístico que sea capaz de encontrar una «buena configuración» en una cantidad de tiempo razonable. La alternativa que se examina a continuación es la del algoritmo genético.

Empleo de un algoritmo genético

Un «algoritmo genético», (AG) es una clase de método de búsqueda estocástico que se aplica con bastante éxito a la resolución de problemas de optimización multidimensionales (ver por ejemplo [SHNK00] y las referencias mencionadas en él). Mientras que la mayoría de los métodos de optimización heurísticos trabajan sobre una única solución candidata, los AG operan sobre un conjunto de soluciones. Cada solución se codifica en una estructura de datos llamada «cromosoma» que incluye los valores que las variables a optimizar. Los AG mantienen este conjunto de soluciones y aplican «operadores genéticos» sobre las mismas (operadores de «cruzamiento» y mutación) con el objetivo de generar nuevos cromosomas en cada iteración «generación» del AG. Se utilizan varios criterios de selección para elegir los cromosomas que se emparejarán entre sí. En la mayoría de los casos, la selección se basa en una función de «idoneidad» que se diseña de acuerdo con la función objetivo que se desea optimizar.

La mayoría de los AG utilizan cadenas de bits para codificar las soluciones candidatas, sin embargo, para muchas aplicaciones como la que aquí se trata, resulta más natural el emplear números en coma flotante. Dependiendo



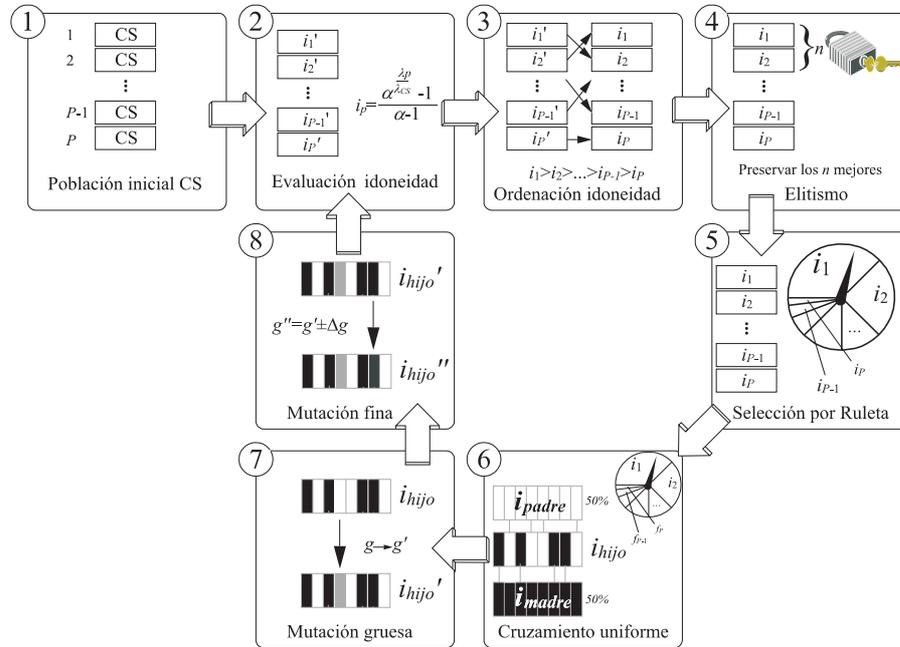


Figura 4.4: Una generación del algoritmo genético, paso a paso.

de la política de control de admisión, el cromosoma contiene unos determinados «genes» que representan el valor correspondiente de los parámetros de configuración de la política ($l_r^n, l_r^h, s_r^n, s_r^h, q_r^n, q_r^h, t_r^n, t_r^h$ dependiendo de la misma).

La estructura del algoritmo genético se detalla en la Fig. 4.4. Los pasos son los siguientes:

- 1. Población inicial** La población inicial (primera generación) consta de P cromosomas (es decir de soluciones candidatas, de 10 a 100 en este caso). En este caso todos los cromosomas representan la política CS. Existen dos razones para ello:

- (a) Intuitivamente cabe pensar que en la mayoría de los casos, la polí-

tica óptima esté más cerca de la política de CS, que de una política de rechazo total.

(b) La idoneidad se define en relación con la capacidad ofrecida por la política CS, (ver siguiente punto).

2. **Evaluación de la idoneidad** Las prestaciones de la política CS se definen en términos de la capacidad λ_{CS} (máxima intensidad de tráfico que se puede ofrecer a dicha política sin incumplir los objetivos de QoS). La idoneidad de un cromosoma se define como

$$i_p = \frac{\lambda_p}{\alpha^{\lambda_{CS}} - 1} \quad (4.11)$$

donde λ_p es la capacidad de la solución representada por el cromosoma p -ésimo, el parámetro α sirve para ponderar la importancia relativa que tienen las soluciones prometedoras respecto a las soluciones mediocres [Mic94]. Un valor de $\alpha = 10$ resulta adecuado.

3. **Ordenación por idoneidad** Tras la evaluación de la idoneidad, se procede a la ordenación de los cromosomas que integran la población a partir de su idoneidad.
4. **Elitismo** Se preservan los n cromosomas con mayor idoneidad para evitar que los operadores genéticos (mutación y cruzamiento) dañen las soluciones más prometedoras. Estos cromosomas no serán substituidos por cromosomas nuevos creados en la siguiente generación. En este caso $n = 1$ para preservar al menos la mejor solución a través de todas las generaciones.
5. **Selección por ruleta** Tras cada iteración del algoritmo genético, la población de cromosomas de la generación anterior se ve substituida por una nueva población de cromosomas (excepto los preservados por el criterio elitista). Cada nuevo cromosoma se construyen a partir de dos cromosomas antiguos (progenitores), seleccionados de manera «proporcional a su idoneidad», es decir: el número medio de veces en que un

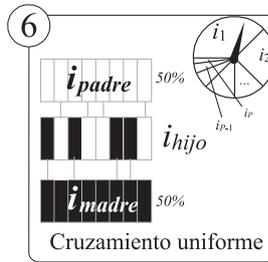


Figura 4.5: Detalle del procedimiento de cruzamiento uniforme.

cromosoma está presente como progenitor coincide con la idoneidad del mismo dividida por la suma de idoneidades de todos los cromosomas de la población. La manera más usual de llevar a cabo este tipo de selección es mediante una selección aleatoria basada en una «ruleta» conceptual, en la que cada cromosoma posee un sector circular proporcional a la idoneidad del mismo. Así, la ruleta se gira tantas veces como sea necesario hasta generar todas las parejas de progenitores.

6. **Cruzamiento uniforme** Se construye cada nuevo cromosoma a partir de la aplicación del operador de cruzamiento a los dos cromosomas progenitores. Se opta por el «cruzamiento uniforme» en el que cada progenitor tiene la mitad de posibilidades de ser elegido para transmitir un determinado gen al cromosoma hijo. La Fig. 4.5 muestra al cromosoma hijo entre sus dos progenitores (arriba y abajo), donde el color de un determinado gen (blanco o negro) indica la procedencia del mismo. La idea detrás del operador de cruzamiento es la de tratar de transmitir a la nueva generación fragmentos de aquellos cromosomas con mayor idoneidad (bajo la idea de que dichos fragmentos son el origen de dicha idoneidad).
7. **Mutación gruesa** Los operadores de mutación evitan que la evolución de la población se estanque. La mutación gruesa es la encargada de substituir un determinado gen g de un cromosoma de la nueva población, substituyéndolo por otro valor g' , ver Fig. 4.6(a).

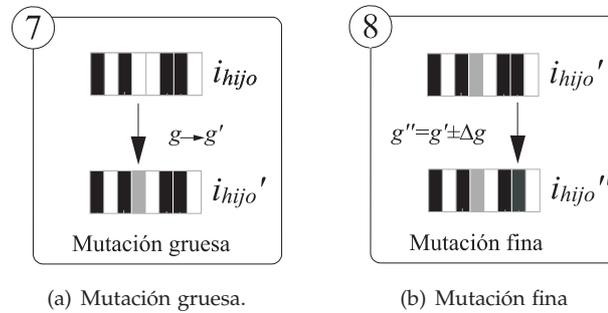


Figura 4.6: Detalle de los procedimientos de mutación.

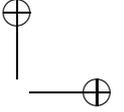
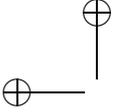
8. **Mutación fina** La mutación fina modifica un determinado gen g , incrementándolo o disminuyéndolo en una cierta cantidad Δg , ver Fig. 4.6(b). Ambos operadores tienen asociados un parámetro de incidencia de la mutación (valores típicos de por ejemplo $p_g = 0.3$, $p_f = 1.0$) respectivamente, que indica la probabilidad de que se aplique cada operador de mutación a un cromosoma nuevo determinado.

9. **Ir al paso 2.**

Todos los valores para los parámetros del AG están seleccionados para lograr un equilibrio entre la velocidad de convergencia y la bondad del subóptimo obtenido, un compromiso siempre presente cuando se emplea esta herramienta de optimización [Mic94]. Los resultados expuestos en la Sección 4.5.1 procedentes del AG, se corresponden a una población media de 20 cromosomas con tasas de mutación gruesa y fina de $p_g = 0.05$, $p_f = 1.0$ respectivamente, donde la convergencia se logró tras un número de generaciones que depende de la cantidad de recursos del sistema (en torno a las decenas cuando $C = 10$ y del orden de centenas cuando $C = 40$). Nótese, además, que políticas de CA cuyos parámetros de configuración son enteros (como IL, GM, MGC) necesitan significativamente menos generaciones que políticas cuyos parámetros pueden tomar cualquier valor dentro de un intervalo continuo (FL, MFGC).

4.5 Comparativa de políticas

El estudio de Ramjee et al. [RTN97] demuestra que para un escenario mono-servicio las políticas *Guard Channel* y *Limited Fractional Guard Channel* (de la familia *trunk reservation*), son óptimas para funciones objetivo de QoS usuales. Los resultados obtenidos por Altman et al. [AJK01] para escenarios multiservicio podrían dar a entender que también una política de *trunk reservation* es la óptima en estos casos, pero sus resultados no son generalizables. Es finalmente, en [BC02] donde se demuestra que la política de admisión óptima, en términos de minimización de una cierta función objetivo, pertenece a la clase de políticas estacionarias aleatorizadas. En este sentido, durante el estudio de las políticas de CA en entornos móviles celulares multiservicio, una cuestión importante que surge es en qué medida son comparables las prestaciones de las políticas con solución en forma de producto (cuya resolución mediante el algoritmo de la convolución se beneficia de un menor esfuerzo computacional) con respecto a las políticas de *trunk reservation* (como se ha visto, en ciertos casos, cercanas a la política óptima). El trabajo previo en escenarios multiservicio fijos de Biswas et al. [BS97] confirma que las políticas *Complete Partitioning*, *Guaranteed Minimum*, *Upper Limit* y *Guard Channel* superan en prestaciones a la política *Complete Sharing* en términos de maximizar una función de ganancia cuando las cotas se imponen sobre la probabilidad de bloqueo de cada clase de peticiones y existen además diferencias significativas entre las clases en términos de demandas de recursos e intensidad de tráfico ofrecido. En sus comparaciones, además, constatan el beneficio añadido de emplear *Guard Channel* cuando los perfiles de las clases son muy dispares (por ejemplo, cuando un clase ofrece relativamente menor intensidad de tráfico que las demás, pero a cambio exige, comparativamente, mayores recursos y menor probabilidad de bloqueo que el resto de clases).



4.5.1 Resultados numéricos

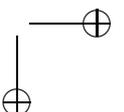
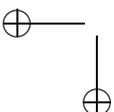
Capacidad del sistema

En esta sección se recogen los resultados de evaluar la capacidad del sistema en presencia de las diferentes políticas de CA mencionadas en la Sección 4.3 optimizadas mediante búsqueda exhaustiva (políticas IL, GM y MGC), mediante el empleo del algoritmo genético descrito en la Sección 4.4.2 (políticas FL, FL+IL, IL+GM, FL+GM, FL+IL+GM y MFGC) y a partir de los resultados recogidos en Pla et al. [PCG04] (para la política RS). Se recuerda que la capacidad del sistema en cada caso se define como $\lambda_{\text{máx}} = \sum_{r=1}^R \lambda_{r(\text{máx})}^n$, es decir: la máxima tasa agregada de llegada de peticiones de nuevas sesiones procedentes de todos los servicios que el sistema puede cursar sin incumplir los objetivos de QoS establecidos en términos de la probabilidad de pérdidas (B_r^n, B_r^h). La capacidad se ha obtenido para los cinco escenarios diferentes (A–E) definidos por las configuraciones de la Tabla 4.1. Los valores de la capacidad se muestran en la tabla 4.2 y están expresados como valores relativos respecto a la capacidad obtenida por la política CS —de la que se muestra la capacidad en valor absoluto—. ⁷ La intensidad de tráfico que se puede ofrecer por cada clase de peticiones se determina directamente a partir del valor de $\lambda_{\text{máx}}$ y los valores del vector de actividad de cada clase de peticiones f_r^n, f_r^h .

Como se observa, la ganancia relativa disminuye cuando el número de recursos C aumenta. Las políticas con solución en forma de producto, en general, no ofrecen tanta capacidad como el resto de políticas estudiadas. Los rangos de ganancia de las políticas de *trunk reservation* oscilan entre el 5% y el 15%, aproximadamente. Finalmente, es notable la similitud que existe entre la capacidad que se puede obtener con la mejor política *MFGC* y la obtenida con la política óptima *RS*.

En general, la complejidad de las políticas estudiadas no es un factor limi-

⁷Por razones de brevedad, políticas con solución en forma de producto fruto de combinaciones de políticas más elementales se nombran en la tabla mediante las abreviaturas C1, C2, C3, C4, haciendo referencia a las políticas FL+IL, FL+GM, IL+GM, FL+IL+GM respectivamente.



Cuadro 4.1: Parámetros de las configuraciones

	Configuración				
	A	B	C	D	E
c_1 (recursos)	1	1	1	1	1
c_2 (recursos)	2	4	2	2	2
f_1	0.8	0.8	0.2	0.8	0.8
f_2	0.2	0.2	0.8	0.2	0.2
B_1^n (%)	5	5	5	1	1
B_2^n (%)	1	1	1	2	1
	A,B,C,D,E				
B_r^h	$0.1B_r^n(max)$				
λ_r^n (llamadas/s)	$f_r\lambda$				
λ_r^h (llamadas/s)	$0.5\lambda_r^n$				
μ_1 (s^{-1})	1				
μ_2 (s^{-1})	3				

Cuadro 4.2: Capacidad (λ_{max} en sesiones/s)

Conf.	C	CS	IL	GM	FL	C1	C2	C3	C4	MGC	MFGC	RS
A	10	1.540	1.13	1.07	1.13	1.14	1.18	1.17	1.18	1.23	1.33	1.34
	20	5.614	1.11	1.06	1.12	1.13	1.15	1.14	1.16	1.26	1.31	1.31
	40	15.760	1.09	1.07	1.10	1.10	1.14	1.13	1.14	1.24	1.25	1.26
B	10	0.366	1.05	1.00	1.18	1.18	1.18	1.05	1.19	1.10	1.15	1.20
	20	2.779	1.03	1.08	1.11	1.11	1.12	1.09	1.13	1.21	1.25	1.25
	40	10.385	1.06	1.04	1.07	1.10	1.09	1.06	1.10	1.21	1.22	1.23
C	10	1.366	1.07	1.05	1.07	1.09	1.10	1.08	1.10	1.11	1.21	1.22
	20	5.772	1.05	1.05	1.08	1.08	1.07	1.06	1.09	1.20	1.21	1.21
	40	17.615	1.05	1.04	1.07	1.07	1.07	1.06	1.07	1.15	1.16	1.16
D	10	1.744	1.03	1.00	1.06	1.06	1.04	1.03	1.06	1.13	1.16	1.17
	20	6.047	1.04	1.00	1.05	1.05	1.05	1.04	1.05	1.13	1.15	1.15
	40	16.543	1.03	1.00	1.04	1.04	1.00	1.03	1.04	1.10	1.11	1.11
E	10	1.540	1.05	1.00	1.06	1.06	1.05	1.05	1.06	1.13	1.17	1.17
	20	5.615	1.04	1.02	1.05	1.05	1.06	1.05	1.06	1.13	1.15	1.16
	40	15.760	1.03	1.02	1.04	1.04	1.05	1.04	1.05	1.10	1.10	1.10

Cuadro 4.3: Configuración de políticas optimizadas con búsqueda exhaustiva

Conf.	C	$l_1^n c_1, l_1^h c_1, l_2^n c_2, l_2^h c_2$	$s_1^n c_1, s_1^h c_1, s_2^n c_2, s_2^h c_2$	$l_1^n, l_1^h, l_2^n, l_2^h$
A	10	4, 4, 4, 8	0, 0, 0, 2	7, 8, 9, 10
	20	9, 8, 6, 20	0, 0, 0, 2	16, 18, 19, 20
	40	19, 15, 10, 40	0, 0, 0, 4	35, 38, 38, 40
B	10	2, 3, 8, 8	0, 0, 0, 0	6, 6, 9, 10
	20	6, 6, 12, 20	0, 0, 0, 4	12, 16, 17, 20
	40	13, 11, 16, 20	0, 0, 4, 8	30, 36, 37, 40
C	10	2, 3, 6, 10	0, 0, 0, 2	6, 8, 8, 10
	20	4, 4, 12, 20	0, 0, 4, 6	15, 19, 18, 20
	40	7, 7, 22, 40	0, 0, 0, 10	33, 38, 38, 40
D	10	6, 6, 4, 10	0, 0, 0, 0	8, 10, 9, 10
	20	11, 10, 6, 20	0, 0, 0, 0	18, 20, 18, 20
	40	22, 18, 10, 12	0, 0, 0, 0	38, 40, 38, 40
E	10	5, 5, 4, 10	0, 0, 0, 0	8, 9, 9, 10
	20	11, 9, 6, 20	0, 3, 0, 2	17, 19, 18, 20
	40	22, 17, 10, 40	0, 0, 0, 2	38, 39, 38, 40

tante a la hora de su implementación porque, como máximo, sólo necesitan almacenar un reducido conjunto de parámetros por cada servicio. En el caso peor, para la política RS, puede demostrarse que el número máximo de variables que se necesitan almacenar equivalen al número de estados más $4R$ [PCG03].

Configuraciones óptimas

La Tabla 4.3 muestra los valores que toman los parámetros de configuración de aquellas políticas optimizadas mediante búsqueda exhaustiva (IL, GM, MGC).⁸ Se aprecia que en la mayoría de los casos, la configuración más adecuada de la política trata de proteger a las peticiones de traspaso del ser-

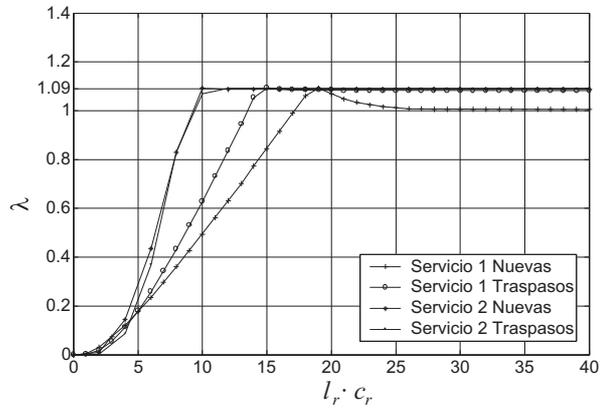
⁸Nótese que los valores de $l_r^{n,h}, s_r^{n,h}$ para las políticas IL y GM respectivamente, están expresados en número de sesiones en curso de la clase de peticiones correspondiente. Dado que el parámetro $l_r^{n,h}$ de la política MGC está expresado en unidades de recurso, en la tabla se muestran los valores correspondientes a los parámetros de configuración de IL y GM, expresados también en unidades de recurso, para facilitar la comparación entre ambas familias de políticas.

vicio 2. En el caso de la política IL, se limita el número de sesiones en curso permitidas al resto de clases de petición; en el de la política GM se reserva explícitamente una cantidad de recursos para las mismas; finalmente, en el caso de la política MGC cuando los recursos libres son escasos se limita el acceso a los mismos por parte del resto de clases de petición.

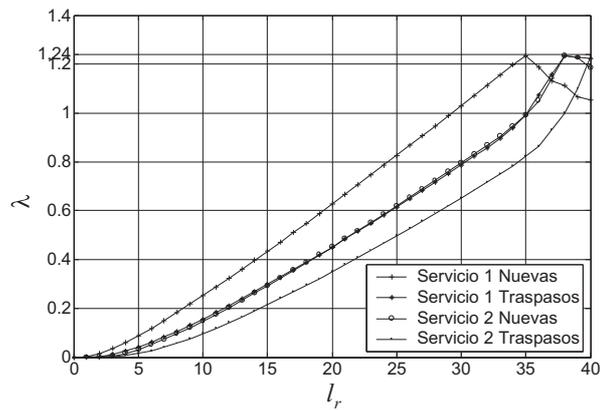
4.5.2 Sensibilidad

Tolerancias en la configuración

Se estudia la sensibilidad frente a tolerancias en los parámetros de configuración de las políticas IL y MGC, como representantes respectivas de políticas con solución en forma de producto y de políticas de *trunk reservation*. La Fig. 4.7(a) muestra la variación de la tasa máxima agregada de llegadas de peticiones de sesiones nuevas λ que puede cursar el sistema en presencia de la política IL sin incumplir los requisitos del QoS. La tasa máxima está expresada de manera relativa a la capacidad de la política CS. Cada trazo corresponde a una configuración de la política en la que se varía el parámetro referido en la leyenda a la vez que se mantienen constantes el resto de parámetros en sus valores óptimos (ver Tabla 4.3). Los resultados son referidos al escenario A de la Tabla 4.1 y $C = 40$ unidades de recurso. Nótese que el valor supremo en todos los casos se corresponde con la capacidad del sistema $\lambda_{\text{máx}} = 1.09$ para la política IL, que se alcanza en cada caso para el valor concreto recogido en la Tabla 4.3, es decir: $l_1^n = 19$, $l_1^h = 15$, $l_2^n = 10$, $l_2^h = 40$. Como se aprecia, $\lambda = \lambda_{\text{máx}}$ de manera significativamente estable dentro de ciertos rangos para los parámetros de la política. Por el contrario, parece que λ es bastante sensible al valor que pueda tomar l_1^n ; esto es debido, principalmente, a que la limitación que ejerce la política IL sobre las peticiones del servicio 1 nuevas redundando beneficiosamente sobre el resto de clases en términos de recursos libres a disposición de las mismas. Este efecto beneficioso se justifica porque la clase de peticiones nuevas del servicio 1 tiene unos objetivos de QoS no demasiado rigurosos. La Fig. 4.7(b) muestra



(a) Política IL



(b) Política MGC

Figura 4.7: Sensibilidad de la tasa máxima a variaciones en los parámetros de configuración de la política.

un estudio análogo al anterior, pero para la política MGC. La capacidad del sistema relativa a la política CS es ahora de $\lambda_{\text{máx}} = 1.24$. A diferencia de la política IL, los trazos para la política MGC describen una forma angulosa con el valor supremo localizado en el vértice del mismo. También conviene

observar que no muy lejos de este vértice, la tasa máxima que puede cursar el sistema se degrada significativamente incluso en comparación con la política CS. Este comportamiento parece sugerir que la especificación de los parámetros de la política MGC requiere de una mayor precisión en comparación con la política IL. Nótese finalmente que la configuración óptima de la política MGC pertenece a las inmediaciones de la política CS, (es decir, $l_1^n = l_1^h = l_2^n = l_2^h = C$) este último indicio queda más patente si se revisa la columna correspondiente a la política MGC en la Tabla 4.3.

Sobrecargas de tráfico

Se estudia el incremento en las probabilidades de pérdidas que experimentan las peticiones de traspaso P_r^h en presencia de diferentes «grados de sobrecarga» (del 1 % al 500 %). El grado de sobrecarga se define como la razón entre la tasa agregada de llegada de peticiones de sesiones nuevas y la capacidad del sistema:

$$\text{Sobrecarga} = \frac{\sum_{r=1}^R \lambda_r^n}{\lambda_{\text{máx}}} \quad (4.12)$$

Sea una política y unos parámetros del sistema dados, se denomina «servicio limitante» (SL) a aquel cuya probabilidad de bloqueo de sesiones nuevas o de trasposos coincide (o casi) con el objetivo ($P_r^n \simeq B_r^n$ o $P_r^h \simeq B_r^h$) cuando se ofrece al sistema una tasa agregada de peticiones de nuevas sesiones igual a la capacidad de dicha política (es decir, $\sum_{r=1}^R \lambda_r^n = \lambda_{\text{máx}}$). Los resultados del estudio constatan que en sistemas bajo las condiciones mencionadas de intensidad de tráfico y en presencia de políticas con solución en forma de producto es generalmente la P_r^h del SL la que coincide con su objetivo B_r^h . Así, cuando ocurre la sobrecarga, usualmente la P_r^h del SL es la primera en incumplir sus objetivos. En cambio, cuando se emplea cualquiera de las políticas de *trunk reservation* con sistemas en condiciones de carga similares a la capacidad, en general las probabilidades de pérdidas de todas las clases de peticiones están cercanas a sus correspondientes objetivos, especialmente en el caso del SL. Bajo ciertos grados de sobrecarga la P_r^h experimentada por el SL suele ser

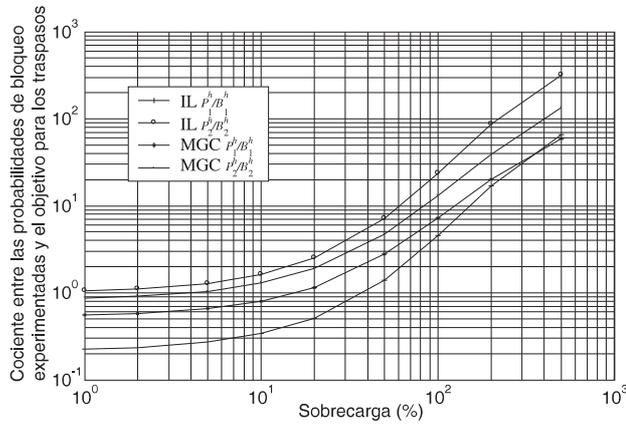


Figura 4.8: Razón entre las probabilidades de bloqueo experimentadas por las peticiones de traspaso y sus respectivos objetivos.

inferior en el caso de políticas de *trunk reservation* en comparación con las políticas con solución en forma de producto. Finalmente, la proporción en que se incrementa P_r^h a medida que aumenta la tasa de llegadas tiende a ser similar para todas las políticas.

La Fig. 4.8 muestra el cociente entre las probabilidades de bloqueo experimentadas por las distintas clases de peticiones de traspaso y sus correspondientes objetivos de probabilidad de bloqueo, respecto a la sobrecarga. Los resultados son referidos al escenario B de la Tabla 4.1 y $C = 20$ unidades de recurso. Por ejemplo, para la política IL un 10 % de sobrecarga, las probabilidades de bloqueo experimentadas por los traspasos del servicio 1, (P_1^h) es de $0.3375B_1^h$ ($B_1^h = 5\%$).

La Fig. 4.9 muestra el incremento relativo de P_r^h para un valor de sobrecarga en relación con el valor experimentado cuando la sobrecarga es la mitad. Los resultados son referidos al escenario B de la Tabla 4.1 y $C = 20$ unidades de recurso. Por ejemplo, para la política IL, la probabilidad de bloqueo experimentada por los traspasos del servicio 1 para una sobrecarga del 10 % es de un 25.23 % superior (1.2523) que el mismo valor para una sobrecarga del 5 %.

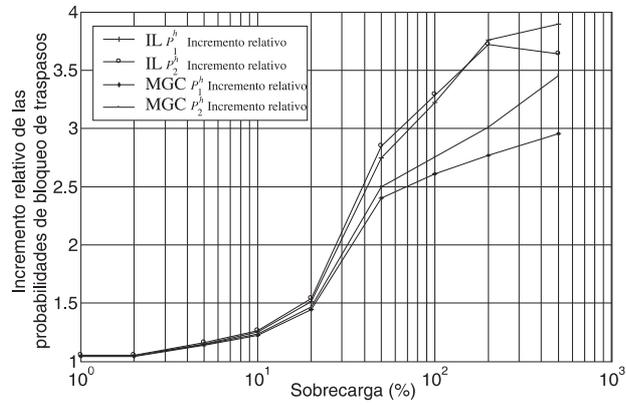
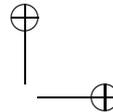
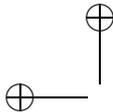


Figura 4.9: Incremento relativo de las probabilidades de bloqueo para traspasos.

Finalmente, se pueden extraer dos conclusiones principales del análisis de sensibilidad a la sobrecarga. Primero, que en general, hacer uso de políticas de control de admisión es conveniente porque introduce un cierto grado de equidad puesto que las clases de peticiones se reparten la penalización procedente del incremento de la probabilidad de bloqueo durante las sobrecargas. En segundo lugar, en general, las políticas de la familia de *trunk reservation* tienden a manejar de manera similar (en la zona de baja a media sobrecarga) o mejor (en la zona de elevada sobrecarga) que las políticas con solución en forma de producto, en el sentido de que el incremento relativo de las probabilidades de bloqueo que experimentan los traspasos es menor que el incremento relativo que se produce con las políticas que producen una solución con forma producto.



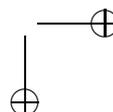
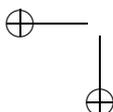
4.6 Algoritmo de escalada para la optimización de políticas

En esta sección se propone un nuevo algoritmo para la optimización del control de admisión. Aunque los procedimientos que se describirán a continuación se pueden trasladar a diversas políticas, es su aplicación a la política MFGC la que se comentará en mayor profundidad. Una de las razones principales es que, como se ha comprobado en la Sección 4.5.1, esta política logra una capacidad significativamente próxima al valor óptimo que consigue la mejor política RS (ver Tabla 4.2). El otro motivo es que la política MFGC requiere definir un número notablemente menor de parámetros para describir completamente la configuración, en comparación con la política RS.

Hasta donde llega el conocimiento del autor, sólo se han propuesto dos algoritmos para el cálculo de la capacidad del sistema en presencia de la política MFGC. Se hará referencia a dichos algoritmos a partir de las iniciales de sus autores:

- Algoritmo HCO, propuesto en Heredia et al. [HUCPOG03c, HUCPOG03b, HUCPOG03a]
- Algoritmo PMC, propuesto en Pla et al. [PMCG05]

Finalmente es necesario subrayar que el algoritmo descrito en esta sección —al igual que los dos anteriores— se limita a hallar los parámetros óptimos de la configuración de la política, por lo que es independiente del método concreto que se emplee para calcular las probabilidades de pérdidas experimentadas por cada clase de peticiones (como el algoritmo de la convolución —si se trata de una de las políticas con solución en forma de producto—, o la resolución de las ecuaciones de balance globales del proceso de Markov mediante Gauss-Seidel, o incluso otras técnicas en las que no se asuma exponencialidad en las distribuciones de los parámetros que definen el modelo del sistema).



4.6.1 Descripción del modelo

El modelo analítico del sistema que se emplea en esta sección es a grandes rasgos el mismo descrito en la Sección 4.2. Sin embargo, para poder comparar las prestaciones del algoritmo en igualdad de condiciones con los algoritmos HCO y PMC, las tasas de llegada de las peticiones de traspaso no se consideran un dato, sino que se calculan a partir de la hipótesis de que existe un equilibrio entre los flujos de entrada y salida de la célula. Aparte se introduce la probabilidad de terminación forzosa para cada servicio (P_r^{ft}) como un parámetro de la QoS. Si se asume una red de células homogéneas, las probabilidades de fallo de handover (P_r^h) y las de terminación forzosa se relacionan a través de la fórmula [HR86]

$$P_r^{ft} = \frac{P_r^h}{\mu_r^s / \mu_r^d + P_r^h} \quad r = 1, \dots, R. \quad (4.13)$$

Se define la tasa agregada de peticiones de nueva sesión como $\lambda^T = \sum_{r=1}^R \lambda_r^n$, relacionada con el vector de actividad mediante la relación $\lambda_r^n = f_r^n \lambda^T$.⁹ Para hacer más compacta la notación se denotará como clase de peticiones i a las llegadas de sesiones nuevas del servicio r (cuando $i = r$), o a las llegadas de peticiones de traspaso del servicio r (cuando $i = r + R$). De este modo el vector $\mathbf{p} = (P_1, \dots, P_{2R})$ representará las probabilidades de bloqueo de las $2R$ clases de peticiones, donde $P_r^n = P_i$, cuando $i = 1, \dots, R$ ó $P_r^h = P_{N+r}$, cuando $i = R + 1, \dots, 2R$.

Se recuerda que el estado del sistema viene descrito por la R -tupla $\mathbf{x} = (x_1, \dots, x_R)$, donde x_r representa el número de sesiones del servicio r en curso dentro del sistema, independientemente de si se iniciaron como peticiones de nueva sesión o de traspaso. Esta distinción es irrelevante cuando se consideran distribuciones exponenciales debido a la propiedad de memoria nula. Nótese que la cantidad de unidades de recurso ocupadas en el estado \mathbf{x} , es $c(\mathbf{x}) = \sum_{r=1}^R x_r c_r$.

⁹En esta ocasión el parámetro f_r^h no está definido puesto que las tasas de llegada de peticiones de traspaso no son un dato.

En presencia de la política MFGC, cuando una petición de la clase i (correspondiente bien al servicio $r = i$ si se trata de una petición de nueva sesión, o bien al servicio $r = i - R$ si se trata de una petición de traspaso) llega al sistema en el estado x , se toman las siguientes decisiones

$$c(x) + c_r \begin{cases} \leq \lfloor t_i \rfloor & \text{aceptar} \\ = \lfloor t_i \rfloor + 1 & \text{aceptar con probabilidad } t_i - \lfloor t_i \rfloor \\ > \lfloor t_i \rfloor + 1 & \text{rechazar.} \end{cases} \quad (4.14)$$

donde t_i son los parámetros de configuración de la política MFGC a optimizar.

4.6.2 Análisis del modelo

El modelo del sistema es un proceso de nacimiento y muerte multidimensional cuyo espacio de estados se denota como S .

Sea r_{xy} la tasa de transición del estado x al estado y y sea e_r un vector cuyas entradas son todas 0 excepto la r -ésima, que es 1.

$$r_{xy} = \begin{cases} a_r^n(x)\lambda_r^n + a_r^h(x)\lambda_r^h & \text{si } y = x + e_r \\ x_r\mu_r & \text{si } y = x - e_r \\ 0 & \text{de lo contrario} \end{cases} \quad (4.15)$$

Los coeficientes $a_r^n(x)$ y $a_r^h(x)$ denotan las probabilidades de aceptar una sesión nueva o traspasada del servicio r respectivamente. Dada una configuración de la política (t_1, \dots, t_{2R}) estos coeficientes se pueden determinar de la siguiente manera

$$a_r^n(x) = \begin{cases} 1 & \text{si } c(x) + c_r \leq \lfloor t_r \rfloor \\ t_r - \lfloor t_r \rfloor & \text{si } c(x) + c_r = \lfloor t_r \rfloor + 1 \\ 0 & \text{si } c(x) + c_r > \lfloor t_r \rfloor + 1 \end{cases} \quad (4.16)$$

y

$$a_r^h(\mathbf{x}) = \begin{cases} 1 & \text{si } c(\mathbf{x}) + c_r < \lfloor t_{R+r} \rfloor + 1 \\ t_{R+r} - \lfloor t_{R+r} \rfloor & \text{si } c(\mathbf{x}) + c_r = \lfloor t_{R+r} \rfloor + 1 \\ 0 & \text{si } c(\mathbf{x}) + c_r > \lfloor t_{R+r} \rfloor + 1 \end{cases} \quad (4.17)$$

De lo anterior, las ecuaciones de balance global se pueden escribir como

$$p(\mathbf{x}) \sum_{\mathbf{y} \in S} r_{\mathbf{x}\mathbf{y}} = \sum_{\mathbf{y} \in S} r_{\mathbf{y}\mathbf{x}} p(\mathbf{y}) \quad \forall \mathbf{x} \in S \quad (4.18)$$

Donde $p(\mathbf{x})$ es la probabilidad en régimen estacionario del estado \mathbf{x} . Los valores de $p(\mathbf{x})$ se obtienen a partir de (4.18) y la ecuación de normalización. De los valores de $p(\mathbf{x})$ las probabilidades de bloqueo se obtiene como

$$P_r = P_r^n = \sum_{\mathbf{x} \in S} (1 - a_r^n(\mathbf{x})) p(\mathbf{x}) \quad P_{R+r} = P_r^h = \sum_{\mathbf{x} \in S} (1 - a_r^h(\mathbf{x})) p(\mathbf{x}) \quad (4.19)$$

Si el sistema se encuentra en equilibrio estadístico las tasas de llegada de peticiones de traspaso se relacionan con las tasas de llegada de peticiones de traspaso y las probabilidades de bloqueo P_i a través de la expresión [LMN94]

$$\lambda_r^h = \lambda_r^n \frac{1 - P_r^n}{\mu_r^s / \mu_r^d + P_r^h} \quad (4.20)$$

Las probabilidades de bloqueo a su vez dependen de las tasas de llegada de traspasos resultando en un sistema de ecuaciones no lineales que se puede resolver mediante un método de iteración de punto fijo como se describe en [HR86, LMN94].

4.6.3 Fundamentos de funcionamiento

El algoritmo descrito trata de calcular la capacidad del sistema (máxima λ_T que es capaz de manejar el sistema sin incumplir los objetivos de QoS impuestos). Al igual que en secciones anteriores, para las clases de peticiones de nueva sesión los requisitos de QoS se siguen definiendo en términos de cotas superiores sus probabilidades de bloqueo (B_r^n), como novedad, las clases

de peticiones de traspaso tiene definidos sus objetivos de QoS en términos de las probabilidades de terminación forzosa (B_r^{ft}). El enfoque usual para sintetizar la política de control de admisión en este sistema multiservicio es mediante la ejecución iterativa de un proceso de análisis (ver Sección 4.4). El proceso de síntesis en esta ocasión tiene como entradas los valores de los siguientes parámetros del sistema: λ_r^n , λ_r^h , μ_r , c_r y C , así como los objetivos de QoS: B_r^n , B_r^{ft} , y como salida devuelve la configuración óptima de los parámetros t_i de la política MFGC. Este proceso de síntesis debe recurrir de manera reiterada a un proceso de análisis que teniendo como entradas el valor de los parámetros del sistema y una configuración (tentativa) de la política produce como salida las probabilidades de bloqueo experimentadas por las distintas clases de peticiones.

Se puede obtener una visión adicional del problema a partir del estudio exhaustivo de la máxima tasa de llegada de peticiones de nuevas sesiones $\lambda_{\max}^T(\mathbf{t})$ que puede cursar cada posible configuración \mathbf{t} de la política sin infringir los requisitos de QoS B_i .

$$\mathbf{t} \rightsquigarrow \lambda_{\max}^T(\mathbf{t}) : \max\{\lambda : P_i(\lambda) \leq B_i\} \quad (4.21)$$

Al resultado de este estudio se denomina el «espacio de soluciones» E de la política, $E = \{(\mathbf{t}; \lambda(\mathbf{t}))\}$. Este espacio es de dimensión

$$\dim(E) = \dim(\mathbf{t}) + 1 = 2R + 1 \quad (4.22)$$

donde su valor máximo

$$\lambda_{\text{CAP}} \rightsquigarrow \exists \mathbf{t}^* : \lambda_{\max}^T(\mathbf{t}^*) = \lambda_{\text{CAP}}, \quad \lambda_{\max}^T(\mathbf{t}) \leq \lambda_{\max}^T(\mathbf{t}^*); \forall \mathbf{t} \quad (4.23)$$

se corresponde con la capacidad de la política λ_{CAP} .

A continuación se explica la obtención de la superficie definida por el espacio de soluciones. Para cada configuración de la política t_i , se computa mediante un proceso de búsqueda binaria similar al descrito en la Sección 3.3.4 que tiene como entrada μ_r , c_r , C y t_i y devuelve las probabilidades de bloqueo P_r^n y P_r^h . El proceso de búsqueda binaria se detiene cuando encuentra la λ_{\max}^T que cumple los objetivos de QoS (B_r^n and B_r^{ft}), $r = 1, \dots, R$.

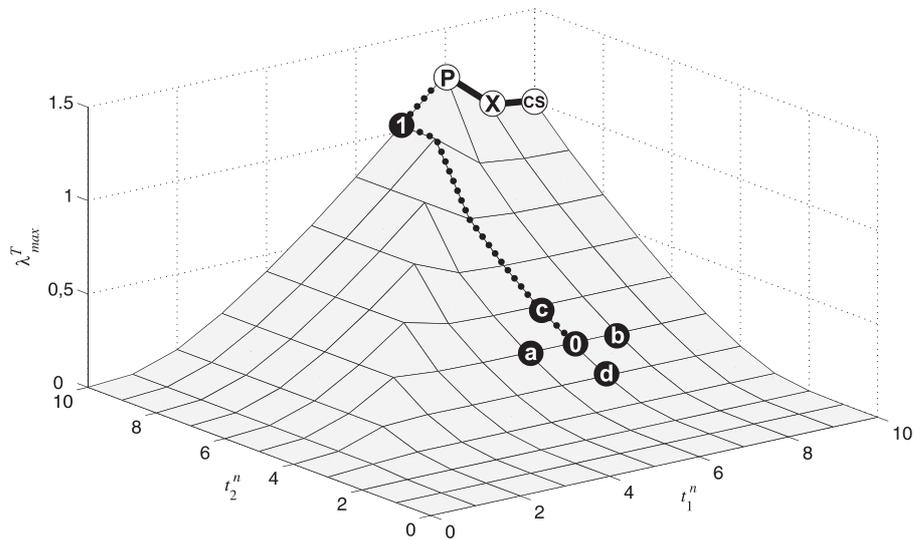


Figura 4.10: Algoritmo de escalada para determinar la configuración óptima de la política MFGC.

Con la intención de ilustrar el algoritmo propuesto se ha escogido un ejemplo simple que permita representar el espacio de soluciones en tres dimensiones. La Fig. 4.10 muestra el espacio de soluciones cuando se emplea la política MFGC en la versión fija del escenario A (luego no existen las clases de peticiones de traspaso respectivas) de la Tabla 4.1 con $C = 10$ unidades de recurso. Se recuerda que los parámetros del escenario A eran $c = (1, 2)$, $f = (0.8, 0.2)$, $\mu = (1, 3)$, $B^n = (0.05, 0.01)$. La configuración de la política se define por dos parámetros t_1 y t_2 . Nótese que la tasa máxima se expresa como valor relativo respecto a la capacidad que se obtiene con la política CS.

La forma del espacio de soluciones que se muestra en la Fig. 4.10 —que exhibe un único máximo— sugiere que un «algoritmo de escalada» (*hill climbing*) podría ser un enfoque eficiente para obtener la configuración óptima de la política MFGC en este escenario. A esto hay que añadir que se ha obtenido el espacio de soluciones para otras diferentes políticas (IL, GM y MGC) en la

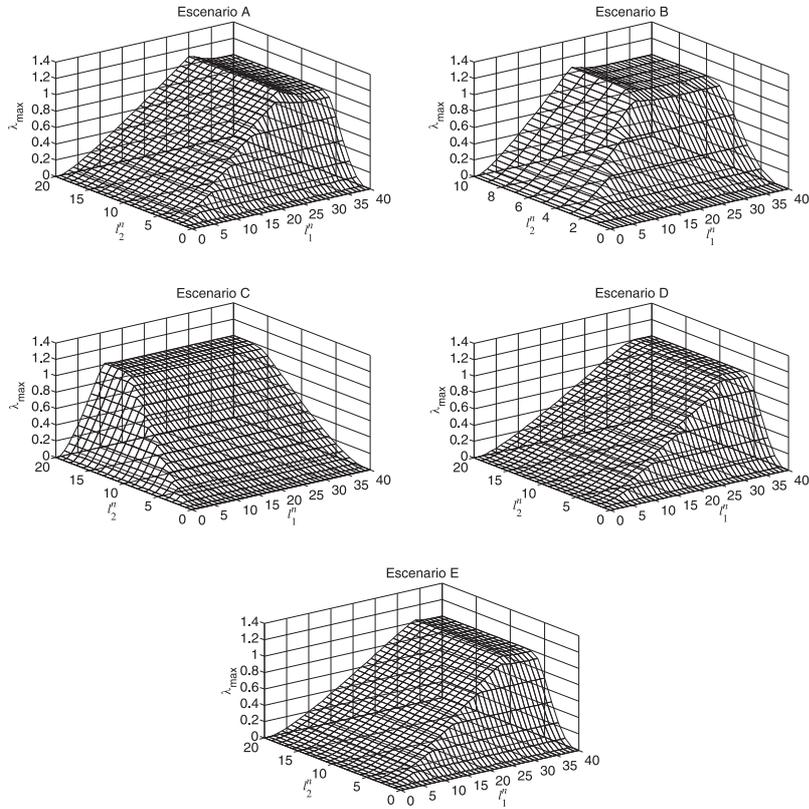


Figura 4.11: Espacios de soluciones, política IL.

versión fija de los escenarios de la Tabla 4.1 y se verifica que siempre existe un valor máximo único para la tasa máxima de llegada y que este valor coincide con la capacidad del sistema (la forma de los diferentes espacios de soluciones se puede apreciar en las Figs. 4.11, 4.12 y 4.13 respectivamente, con $C = 40$ unidades de recurso).

La Fig. 4.10 muestra el funcionamiento del algoritmo de escalada:

1. Dado un punto de partida en un espacio de búsqueda 2R-dimensional

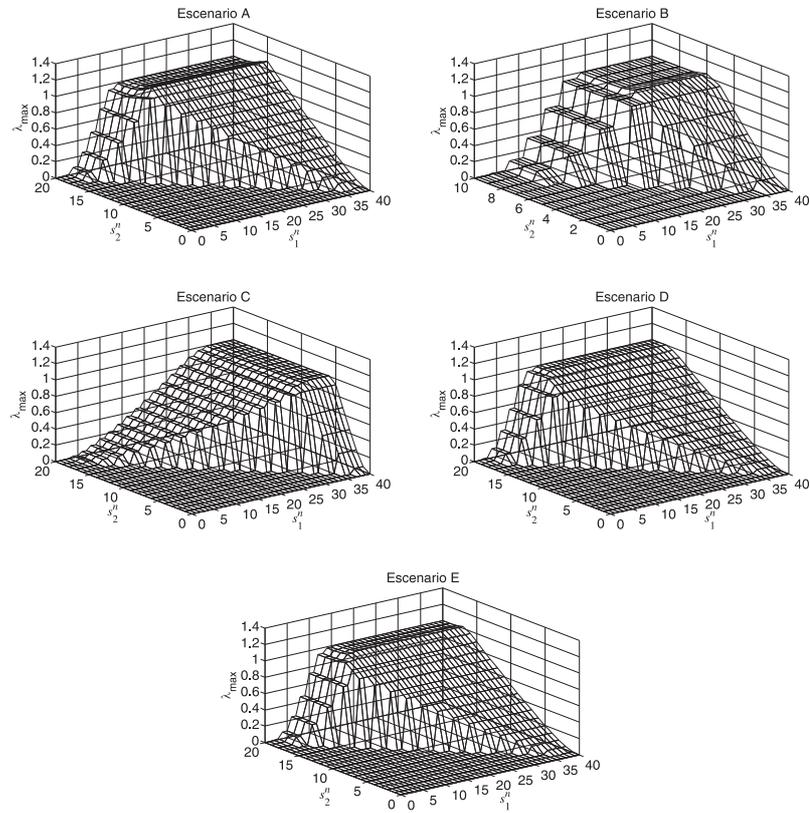


Figura 4.12: Espacios de soluciones, política GM.

(por ejemplo, el punto **0**), el algoritmo de escalada comienza calculando el valor de la función (λ_{max}^T), y de las probabilidades de bloqueo para las diferentes clases de peticiones (P_r^n y P_r^h).

2. Se selecciona la dimensión con la mayor pendiente tal y como se describe más abajo (en este caso t_2^n).
3. El algoritmo busca el máximo a lo largo de dicha dimensión (en este caso **1**). Nótese que se trata de un problema de maximización a lo largo

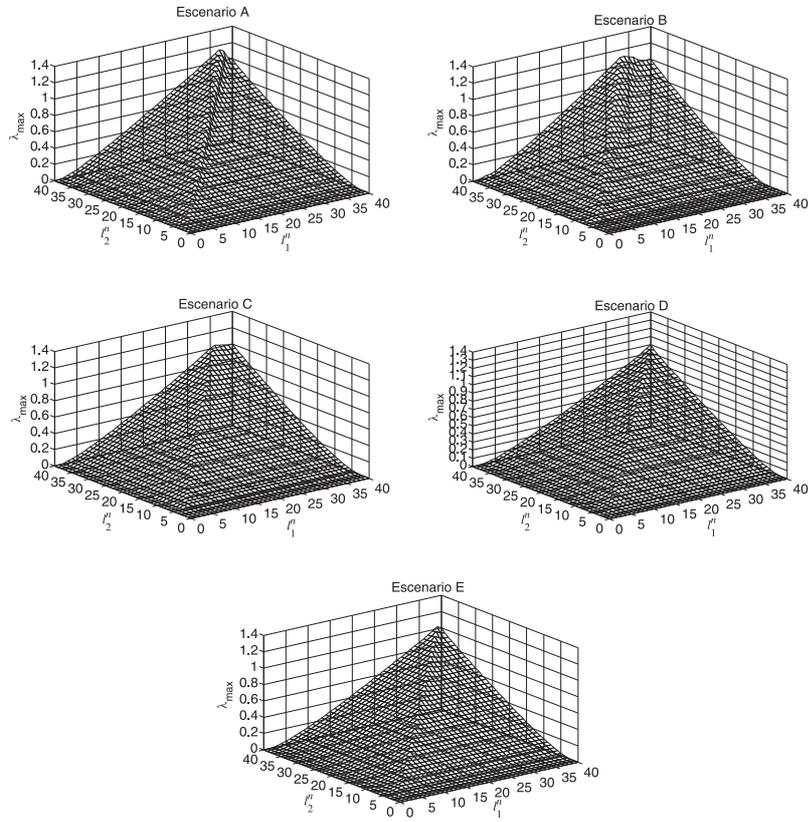


Figura 4.13: Espacios de soluciones, política MGC.

de una línea.

4. Retornar al paso 1 hasta que se halla el máximo local \mathbf{P} con la precisión deseada (la progresión del algoritmo se muestra con una línea punteada).

Para el algoritmo de escalada explicado surgen dos cuestiones importantes:

- Cómo se realiza la selección de la dimensión con mayor pendiente
- Cómo se lleva a cabo la búsqueda del máximo

Cuando se aplican al problema objeto de estudio métodos basados en el gradiente, la función debe evaluarse para al menos uno de los dos vecinos adyacentes en cada una de las $2R$ dimensiones (puntos **a**, **b**, **c** y **d**), seleccionando a continuación la dimensión más escarpada como aquella para la que el valor de la función es el superior (**c**). Sin embargo, si el proceso de búsqueda binaria que calcula la tasa máxima se lleva a cabo con una precisión reducida, o los vecinos están lo suficientemente cercanos al punto considerado, entonces este método es impracticable, principalmente porque los valores de la función para los vecinos son idénticos al punto considerado, dando ninguna información en absoluto (por ejemplo, los puntos **a** y **b**). Como este puede ser el caso durante el curso de la ejecución del algoritmo de escalada, se requiere de otro enfoque para determinar la dimensión de mayor pendiente. Se define la «distancia relativa» al objetivo de QoS B_i de una clase de peticiones i que está experimentando una probabilidad de bloqueo P_i (suponiendo que cumple su objetivo, es decir, $P_i < B_i$), al cociente $(B_i - P_i)/B_i$. Como, generalmente, la distancia relativa al objetivo de QoS de una clase de peticiones es un indicio de que las peticiones de dicha clase toleran un aumento en su probabilidad de bloqueo sin incumplir los objetivos de QoS (proporcionando así una capacidad adicional para el resto de clases), se elige como dimensión con mayor pendiente el parámetro de configuración t_i asociado con la clase i que maximiza la distancia relativa al objetivo de QoS. Además, este enfoque tiene un beneficio adicional en términos de complejidad computacional: ya no es preciso calcular la tasa máxima en los puntos del vecindario para encontrar la dimensión con mayor pendiente.

En su búsqueda del máximo a lo largo de la dimensión con mayor pendiente, el algoritmo propuesto realiza una serie de pasos unitarios y se detiene cuando alcanza el máximo de la función. Cuando el espacio de soluciones es continuo, como sucede con la política MFGC, se requiere de un proceso de refinamiento gradual que reduzca el tamaño de los pasos una vez que

se ha encontrado una región prometedora que quizás se encuentre en las inmediaciones del óptimo. Se puede obtener una reducción adicional de la complejidad computacional si se observa que la configuración óptima (punto **P**) para cualquier política suele encontrarse cerca de la configuración de la política CS (punto **CS**), y por lo tanto es una buena idea seleccionar esta configuración como el punto de partida para el algoritmo. La Fig. 4.10 ilustra la progresión típica del algoritmo propuesto (línea sólida) comenzando por la configuración de CS (punto **CS**) hacia el punto **X** y terminando en el máximo (punto **P**).

4.6.4 Descripción del algoritmo

El problema de optimización de la capacidad se puede precisar formalmente de la siguiente manera

Dado: $C, c_r, f_r, \mu_r^c, \mu_r^r, B_r^n, B_r^{ft}; r = 1, \dots, R$

Maximizar: $\lambda^T = \sum_{1 \leq r \leq R} \lambda_r^n, \lambda_r^n = f_r \lambda^T$
 encontrando configuración apropiada de la política MFGC $t_i; i = 1, \dots, 2R$

Sujeto a: $P_r^n \leq B_r^n, P_r^{ft} \leq B_r^{ft}; r = 1, \dots, R$

Se propone un algoritmo para resolver el problema de optimización de la capacidad. El algoritmo tiene una parte principal (Algoritmo 1 `solveMFGC`) desde la que se llama al procedimiento `capacity` (ver Algoritmo 2). El procedimiento `capacity` a su vez, llama a otro procedimiento (MFGC) que calcula las probabilidades de bloqueo.

Para simplificar la notación se introduce la $2R$ -tupla $p_{max} = (B_1^n, \dots, B_R^n, B_1^h, \dots, B_R^h)$ como las cotas superiores para las probabilidades de bloqueo, donde el valor de B_r^h proviene de

$$B_r^h = \frac{\mu_r^s}{\mu_r^d} \frac{B_r^{ft}}{1 - B_r^{ft}} \quad (4.24)$$

Siguiendo la convención común se usan fuentes en **negrita** para representar vectores en el pseudocódigo del algoritmo.

Algoritmo 1 $(\lambda_{max}^T, t_{opt}) = \text{solveMFGC}(C, p_{max}, c, \mu_s, \mu_d)$ (calcula los parámetros de configuración de la política MFGC)

```

1:  $\epsilon_2 :=$  < precisión deseada >
2:  $current\epsilon_2 := 1$ 
3:  $point := C$ 
4:  $direction := -1$ 
5:  $step := (1, 1, \dots, 1)$  <size  $2R$ >
6:  $steepest := 0$ 
7:  $changeOfDirection := FALSE$ 
8:  $t_{opt} := (C, C, \dots, C)$ ;  $t := t_{opt}$ 
9:  $\lambda_{max}^T := \mathbf{0}$ ;  $\lambda^T := \mathbf{0}$ ;
10:  $p_{opt} := \mathbf{0}$ ;  $p := \mathbf{0}$ ;
11:  $dp_{opt} := \mathbf{0}$ ;  $dp := \mathbf{0}$ ;
12:
13:  $(\lambda_{max}^T, p_{opt}) := \text{capacity}(p_{max}, t_{opt}, \mu_s, \mu_d, c, C)$ 
14:  $dp_{opt} := (p_{max} - p_{opt}) / p_{max}$ ;  $dp := dp_{opt}$ 
15:  $current\epsilon_2 := \max(dp_{opt})$ 
16:  $steepest :=$  < la clase  $i$  que maximiza  $dp_{opt}(i)$  >
17:
18: while  $current\epsilon_2 > \epsilon_2$  do
19:    $point := t_{opt}(steepest)$ 
20:    $direction := -1$ 
21:   if  $step(steepest) < > 1$  then
22:      $step(steepest) = 0.5$ 
23:   end if
24:    $changeOfDirection := FALSE$ 
25:
26:   repeat
27:     if  $direction = -1$  then
28:        $point = point - step(steepest)$ 
29:     else
30:        $point = point + step(steepest)$ 
31:     end if
32:
33:      $t := t_{opt}$ ;  $t(steepest) := point$ ;
34:      $(\lambda^T, p) := \text{capacity}(p_{max}, t, \mu_s, \mu_d, c, C)$ 
35:      $dp := (p_{max} - p) / p_{max}$ ;
36:

```

```

37:   if  $\lambda^T \geq \lambda_{max}^T$  then
38:      $t_{opt}(\text{steepest}) := \text{point}; \lambda_{max}^T := \lambda^T;$ 
39:      $p_{opt} := p; dp_{opt} := dp;$ 
40:   end if
41:
42:   if  $dp(\text{steepest}) > \varepsilon_2$  then
43:     if direction = -1 then
44:       if changeOfDirection then
45:          $step(\text{steepest}) := step(\text{steepest})/2$ 
46:       end if
47:     else
48:        $step(\text{steepest}) := step(\text{steepest})/2$ 
49:       direction := -1
50:       changeOfDirection := TRUE
51:     end if
52:   else
53:     if  $\lambda^T < \lambda_{max}^T$  then
54:       if direction = +1 then
55:         if changeOfDirection then
56:            $step(\text{steepest}) := step(\text{steepest})/2$ 
57:         end if
58:       else
59:          $step(\text{steepest}) := step(\text{steepest})/2$ 
60:         direction := +1
61:         changeOfDirection := TRUE
62:       end if
63:     end if
64:   end if
65: until  $(dp(\text{steepest}) < \varepsilon_2)$  AND  $(\lambda^T \geq \lambda_{max}^T)$ 
66:
67:  $steepest := \langle \text{la clase } i \text{ que maximiza } dp_{opt}(i) \rangle$ 
68:  $current\varepsilon_2 := \max(dp_{opt})$ 
69: end while

```

Algoritmo 2 $(\lambda_{max}^T, p) = \text{capacity}(p_{max}, t, \mu_s, \mu_d, c, C)$

INPUTS: $p_{max}, t, \mu_s, \mu_d, c, C$

OUTPUTS: λ_{max}^T, p

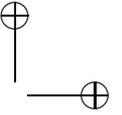
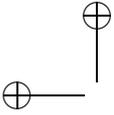
- 1: $\varepsilon_1 := \langle \text{precisión deseada} \rangle$
- 2: $current\varepsilon_1 := 1$

```

3:  $L := 0$ 
4:  $U := \langle \text{high value} \rangle$ 
5:  $\text{meetQoSrequirements} := \text{FALSE}$ 
6:
7: while ( $\text{current}\varepsilon_1 > \varepsilon_1$ ) OR NOT( $\text{meetQoSrequirements}$ ) do
8:    $\lambda_{max}^T := (U + L)/2$ 
9:    $p := \text{MFGC}(t, \lambda_{max}^T, \mu_s, \mu_d, c, C)$ 
10:   $\text{current}\varepsilon_1 := \min((p_{max} - p)/p_{max})$ 
11:  if  $\text{current}\varepsilon_1 < 0$  then
12:     $U := \lambda_{max}^T$ 
13:     $\text{meetQoSrequirements} := \text{FALSE}$ 
14:  else
15:     $L := \lambda_{max}^T$ 
16:     $\text{meetQoSrequirements} := \text{TRUE}$ 
17:  end if
18: end while

```

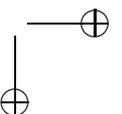
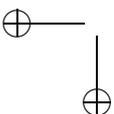
El algoritmo `solveMFGC`, comienza calculando la configuración de la política CS como el punto de partida (líneas 13–14), y selecciona la clase de peticiones i que maximiza $(B_i - P_i)/B_i$ como la dimensión con mayor pendiente, (línea 16). El bucle completo de escalada comienza en la línea 18, y el bucle de maximización a lo largo de una dimensión comienza en la línea 26. Nótese (línea 21) que en la primera ocasión en que se elige una dimensión como la de mayor pendiente, el paso de escalada es igual a 1, sin embargo, si una dimensión ya ha sido elegida previamente, el paso inicial de escalada se reduce a 0.5 porque se asume cierta localidad a la configuración óptima. Las líneas (37–64) llevan a cabo el algoritmo de escalada a lo largo de la dimensión con mayor pendiente. Esta subrutina lleva a cabo tareas como cambiar la dirección de los sucesivos pasos y el refinamiento de los mismos una vez que se llega a una configuración prometedora. El algoritmo `capacity` es básicamente, una búsqueda binaria de λ_{max}^T que llama al procedimiento MFGC en cada iteración con el propósito de calcular las probabilidades de bloqueo.



Acerca del procedimiento MFGC

El procedimiento MFGC se invoca desde el bucle más interno del algoritmo y se emplea para calcular las probabilidades de bloqueo, $p := \text{MFGC}(t, \lambda_n, \mu_s, \mu_d, c, C)$, resolviendo la cadena de Markov de tiempo continuo (CTMC) que modela el sistema mediante el método de Gauss-Seidel. La mayor parte de la complejidad computacional del algoritmo descrito en esta sección viene de resolver en múltiples ocasiones la CTMC, por lo tanto la diferencia entre los diferentes algoritmos radica básicamente en cuantas veces debe resolverse la CTMC. Nótese que en la Sección 4.6.5 de evaluación de la complejidad computacional sólo se consideran dos servicios. Para escenarios con un número superior de servicios, la cadena de Markov tendría $2R$ dimensiones (peticiones nuevas y de traspaso). Para obtener resultados provechosos el número de unidades de recurso del sistema tendría que estar dimensionado apropiadamente lo que causaría una explosión en el espacio de estados convirtiendo así en impracticable la evaluación numérica de cualquiera de los algoritmos que aquí se comparan. Para comparar los algoritmos en escenarios con un número mayor de dimensiones se requeriría abordar de manera más eficiente el problema de dimensionalidad inherente a estos escenarios mediante el empleo de algún método aproximado para resolver la CTMC asociada. No obstante, resolver la cadena de Markov con una precisión inferior tendría un impacto en el comportamiento de los algoritmos, que puede variar de un algoritmo a otro. Un estudio detallado del comportamiento desconocido aún de estos algoritmos en presencia de soluciones aproximadas de la CTMC está fuera del ámbito de este capítulo.

Al mismo tiempo, para el cálculo de las probabilidades de bloqueo, se requiere de un procedimiento adicional de iteración de punto fijo para obtener el valor de las tasas de llegada de las peticiones de traspaso (ver el final de la Sección 4.6.2). En cada iteración se deber resolver un proceso multidimensional de nacimiento y muerte. Resolver este proceso, que en general tendrá un número elevado de estados, constituye la parte que genera más costes de computación del algoritmo.



Se hace uso de la misma mejora explicada en [PMCG05] para eliminar la iteración de punto fijo necesaria para calcular las tasas de llegada de los traspasos. Cada ejecución de `capacity` encuentra una λ_{max}^T (dentro de una tolerancia) tal que $p \leq p_{max}$. Así, en lugar de emplear (4.20) para calcular λ_r^h se emplea la expresión

$$\lambda_r^h = \lambda_r^n \frac{1 - B_r^n}{\mu_r^s / \mu_r^d + B_r^h} \quad (4.25)$$

Aunque (4.20) y (4.25) parecen muy similares existe una diferencia substancial entre las dos. En la Ecuación (4.25), λ_r^h está definida explícitamente mientras que en (4.20) no lo está, debido a que tanto P_r^n como P_r^h dependen de λ_r^h . Nótese que $p = p_{max}$ sólo cuando λ_{max}^T es igual a la capacidad del sistema (dentro de una tolerancia), pero usar 4.25 reduce considerablemente los costes computacionales y por lo tanto acelera la velocidad de convergencia del algoritmo.

Se usa la expresión (4.25) debido a que mediante una adecuada configuración de los parámetros de la política MFGC es posible cumplir los objetivos de QoS con elevada precisión (siempre y cuando exista una solución factible). Queda claro que cuando la tasa agregada de llegadas iguala a la capacidad del sistema el valor de los parámetros de configuración son tales que las probabilidades de bloqueo percibidas por las diferentes clases de peticiones están muy cerca de sus objetivos. Por lo tanto, incluso si las tasas de peticiones de traspasos calculadas al principio del algoritmo puede ser aproximadas, su precisión mejora a medida que el algoritmo progresa hacia el máximo.

4.6.5 Evaluación numérica de la complejidad computacional

Se compara la complejidad computacional del algoritmo propuesto, con la de los algoritmos previos de Heredia et al. [HUCPOG03a] (algoritmo HCO) y Pla et al. [PMCG05] (algoritmo PMC).

El algoritmo HCO requiere el «orden de prioridad» óptimo (*optimal prioritization order*) como entrada, es decir: la lista de las clases de peticiones

ordenadas por sus prioridades relativas. Recuérdese que para un sistema con R servicios existen $2R$ clases de peticiones distintas. La política MFGC se define (ver Sección 4.3.2) por la $2R$ -tupla $\mathbf{t} = (t_1, \dots, t_{2R})$, donde el parámetro de la configuración $t_i \in \mathbb{R}$ representa la cantidad media de recursos que tiene a su disposición la clase de peticiones i . Si \mathbf{t}_{opt} es la configuración de la política para la que se logra la capacidad del sistema, el orden de prioridad óptimo es la permutación $\sigma^* \in \Sigma$, $\Sigma := \{(\sigma_1, \dots, \sigma_{2R}) : \sigma_i \in \mathbb{N}, 1 \leq \sigma_i \leq 2R\}$, tal que $t(\sigma_1^*) \leq t(\sigma_2^*) \leq \dots \leq t(\sigma_{2R}^*) = C$, donde $t(\sigma_i^*)$ es el elemento σ_i^* de \mathbf{t}_{opt} y C es la cantidad total de unidades de recurso del sistema. Seleccionar el orden de prioridad óptima es una tarea complicada dado que depende tanto de los requisitos de QoS como de las características del sistema como se apunta en [HUCPOG03a]. En general existen $(2R)!$ órdenes de prioridad diferentes. En [HUCPOG03a] los autores proporcionan ciertas directrices para contruir una lista parcialmente ordenada de órdenes de prioridad, de acuerdo con la verosimilitud de ser los óptimos. A continuación se sigue un procedimiento de prueba y error empleando elementos sucesivos de la lista hasta que se encuentra el orden de prioridad óptimo. Para cada elemento se ejecuta el algoritmo HCO y si después de un número significativo de iteraciones no ha convergido, se prueba otro orden de prioridad.

El algoritmo PMC no requiere de ningún conocimiento a priori. De hecho, tras obtener la configuración óptima de la política MFGC \mathbf{t}_{opt} para la que se obtiene la capacidad del sistema, el orden de prioridad se puede recuperar mediante una mira inspección de la solución obtenida. Además, a partir de ejemplos numéricos se demuestra en [PMCG05] que el algoritmo PMC es más eficiente que el algoritmo HCO incluso cuando a este último se le proporciona inicialmente el orden de prioridad óptimo. En [PMCG05] el problema de optimización se formula como un problema de programación no lineal que trata de determinar los parámetros de configuración de la política MFGC que maximizan las tasas de llegada de peticiones a la vez que mantienen las probabilidades de bloqueo por debajo unas cotas especificadas, adicionalmente se proporciona un algoritmo para resolver el problema de programación no lineal. Debido a que, en general, las probabilidades de bloqueo son funciones

no monótonas tanto respecto a la carga ofrecida como a los parámetros de la configuración de la política, encontrar la solución óptima no es una tarea sencilla y no se proporciona evidencia que apoye que el algoritmo converge en todas las posibles configuraciones del escenario.

Para los ejemplos numéricos se considera un sistema con dos servicios ($C = 2$), y para valorar el impacto de la movilidad sobre la complejidad computacional, se consideran cinco escenarios diferentes (A–E) con factores de movilidad (μ_r^d / μ_r^s) variables. El conjunto de parámetros que definen al escenario A son: $c = (1, 2)$, $f = (0.8, 0.2)$, $\mu_s = (1/180, 1/300)$, $\mu_d = (1/900, 1/1000)$, $B^n = (0.02, 0.02)$, $B^{ft} = (0.002, 0.002)$; todas las tolerancias se fijan a $\epsilon = 10^{-2}$. De (4.24), $B^h \approx (0.01002, 0.00668)$ por lo que $p_{max} \approx (0.02, 0.02, 0.01002, 0.00668)$.

Para el resto de escenarios los parámetros tienen los mismos valores que los utilizados en el escenario A excepto μ_r^d , que se varía para obtener cuatro combinaciones diferentes de factores de movilidad: B) $\mu_1^d = 0.2\mu_1^s$, $\mu_2^d = 0.2\mu_2^s$; C) $\mu_1^d = 0.2\mu_1^s$, $\mu_2^d = 1\mu_2^s$; D) $\mu_1^d = 1\mu_1^s$, $\mu_2^d = 0.2\mu_2^s$; E) $\mu_1^d = 1\mu_1^s$, $\mu_2^d = 1\mu_2^s$.

En la Tabla 4.4 figura una comparación entre el número de operaciones de coma flotante (*flop*) que requiere la ejecución de los algoritmos HCO, PMC y el algoritmo propuesto, bajo el entorno de cálculo numérico/lenguaje de programación Matlab. A los tres algoritmos se les implementó la técnica de aceleración (ver la Sección 4.6.4). Vale la pena reseñar que, como era de esperar, en todos los casos, los valores obtenidos para la capacidad óptima por medio de los diferentes algoritmos están dentro de la tolerancia impuesta.

La cuenta de *flops* de la Tabla 4.4 para los algoritmos HCO y PMC está extraída directamente de Pla et al. [PMCG05]. En [PMCG05] se comenta que al algoritmo HCO se le proporciona el orden de prioridad como dato de entrada por lo que no se necesita realizar la búsqueda previa que los autores de [HUCPOG03a] proponen, lo cual es una ventaja substancial en términos de coste computacional. Además, como en su versión original el algoritmo HCO no implementa la técnica de aceleración presentada en [PMCG05], la cuenta de *flops* sería bastante superior a la reflejada en la

Cuadro 4.4: Comparación del algoritmo HCO y PMC con el algoritmo propuesto (*our algorithm*), en Mflops.

C	HCO				PMC				Our algorithm			
	5	10	20	TOT	5	10	20	TOT	5	10	20	TOT
A	2.00	20.00	156.00	178.00	0.39	4.53	46.60	51.52	0.21	1.17	9.28	10.66
B	2.08	17.54	74.33	93.95	0.35	4.42	53.64	58.41	0.27	1.92	8.70	10.89
C	2.67	14.06	147.13	163.86	0.34	3.87	43.01	47.22	0.26	1.28	12.90	14.44
D	1.12	24.54	110.41	136.07	0.38	3.93	47.95	52.26	0.23	1.74	12.36	14.33
E	2.24	16.86	121.39	140.49	0.31	3.93	45.92	50.16	0.26	2.39	13.56	16.21
TOT	28.11	93.00	609.26	730.37	1.77	20.68	237.12	259.57	1.23	8.50	56.80	66.53

Table 4.4. En [PMCG05] se muestra que la técnica de aceleración divide la cuenta de *flops* en los cinco escenarios estudiados por un factor medio de aproximadamente tres.

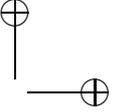
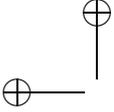
El algoritmo propuesto en esta sección se comporta mejor que el resto. Los factores de ganancia varían entre 4.9–17 con respecto al algoritmo HCO y entre 1.2–6.3 con respecto al algoritmo PMC, con una ganancia media de 10.3 y 2.81 sobre los algoritmos HCO y PMC, respectivamente.

4.7 Conclusiones

En este capítulo se ha determinado la capacidad que el operador puede esperar de un sistema móvil celular multiservicio en presencia de diferentes políticas de control de admisión. Debido a la multidimensionalidad y el comportamiento no monótono del sistema estudiado, la determinación de la configuración óptima de la política (aquella que maximiza el tráfico ofrecido que puede cursar el sistema sin incumplir los objetivos de QoS impuestos) se convierte en un problema complejo y costoso en términos computacionales. La forma del espacio de soluciones de las políticas más comunes sugiere que un algoritmo de escalada puede ser una alternativa prometedora para optimización de dichas políticas, en términos de la reducción del coste de cálculo que

conllevaría el cálculo de la configuración óptima de cada política. Para la capacidad del sistema, los resultados muestran que las políticas de la familia de *trunk reservation* superan en prestaciones a la familia de políticas con solución en forma de producto, con unos rangos de mejora entre el 5–10%. Sin embargo, la ganancia relativa entre las políticas disminuye cuando el número de unidades de recurso C se incrementa. Además, la forma del espacio de soluciones de las políticas de *trunk reservation* muestra que se requiere una mayor precisión a la hora de determinar la configuración óptima, en comparación con las políticas en forma de producto. Asimismo, las políticas con solución en forma de producto presentan la ventaja de que los modelos obtenidos son insensibles al tipo de distribución considerada para la variable aleatoria que describe el tiempo de utilización de los recursos, de la que solamente interesa el valor medio [Ive87]. Dado que en la práctica los parámetros del sistema se deben estimar y son no estacionarios, las políticas de *trunk reservation* podrían llegar a ser menos atractivas según la situación. Del mismo modo se ha estudiado también la sensibilidad de las políticas a las sobrecargas y se ha encontrado que, en general, las políticas de *trunk reservation* manejan mejor la sobrecarga.

Adicionalmente se ha propuesto un nuevo algoritmo para el cálculo de la configuración óptima de la política *Multiple Fractional Guard Channel* (MFGC) en redes móviles celulares multiservicio —aunque podría emplearse para otras políticas de control de admisión con espacio de soluciones monótono—. Comparado con dos algoritmos publicados recientemente (HCO y PMC), el algoritmo de escalada propuesto, que está basado en un enfoque sencillo e intuitivo, es menos costoso computacionalmente en todos los escenarios estudiados. Además, el concepto del espacio de soluciones desvela una perspectiva novedosa en el problema de determinar los valores óptimos de los parámetros de configuración de la política MFGC —de por sí una contribución significativa— proporcionando además, una evidencia heurística de que el algoritmo encuentra la solución óptima y converge en todos los escenarios.

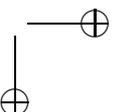
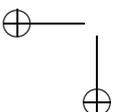


Capítulo 5

Esquema adaptativo de control de admisión

El Capítulo 4 pone de manifiesto que la optimización del Control de Admisión de Sesiones (CAS) es un mecanismo eficaz para maximizar el tráfico que la red es capaz de soportar (capacidad del sistema), a la vez que se satisfacen los objetivos de Calidad de Servicio (QoS) fijados y las demandas de recursos exigidas por los múltiples servicios que compondrán las futuras redes móviles celulares. El CAS especifica la acción (aceptación/rechazo) que debe tomarse en cada estado del sistema cuando llega una petición de sesión nueva o de traspaso. En el Capítulo 3 se vio que incorporar el impacto de la movilidad del terminal en el diseño del CAS complica todavía más el dimensionado de la red, debido a que se debe asegurar que existirán suficientes recursos libres a disposición de la sesión en todas aquellas células que visite el terminal con sesión en curso, mientras dure la misma.

En el Capítulo 4 también se muestra que las capacidades que se pueden obtener con políticas de la familia *trunk reservation* son superiores respecto a la familia de políticas con solución en forma de producto; en particular, las prestaciones de la política *Multiple Fractional Guard Channel* resultan muy

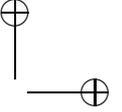
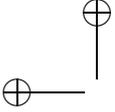


parecidas a las que se obtendrían en presencia de una política óptima. En el caso de todas las políticas comentadas en el capítulo anterior, se requieren dos conjuntos de parámetros para determinar la capacidad del sistema: aquellos que describen los servicios (en este caso como procesos de Markov) y aquellos que definen los objetivos de QoS (por ejemplo en términos de la probabilidad de bloqueo máxima tolerable por cada clase de peticiones). Hasta el momento, el enfoque aplicado para el diseño de políticas es el que se emplea tradicionalmente durante la fase de planificación de la red: considerar que los parámetros que describen a los servicios son estacionarios.¹ Bajo esta hipótesis diseñar una política de CAS estática para el peor caso resulta una opción adecuada.

Sin embargo, durante la fase de operación real de la red los parámetros de los servicios siguen un comportamiento marcadamente «no estacionario» que se experimenta en el sistema como periodos de carga superior a la capacidad que se había previsto durante la fase de planificación («sobrecarga»), o inferior a la mismas («subcarga»). En estos escenarios, las políticas de CAS estáticas no son la solución más adecuada —puesto que la QoS obtenida puede diferir considerablemente de los objetivos establecidos— y se debe recurrir a un «esquema adaptativo» que se encargue de ajustar dinámicamente la política de CAS, bien en función de estimaciones periódicas de los parámetros de los servicios o de la QoS experimentada, o bien recurriendo a informes históricos actualizados acerca de los patrones de tráfico.

En el presente capítulo se propone un esquema adaptativo de control de admisión que ajusta de forma apropiada los parámetros de configuración de dos políticas de la familia *trunk reservation* —*Multiple Guard Channel* (MGC) y *Multiple Fractional Guard Channel* (MFGC)—, en función de la QoS percibida por los usuarios. El esquema propuesto es adaptativo en el sentido de que si la carga ofrecida está por encima de la capacidad del sistema, o el número

¹En un proceso estrictamente estacionario, la distribución de probabilidades en un instante determinado del tiempo es la misma que para cualquier otro instante arbitrario. En consecuencia, momentos de la variable aleatoria tales como la media y la varianza, si están definidos, tampoco varían con el instante de tiempo.

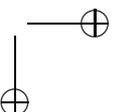
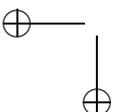


de unidades de recurso disminuye (o se verifican ambas a la vez), la política de CAS reaccionará tratando de hacer cumplir el objetivo de QoS para tantas clases como sea posible. En consecuencia el esquema propuesto se puede emplear tanto para sistemas con una cantidad de recursos fija (basados en FDMA/TDMA) como para sistemas limitados por interferencia (basados en CDMA) donde los recursos percibidos son variables. Aspectos en los que se ha puesto especial énfasis a la hora del diseño y evaluación del esquema son la velocidad de adaptación a las nuevas condiciones de tráfico, el soporte adecuado del multiservicio, y la rigurosidad con la que se respetan los objetivos de QoS exigidos por el operador.

El resto del capítulo está estructurado de la siguiente manera: en la Sección 5.1 se pone la materia en antecedentes, en la Sección 5.2 se describe el modelo del sistema —definiendo asimismo las políticas de CAS empleadas—, en la Sección 5.3 se presentan los fundamentos detrás del esquema adaptativo y se explica su funcionamiento, en la Sección 5.4 se procede a caracterizar analíticamente el modelo del sistema en presencia del esquema adaptativo, en la Sección 5.5 se presentan los detalles de dos esquemas adaptativos previamente propuestos por Zhang et al. [ZL01] y Wang et al. [WZZZ03] para escenarios monoservicio, en la Sección 5.6 se evalúan comparativamente las prestaciones del esquema diseñado frente a estos dos esquemas previos en entornos monoservicio, asimismo se muestran las prestaciones obtenidas por el esquema propuesto en escenarios multiservicio (tanto estacionarios como no estacionarios) y finalmente en la Sección 5.7 se hace un resumen de este capítulo y se presentan las conclusiones.

5.1 Antecedentes

Recientemente, se han propuesto numerosos esquemas adaptativos de CAS para redes móviles celulares. En estos esquemas la configuración de la política de CAS se adapta periódicamente de acuerdo a estimaciones del tráfico o de los parámetros de QoS. Una parte significativa de este capítulo tiene como

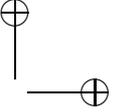
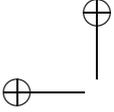


objetivo la comparación del esquema aquí introducido con dos propuestas relevantes anteriores en escenarios monoservicio [ZL01, WZZZ03] con un enfoque que podría considerarse similar al que en este trabajo se utiliza. Zhang et al. en [ZL01] proponen un algoritmo de cuatro parámetros que ajusta la política *Guard Channel*² (GC) a partir de estimaciones de la probabilidad de bloqueo percibida por las peticiones de traspaso. Para ello definen una ventana temporal de dos horas durante la cual el sistema acumula la información necesaria para el cálculo de las estimaciones. Sin embargo, este periodo de medida resulta excesivamente largo si se desea distinguir adecuadamente el comportamiento dinámico de una red celular en operación. Además, el valor de los parámetros propuestos en [ZL01] provocan un funcionamiento incorrecto del algoritmo para ciertos perfiles de tráfico [WZZZ03] (es decir, no se cumplen los objetivos de QoS). Para superar estas limitaciones, Wang et al. en [WZZZ03] proponen un algoritmo adaptativo aleatorizado de tres parámetros similar en cierta manera a *Random Early Detection*.³ Su principal ventaja, comparado con el esquema de [ZL01], es que permite reducir la probabilidad de bloqueo de las peticiones de sesiones nuevas una vez que se ha alcanzado el régimen permanente y por lo tanto, consigue aumentar la utilización de los recursos. No obstante, el periodo de convergencia es todavía del orden de horas. El esquema que se propone en este capítulo, está basado en una adaptación probabilística similar a la de [WZZZ03] pero presenta un tiempo de convergencia considerablemente más corto debido a que no recurre a intervalos de medida temporales para estimar la QoS experimentada por las clases; en su lugar lleva a cabo el ajuste probabilístico basándose exclusivamente en la última decisión del CAS, lo que permite obtener un ajuste continuo de los parámetros de la política. Asimismo, el esquema propuesto en este capítulo se diferencia de las dos propuestas previas por:

1. su simplicidad, puesto que no recurre a intervalos de medida para es-

²La política *Guard Channel* es un caso particular de aplicación de la política *Multiple Guard Channel* a un escenario monoservicio.

³*Random Early Detection* es un algoritmo de gestión de colas que se puede aplicar a la prevención de la congestión.

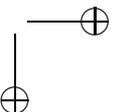
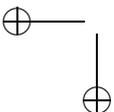


timar la QoS experimentada por cada clase de peticiones.

2. su aplicabilidad tanto a redes monoservicio como multiservicio.
3. el soporte que ofrece a la definición por parte del operador de prioridades para las diferentes clases de peticiones.
4. la posibilidad que ofrece al operador de dejar sin precisar el objetivo de QoS de la clase de peticiones de prioridad ínfima, para convertirla así en una clase «*best effort*» sobre la que puede ser útil concentrar la penalización que ocurre inevitablemente durante los episodios de sobrecarga.
5. la alta precisión con la que se respetan —y se hacen cumplir— los objetivos de QoS aún durante episodios de sobrecarga o subcarga. Así, el tratamiento diferenciado que reciben las diferentes clases logra garantizar que aquellas de mayor prioridad mantendrán sus objetivos, posiblemente a costa de la QoS de las clases de prioridad inferior.

En [YL97, RSAK99, YK02] se estudian mecanismos adaptativos de CAS tanto en escenarios monoservicio como multiservicio, pero en un contexto diferente al estudiado en este capítulo. En estas propuestas, el ajuste de la configuración de la política de CAS se basa en estimaciones del patrón de movilidad y de la tasa de llegada de las peticiones de traspaso, derivadas ambas del número de llamadas en curso en las células vecinas. Sin embargo no se proporciona un guía explícita de cómo debería configurarse la ventana temporal de medida. Otro ejemplo de esquema adaptativos propuestos en un contexto diferente es [HKK04], donde se consideran aplicaciones multimedia de tasa adaptativa. Es aceptable pensar que las prestaciones un esquema como el que se propone aquí aumentarán, bien si se les incorpora bien información de predicción, o bien si se contempla la introducción de mecanismos de degradación progresiva del ancho de banda.

Una observación extensible a la mayoría de los estudios dedicados a esquemas adaptativos es la atención casi exclusiva que dedican al funciona-



miento del esquema en régimen permanente, en contraposición con las escasas evidencias que suelen proporcionar del comportamiento del esquema en régimen transitorio. Es por ello que se juzgue insuficiente la exploración que se ha llevado a cabo en la literatura de una característica fundamental de los esquemas adaptativos como es su velocidad de convergencia a las nuevas condiciones de operación.

5.2 Descripción del modelo

Se considera el caso homogéneo en el que todas las células son estadísticamente idénticas e independientes, por lo que las prestaciones globales del sistema se pueden analizar limitando el enfoque a una única célula. No obstante, dado que el esquema propuesto es adaptativo se podría desplegar en escenarios no homogéneos.

El modelo empleado guarda similitud con el planteado en el Capítulo 4. En cada célula un conjunto de R servicios diferentes compiten por R unidades de recurso, donde el significado de una unidad de recurso depende de la implementación concreta del interfaz radio. Para cada servicio, se distinguen las llegadas de peticiones de nuevas sesiones y de traspasos, lo que define $2R$ clases de peticiones. Por conveniencia, se denota como s_i a la clase i , $1 \leq i \leq 2R$. Adicionalmente se denota como $s_r^n \mid s_r^h$ a la clase asociada a las peticiones de sesión nueva | de traspaso del servicio r , $1 \leq r \leq R$, siendo $s_r^n = s_r$ y $s_r^h = s_{r+R}$, $1 \leq r \leq R$.

Por tratabilidad matemática se efectúan las suposiciones comunes para modelar el sistema como un proceso de Markov. Abusando de la definición de proceso de Poisson, se dirá que para cada servicio r , las peticiones de sesión nueva | de traspaso llegan de acuerdo con un proceso de Poisson con tasa variable con el tiempo $\lambda_r^n(t) \mid \lambda_r^h(t)$. Además, aunque el esquema propuesto no requiere ninguna relación entre $\lambda_r^h(t)$ y $\lambda_r^n(t)$, por simplicidad se supondrá que $\lambda_r^h(t)$ es una fracción constante de $\lambda_r^n(t)$. Las peticiones del servicio r requieren c_r unidades de recurso por sesión. Como cada servicio

tiene asociado dos clases de peticiones, si se denota como d_i a la cantidad de unidades de recurso que requiere una clase por cada sesión, entonces $c_r = d_r = d_{r+R}$, $1 \leq r \leq R$. Para aquellas clases donde las fuentes emiten tráfico de tasa variable, el valor de c_r se corresponde con el ancho de banda efectivo de la sesión [BS97, EE99] (ver también la nota al pie de la página 10).

Tanto la duración como el tiempo de residencia en una célula de una sesión del servicio r están distribuidos exponencialmente con tasas μ_r^s y μ_r^d respectivamente, de ahí que el tiempo de ocupación de recursos de la célula para una sesión del servicio r esté también distribuido exponencialmente con tasa $\mu_r = \mu_r^s + \mu_r^d$. Llegado este punto, es necesario insistir en que la operación del esquema propuesto está basada en las ecuaciones de balance descritas en la Sección 5.3.2, que se verifican para cualquier proceso de llegada y distribución del tiempo de ocupación de recursos.

Se denotan como P_i , $1 \leq i \leq 2R$, a las probabilidades de bloqueo percibidas por las peticiones de la clase i y como $P_r^n = P_r \mid P_r^h = P_r$ la probabilidad de bloqueo percibida por las peticiones de nueva sesión | de traspaso del servicio r . El objetivo de QoS se expresa como cotas superiores para las probabilidades de bloqueo de cada clase de peticiones, por lo que se denota mediante $B_r^n \mid B_r^h$ la cota para las probabilidades de bloqueo de nuevas sesiones | de traspasos. Sea el vector de sesiones en curso $x := (x_1, \dots, x_R)$, donde x_r es el número de sesiones del servicio r en progreso en la célula iniciadas bien como nueva o como traspasada. Se denota como $c(x) = \sum_{r=1}^R x_r c_r$ al número de unidades de recurso ocupadas en el estado x .

Se denota como $\lambda_{\text{máx}}$ a la capacidad del sistema, es decir, la λ máxima que se puede ofrecer al sistema a la vez que se cumplen los objetivos de QoS, donde λ es la tasa agregada de llegadas de peticiones nuevas $\lambda = \sum_{r=1}^R \lambda_r^n$, donde $\lambda_r^n = f_r \lambda$ y $\sum_{r=1}^R = 1$. Definir las penetraciones de cada servicio (f_r) es un enfoque común en el estudio de estos sistemas [BS97].

Tanto la definición de la política MGC como MFGC es idéntica a la presentada en la Sección 4.3.2, aunque se cree necesario incluir a continuación un breve recordatorio por razones de completitud:

MGC (*Multiple Guard Channel*) [CC97, LLC98]

Se asocia un parámetro $l_i \in \mathbb{N}$ a cada clase de peticiones i . Se acepta una sesión de la clase i que llega en el estado x si $c(x) + d_i \leq l_i$ y se rechaza en caso contrario. Equivale a establecer un umbral respecto a la ocupación del sistema, por encima del cual se rechazan las sesiones de dicha clase de peticiones. La particularización de esta política para un sistema monoservicio se conoce como *Guard Channel* (GC) [PG85, HR86].

MFGC (*Multiple Fractional Guard Channel*) [HUCPOG03a]

Se asocia un parámetro $t_i \in \mathbb{R}$ a cada clase de peticiones i . Se acepta una sesión de la clase i que llega en el estado x si $c(x) + d_i \leq \lfloor t_i \rfloor$ donde « \lfloor » y « \rfloor » denotan al entero inmediatamente inferior al que encierran; si lo anterior no se cumple, se acepta con probabilidad $t_i - \lfloor t_i \rfloor$ si $c(x) + d_i = \lfloor t_i \rfloor + 1$; de lo contrario, se rechaza. La particularización de esta política para un sistema monoservicio se conoce como *Fractional Guard Channel* (FGC) [RTN97].

Finalmente, en la Tabla 5.1 figura un resumen de la notación empleada en el modelo.

5.3 Esquema adaptativo

La mayoría de los esquemas adaptativos comentados en la Sección 5.1 para sistemas monoservicio utilizan una estrategia de reserva basada en políticas que hacen uso de «canales de guarda»,⁴ —(*guard channel*)—, aumentando el número de canales de guarda asignados a una clase de peticiones cuando no se consigue el objetivo de QoS acordado. Sin embargo la extensión de este heurístico a un escenario multiservicio no es directa, habida cuenta que el

⁴Se define el número de canales de guarda asignados a la clase de peticiones i por parte de una política MGC | MFGC como $C - l_i$ | $C - t_i$. Nótese que en el caso de la política MFGC, se trata de un número real.

Cuadro 5.1: Notación usada para describir el modelo.

Notación	Significado
C	Número de unidades de recurso.
R	Número de servicios.
r	Índice del servicio.
i	Índice de la clase de peticiones.
s_i	La clase i .
s_r^n, s_r^h	Las clases de peticiones del servicio r nuevas y de traspaso.
$\lambda_r^n(t), \lambda_r^h(t)$	Tasas de llegada para nuevas y traspasos del servicio r .
λ_{max}	Capacidad del sistema.
f_r	Penetración del servicio r .
$1/\mu_r$	Tiempo de ocupación de recursos.
c_r, d_i	Unidades de recurso requeridas por las peticiones del servicio r y las de la clase i .
P_i	Probabilidad de bloqueo percibida por la clase i .
P_r^n, P_r^h	Probabilidades de bloqueo del servicio r (clase de peticiones nuevas y de traspaso).
B_i	Probabilidad de bloqueo objetivo para la clase i .
B_r^n, B_r^h	Probabilidad de bloqueo objetivo para el servicio r (clase de peticiones nuevas y de traspaso).
x_r	número de sesiones en progreso del servicio r .
\mathbf{x}	Vector de sesiones en curso.
$c(\mathbf{x})$	Número de unidades de recurso ocupadas.
$l_r^{n,h}, t_r^{n,h}$	Parámetro de configuración de la políticas MGC y MFGC para el servicio r .
l_i, t_i	Parámetro de configuración de las políticas MGC y MFGC asociados a la clase i .

parámetro de configuración de una política asociado a la clase i no afecta exclusivamente a la QoS percibida (P_i) por dicha clase de peticiones. Para aclarar este punto se incluye la Fig. 5.1 donde se muestra la dependencia de P^n , P^h respecto tanto al parámetro l^n , como l^h de la política MGC.⁵ Los parámetros del sistema monoservicio son los de la Tabla 5.2 con tasa ofrecida $\lambda^n = 0.175$. Como se puede apreciar en la Fig. 5.1(b), claro que la QoS percibida por los traspasos (P^h) se ve notablemente afectada por el valor del parámetro de la política asociado a la clase de peticiones nuevas (l^n). No obstante, la Fig. 5.1(a) muestra que la influencia exacta que ejercen los parámetros de la política sobre la QoS de las clases depende estrechamente de la identidad de la clase, siendo, en ocasiones, despreciable (en este caso el bloqueo experimentado por las peticiones nuevas es «casi» independiente del parámetro l^h). Así, pese a que la extensión del heurístico tradicional a escenarios multiservicio no es exactamente directa, tampoco es razonable pensar que dicha extensión sea considerablemente distinta a la anterior.

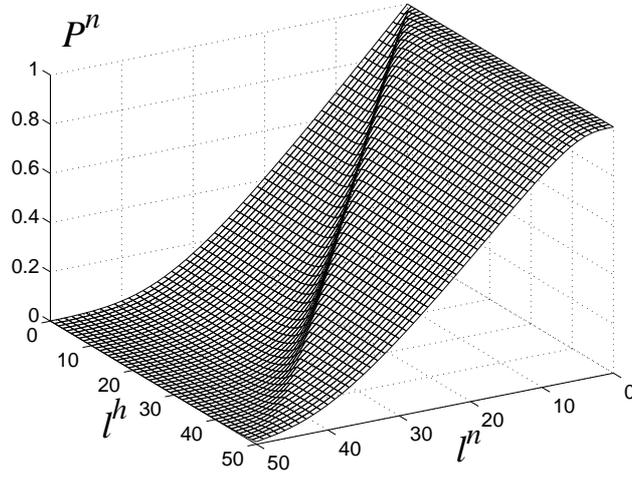
Cuadro 5.2: Configuración de un escenario monoservicio

C	50 (unidades de recurso)
$B^h \%$	1
λ^h	$0.2\lambda^n$ (sesiones/segundo)
μ	$1/180$ (segundos ⁻¹)
c	1 (unidad de recurso · sesión/segundo)

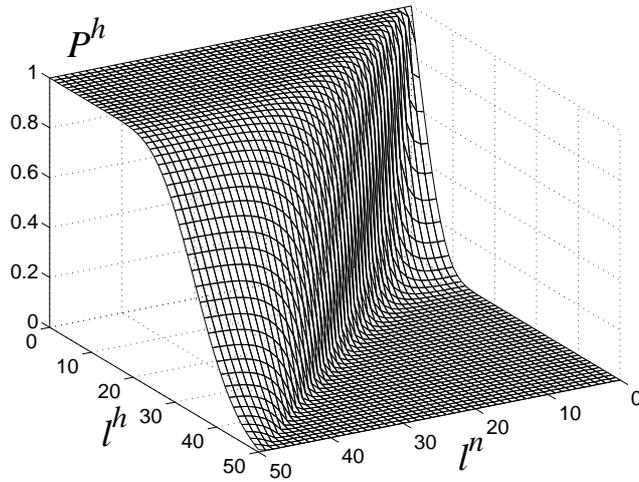
El esquema propuesto se ha diseñado para resolver este extremo. Para este propósito se clasifican las diferentes clases en dos categorías genéricas:

- varias clases «sensibles» para las que se deben cumplir ciertos objetivos de QoS.

⁵Obsérvese que a lo largo de todo el capítulo y para ahorrar notación, en el caso monoservicio ($R = 1$) se sobreentiende P^n , l^h , μ, \dots etc. como P_1^n , l_1^h , μ_1, \dots etc. es decir, en general, sin el valor de r que indicaría el servicio.



(a) Peticiones nuevas



(b) Peticiones de traspaso

Figura 5.1: Dependencia de la probabilidad de bloqueo con los parámetros de configuración de la política MGC.

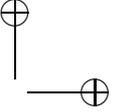
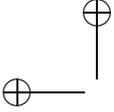
- una clase *best effort* (BEC) sin objetivos de QoS definidos.

Adicionalmente, en un sistema multiservicio el operador puede definir a su voluntad y conveniencia, una sucesión de de prioridades para las clases sensibles con el propósito de dar mayor protección a las clases de importancia superior. Nótese que aunque la BEC percibe una probabilidad de bloqueo impredecible, cada petición que se acepta de dicha clase recibe la cantidad exacta de recursos solicitada, al igual que el resto de clases que acceden al sistema.

Sea $\mathbf{s} = (s_1, \dots, s_{2R})$ el conjunto de todas las clases, $\Pi := \{(\pi_1, \dots, \pi_{2R}) : \pi_i \in \mathbb{N}, 1 \leq \pi_i \leq 2R\}$ el conjunto de todas las posibles permutaciones y $\pi^* \in \Pi$ el orden de prioridades definido por el operador, entonces a $\mathbf{s}^* = (s_{\pi_1}, \dots, s_{\pi_{2R}})$ se le denomina el «orden de prioridad», siendo s_{π_1} la «clase de mayor prioridad» —*Highest Priority Class* (HPC)— y $s_{\pi_{2R}}$ la «clase de menor prioridad» —*Lowest Priority Class* (LPC)—. Si existe una BEC, esta clase será la LPC dentro del orden de prioridad. En este capítulo se considerarán ambas implementaciones: bien que la LPC se considera una clase sensible o bien que se trata de la BEC.

5.3.1 Fundamentos

Por motivos de claridad, durante la descripción de los fundamentos del esquema propuesto se supondrá que los procesos de llegada son estacionarios y que el sistema ha alcanzado el régimen permanente. Si se asume que el objetivo de QoS para la clase i se puede expresar como $B_i = b_i/o_i$, donde $b_i, o_i \in \mathbb{N}$, es de esperar que cuando $P_i = B_i$ la clase i experimentará b_i peticiones rechazadas y $o_i - b_i$ peticiones admitidas, de cada o_i peticiones ofrecidas. Por ejemplo, un objetivo de QoS para clase i de $B_i = 1/100$ implicaría que $b_i = 1$ y $o_i = 100$, por lo que en régimen permanente se percibirán, en media, 1 petición rechazada y 99 peticiones admitidas de cada 100 peticiones ofrecidas al sistema.



Parece intuitivo pensar que el esquema adaptativo debería mantener los parámetros de configuración de la política asociados a aquellas clases que cumplen⁶ sus objetivos de QoS. En consecuencia, considerando valores enteros para los parámetros de configuración —como en el caso de los l_i de la política MGC—, se propone realizar el siguiente ajuste probabilístico cada vez que se procesa la llegada de una petición:

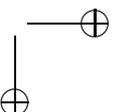
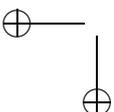
- si se acepta la petición de la clase i entonces $l_i := l_i - \Delta l_i$ con probabilidad $1/(o_i - b_i)$
- si se rechaza la petición de la clase i entonces $l_i := l_i + \Delta l_i$ con probabilidad $1/(b_i)$

Donde $\Delta l_i \in \mathbb{N}$ es un parámetro constante que define la cuantía del ajuste realizado sobre el parámetro de configuración de la política l_i .

Siguiendo este procedimiento de ajuste, cuando se cumple el objetivo de QoS para la clase i , el número de veces que se incrementa el valor de l_i y el número de veces en que se disminuye coinciden, por lo que el resultado sobre l_i —en media— es cero y el parámetro de configuración permanece así fijo. Finalmente, nótese que cuando los procesos de llegada no son estacionarios, el esquema adaptativo ajustará continuamente la QoS percibida por cada clase con el objetivo de cumplir su objetivo, adaptándose a cualquier mezcla de tráfico (es decir, de agregado de tráfico con factores de penetración f_r cualesquiera).

En el caso de la política MFGC, el procedimiento es similar, definiéndose ahora otro parámetro constante $\Delta t_i \in \mathbb{R}$ que especifica la magnitud del ajuste realizado sobre t_i :

⁶En sentido estricto, si una clase de peticiones percibe una probabilidad de bloqueo por debajo del objetivo se puede decir que sus objetivos de QoS se están cumpliendo; excepcionalmente, durante la descripción del esquema adaptativo, «cumplir los objetivos» equivaldrá a que la QoS percibida por la clase iguale el valor de objetivo establecido.



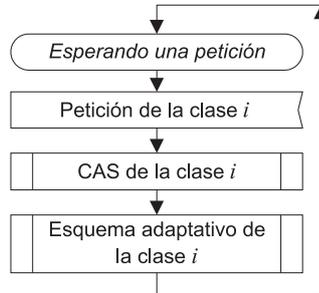


Figura 5.2: Operación del esquema adaptativo de CAS.

- si se acepta la petición de la clase i entonces $t_i := t_i - \Delta t_i$ con probabilidad $1/(o_i - b_i)$
- si se rechaza la petición de la clase i entonces $l_i := l_i + \Delta t_i$ con probabilidad $1/(b_i)$

5.3.2 Funcionamiento

El funcionamiento general del esquema propuesto se muestra en la Fig. 5.2. Cuando llega al sistema una petición de la clase i , el CAS decide su admisión o rechazo y esta decisión se emplea en el esquema adaptativo para ajustar la configuración de la política de CAS.

La Fig. 5.3 muestra con mayor detalle el funcionamiento de la política de CAS y del esquema adaptativo. En la parte izquierda de la Fig. 5.3 se presentan los pormenores de la política de CAS. Como se muestra, para admitir una petición de la clase i primero se comprueba que se dispongan de al menos d_i unidades de recurso libres. Nótese que una vez que se ha verificado este punto, las peticiones de la HPC son admitidas siempre, mientras que para la admisión de peticiones del resto de clases se debe, además, de cumplir la condición de admisión impuesta por la política. En la parte derecha de la Fig. 5.3 están representados los detalles del esquema adaptativo. En general,

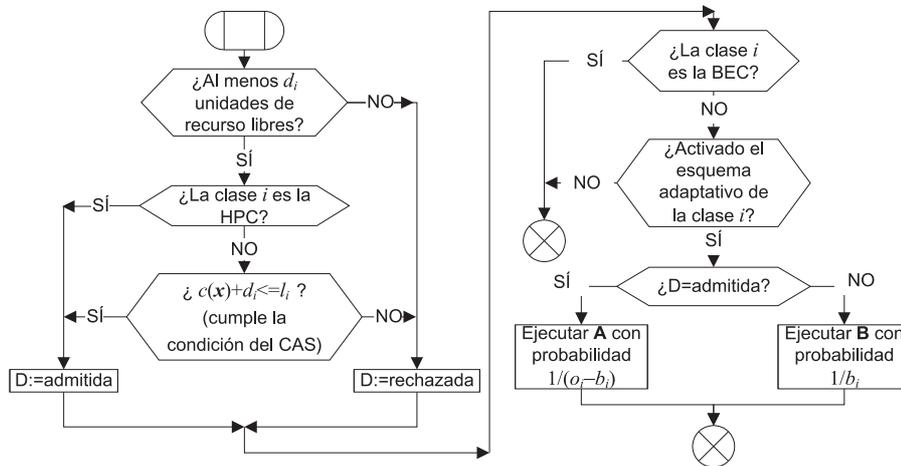


Figura 5.3: Operación del esquema adaptativo de CAS (detalle).

el esquema asociado a cada clase i se encuentra siempre activado (excepto para la BEC), pero para que el esquema adaptativo sea capaz de garantizar que el objetivo de QoS se verifica siempre, en particular durante los episodios de sobrecarga o de cambios en la mezcla de tráfico, el ajuste probabilístico descrito con anterioridad requiere de mecanismos adicionales.

Cuando el objetivo de QoS de la clase i no se cumple, la configuración de la política MGC se puede ajustar mediante dos mecanismos diferentes. La «manera directa» es incrementar l_i (el parámetro de configuración de la política asociado a la clase i). Sin embargo como el valor máximo que este parámetro puede alcanzar es C (es decir, cuando $l_i = C$ se está proporcionando a la clase i el acceso libre a los recursos disponibles), por lo que forzar $l_i > C$ no proporciona beneficios adicionales. En estos casos, una «manera indirecta» de ayudar a la clase i es limitar el acceso a recursos de las clases de menor prioridad mediante una reducción del valor de sus parámetros de configuración asociados. Queda claro que cuando una clase i de mayor prioridad necesita ajustar los parámetros de configuración de una clase j de prioridad menor, el esquema adaptativo debe ajustar l_j sólo cuando ocurren

llegadas de la clase i , mientras que no se deben efectuar ajustes cuando ocurren llegadas de la clase j . Para poder operar de este modo se desactiva el esquema adaptativo asociado a la clase j .

Como se muestra en la Fig. 5.4(b), tras el rechazo de una petición el esquema adaptativo emplea primero el ajuste directo y si el mismo ha dejado de ejercer efecto recurre al ajuste indirecto, en cuyo caso los esquemas adaptativos de las clases de menor prioridad se deben desactivar convenientemente, es decir, comenzando por la LPC y si ello es necesario continuando por las clases de prioridad inmediatamente superior a la del LPC. Esto es necesario porque de lo contrario los esquemas adaptativos asociados a las clases de menor prioridad terminarían eventualmente por deshacer los ajustes indirectos realizados desde las clases de mayor prioridad.

La Fig. 5.4(a) muestra el procedimiento inverso. Nótese que es necesario definir los mecanismos correctos para reactivación de aquellos esquemas desactivados: en este caso, cuando el parámetro de configuración de la clase π_k permite el acceso a recursos, se procede a activar el esquema adaptativo de la clase π_{k-1} . Recuérdese que cuando la LPC es una BEC su esquema adaptativo no se activa nunca. Finalmente, como se aprecia, se permite a l_i tomar valores por encima de C y por debajo de 0 ; de esta manera se ayuda a contabilizar de manera precisa cómo está afectando a l_i el total de ajustes realizados hasta dicho instante.

5.4 Análisis

Las prestaciones de la red móvil celular multiservicio —tanto en régimen transitorio como permanente— se pueden obtener a partir de la «cadena de Markov de tiempo continuo» (CTMC) que describe el funcionamiento del sistema en presencia del esquema adaptativo de CAS propuesto. En ambos regímenes, P_i se define como el porcentaje de tiempo durante el que la llegada una petición de la clase i se rechazaría.

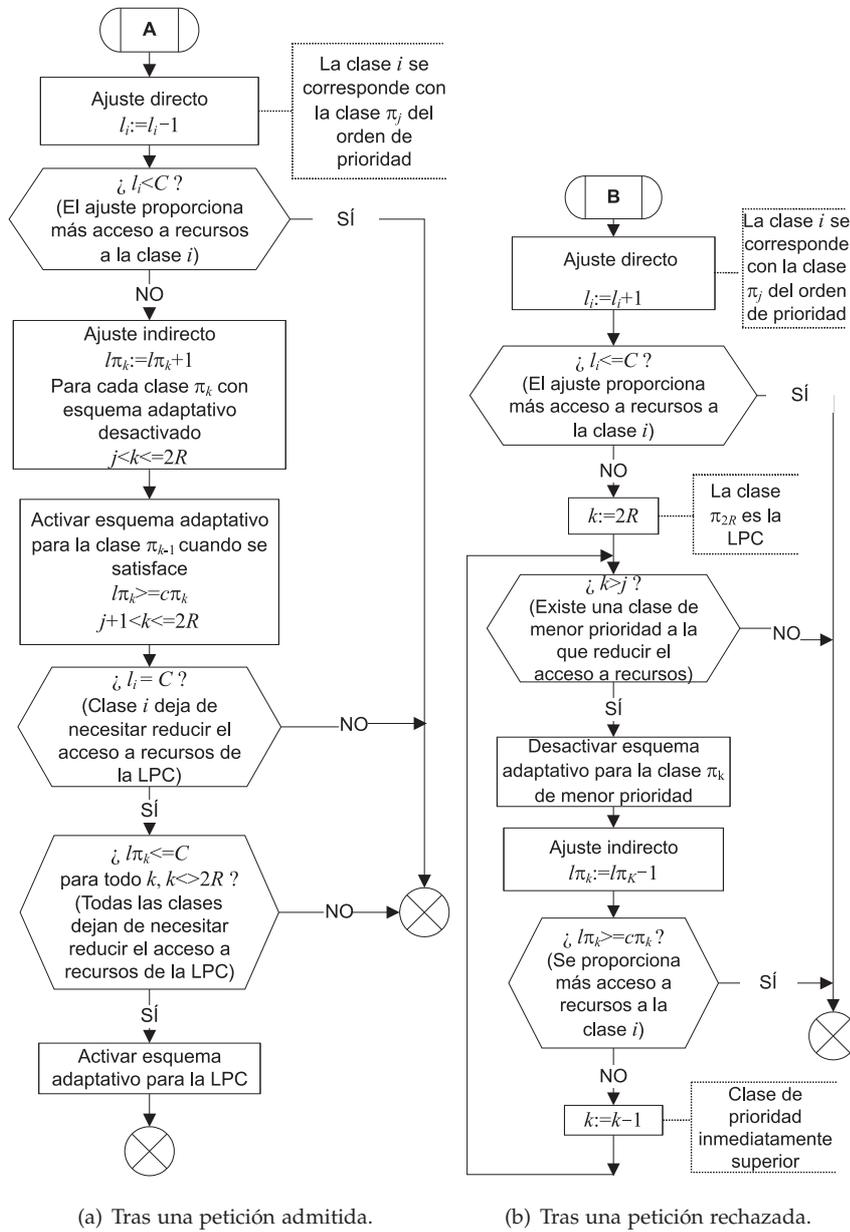


Figura 5.4: Algoritmos de ajuste de la política de CAS.

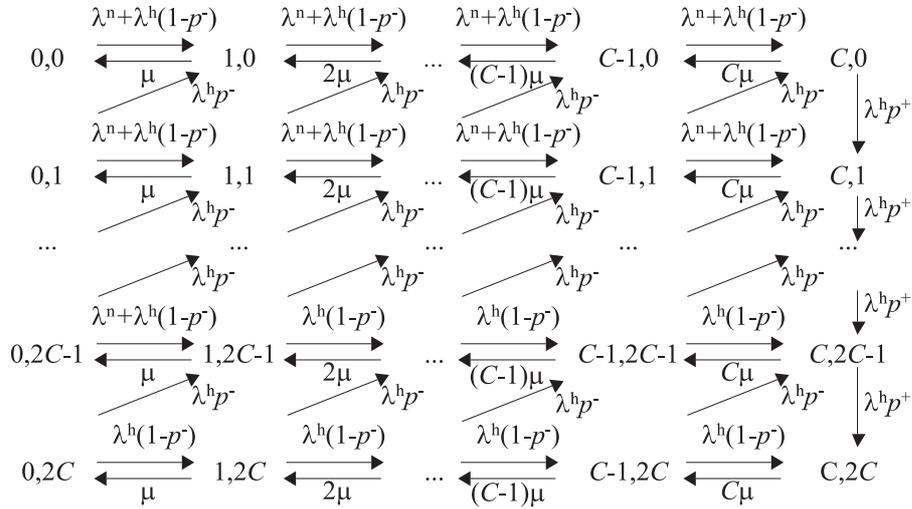
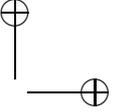
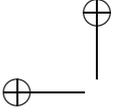


Figura 5.5: Diagrama de estados de la CTMC en un escenario con dos clases.

En general, se tiene una CTMC multidimensional cuyo espacio de estados viene dado por $(x_1, \dots, x_R, l_1, \dots, l_{2R})$. Dado que un diagrama multidimensional de esta cadena sería difícil de representar, se muestra a modo de ejemplo la CTMC bidimensional de la Fig. 5.5. El sistema representado sólo tiene un servicio —y por lo tanto dos clases s^h y s^n —, con $d = 1$ y C unidades de recurso. Se asume que s^h es la HPC y por lo tanto sus peticiones se aceptan siempre que existan suficientes recursos disponibles, mientras que s^n es la BEC. El vector de estado del sistema se define como $\{c(x), l^h\}$, donde $c(x)$ es el número de unidades de recurso ocupadas. En este sistema, l^h se ajusta siguiendo el procedimiento probabilístico descrito con anterioridad, mientras que $l^n = C - \max\{0, (l^h - C)\}$. Nótese que durante los episodios de subcarga $l^n = C$, pero durante los episodios de sobrecarga la clase s^h podría tener que recurrir al manera indirecta de ajustar μ , en cuyo caso l^n se disminuye de la manera correspondiente. Si el objetivo de QoS para s^h se expresa como $B = b/o$, entonces $p^- = 1/(o - b)$ y $p^+ = 1/b$.



5.5 Esquemas adaptativos comparados

De ahora en adelante se hará referencia al esquema propuesto en [ZL01] como ZL, al propuesto en [WZZZ03] como WZZZ, mientras que al esquema que se propone a lo largo de este capítulo se le denominará «nuestro esquema». Los esquemas ZL y WZZZ se diseñaron para la política GC (particularización para un único servicio de la política MGC), y al igual que con el esquema propuesto, las peticiones de traspaso entrantes se aceptan siempre que existan suficientes recursos disponibles en el sistema para cursar la petición.⁷

A continuación se realiza una breve descripción de los esquemas ZL y WZZZ. Asimismo se presentan los detalles de nuestro esquema especificados para un sistema monoservicio, dado que es este caso particular el que se compara en la Sección 5.6.1 con los dos esquemas antes mencionados.

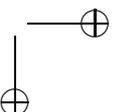
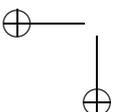
5.5.1 Esquema ZL

El esquema ZL consta de cuatro parámetros, denominados α_u , α_d , N y τ . Su funcionamiento es el siguiente:

- si se rechaza una petición de traspaso y se verifica que $P^h \geq \alpha_u B^h$, entonces $l_i := l_i - 1$; de lo contrario no se realiza ninguna acción
- si se verifica que durante N peticiones de traspaso consecutivas $P^h \leq \alpha_d B^h$, entonces $l_i := l_i + 1$; de lo contrario no se realiza ninguna acción

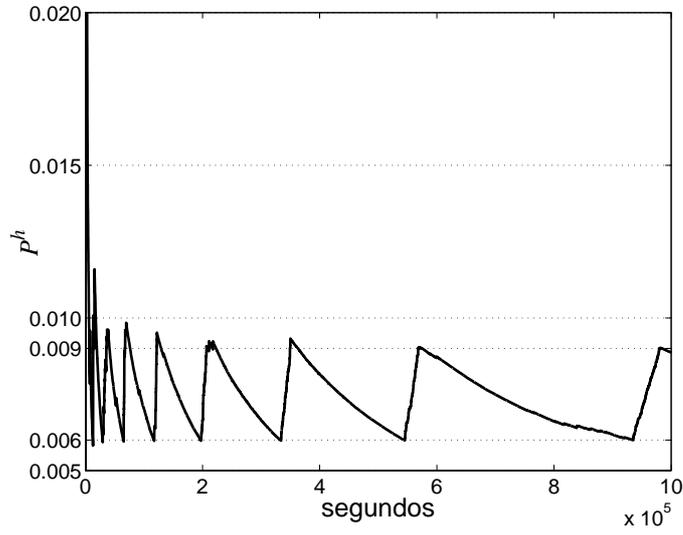
Este esquema, al igual que el esquema WZZZ (ver Sección 5.5.2) estima el cociente de peticiones rechazadas respecto al número total de peticiones de

⁷Se recuerda que en el algoritmo descrito en la Sección 5.3.2 la HPC recibe precisamente dicho tratamiento. Puesto que la interrupción de una sesión en curso es más molesta para el usuario que el rechazo de una nueva sesión, en un escenario monoservicio, la clase de peticiones de traspaso es la HPC.

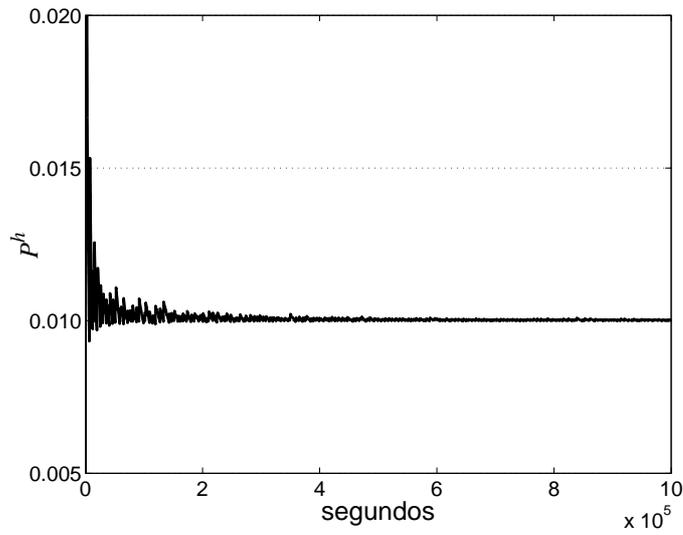


traspaso durante un periodo de actualización τ de tamaño fijo. En [ZL01] se sugiere que, con el fin de obtener una determinada precisión, conforme la probabilidad de bloqueo objetivo sea menor, mayor debe ser el intervalo de estimación. Debido al reducido valor requerido para B^h , este periodo de actualización podría ser excesivamente largo (por ejemplo de $\tau = 2$ horas en el caso del esquema ZL). Sin embargo este esquema define de forma ambigua el «estimador», es decir cómo se mide el cociente de peticiones rechazadas respecto al número total de peticiones de traspaso. Se plantean dos posibilidades: i) el enfoque sencillo consiste en poner a cero el valor del cociente cada τ unidades de tiempo; ii) el enfoque complejo consiste en medir el cociente sobre las últimas τ unidades de tiempo y por tanto necesita mantener una tabla de eventos. Esta ambigüedad lleva a los autores de [WZZZ03] a adoptar $\tau \rightarrow \infty$ en el estudio comparativo que hace de ambos esquemas — ZL y WZZZ— eliminando así la dependencia del parámetro τ (es decir, el mencionado cociente equivale a la probabilidad de bloqueo, P^h , experimentada hasta el momento). Esta alternativa para el estimador se asume en la evaluación comparativa de la Sección 5.6.1, tanto para el esquema ZL como para el WZZZ. Nótese que, en contraste, el esquema propuesto no se basa en intervalos de medida.

Adicionalmente, durante el desarrollo de la evaluación comparativa se ha descubierto también que el esquema ZL no precisa de los parámetros α_u y α_d para obtener las prestaciones deseadas ($P^h \leq B^h$). Los valores que se sugieren en [ZL01] son $\alpha_u = 0.9$ y $\alpha_d = 0.6$. Aunque estos parámetros logran el objetivo de mantener $\alpha_d B^h \leq P^h \leq \alpha_u B^h$, también evitan que P^h alcance el régimen permanente, y por lo tanto P^h oscila entre los dos valores frontera tal y como se muestra en la Fig. 5.6(a). Se ha hallado que usando $\alpha_u = \alpha_d = 1.0$ se permite que el esquema adaptativo alcance el régimen permanente con $P^h = B^h$, tal y como se muestra en la Fig. 5.6(b). Ambas figuras se han obtenido usando los parámetros de la Tabla 5.2 de la página 108 con $\lambda^n = 0.333$. Las modificaciones realizadas sobre los valores de los parámetros facilitan la comparación en igualdad de condiciones entre el esquema ZL y el esquema propuesto en este capítulo. Finalmente, la Sección 5.6.1 también aporta indi-



(a) $\alpha_u = 0.9, \alpha_d = 0.6$



(b) $\alpha_u = \alpha_d = 1.0$

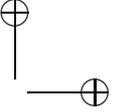
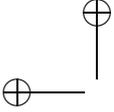
Figura 5.6: Ejemplo de la probabilidad de bloqueo de traspasos para dos instancias de ejecución distintas del esquema ZL con carga estacionaria.

cios de que el parámetro N es perjudicial, puesto que esperar N peticiones de traspaso consecutivas para realizar el ajuste del esquema (con $N > 1$) dilata el tiempo de convergencia del esquema sin obtener ninguna compensación evidente a cambio.

5.5.2 Esquema WZZZ

Para minimizar el número de parámetros, mejorar la adaptabilidad del sistema a diferentes perfiles de tráfico y mejorar el tiempo de respuesta del sistema, en [WZZZ03] se proponen dos nuevos esquemas adaptativos probabilísticos basados en el esquema ZL. Aquí se tratará el primero de ellos, puesto que la especificación que se proporciona del otro en [WZZZ03] no permite su reproducción. El esquema WZZZ consta de tres parámetros: α_u , α_d y P_{inc} . Los dos primeros tienen el mismo significado que en el esquema ZL, mientras que el último de los tres es la probabilidad de disminuir l^n . A modo de resumen sólo se precisará aquí que el esquema WZZZ, ligeramente más complejo que el ZL, realiza ajustes probabilísticos de la configuración del CAS sólo para cada petición de traspaso bloqueada.

La evaluación comparativa muestra, que al igual que con el esquema ZL, los parámetros α_u y α_d no se precisan para hacer cumplir las especificaciones de QoS impuestas para la clase de peticiones de traspaso. Por otro lado, el valor sugerido de $P_{inc} = 0.2$ da la impresión de proteger excesivamente a los trasposos, resultando en valores extremadamente pequeños de P^h , a cambio de P^n muy elevada. Este fenómeno se debe a que P_{inc} controla la velocidad a la que el esquema WZZZ impide a la clase de peticiones de nuevas sesiones el acceso a los recursos. Así, en lugar de los valores recomendados de $P_{inc} < 0.5$, se ha comprobado que tomar $P_{inc} = 1.0$ suele ser un valor razonable. Finalmente, el hecho de que el esquema sólo realice ajustes probabilísticos por cada petición de traspaso *bloqueada* (al contrario que el esquema ZL que los realiza para cada petición de traspaso *ofrecida*) conlleva que la duración del intervalo de convergencia que experimenta el esquema WZZZ sea aún más larga que con el esquema ZL.



5.5.3 Nuestro esquema (caso monoservicio)

Dado que la terminación abrupta de una sesión en curso es más perjudicial para el sistema que el rechazo de una nueva sesión, en un escenario monoservicio, la clase de peticiones de traspaso (s^h) es la HPC y en consecuencia la clase de peticiones de nueva sesión (s^n) se identifica con la LPC.

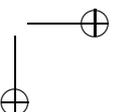
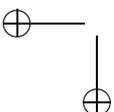
En la comparativa de la Sección 5.6.1 se estudian las dos formas de tratar a la LPC:

1. La LPC tiene objetivos de QoS, a cumplir siempre que sea posible (durante los episodios de subcarga). En este escenario se ajusta tanto el valor de l^h como l^n de la política.
2. La LPC es una BEC sin objetivo de QoS. En este escenario sólo se ajusta l^n , (el valor de l^h no está definido).

Mientras que el segundo enfoque ha recibido mucha atención en la literatura (es el tratamiento que se da en [ZL01, WZZZ03]), el primer enfoque —hasta donde llega el conocimiento del autor— no se ha propuesto con anterioridad.

Particularizar nuestro esquema adaptativo multiservicio (Fig. 5.3) al caso monoservicio da lugar a sendas instancias del mismo asociadas a la clase de peticiones de nuevas sesiones (Fig. 5.7) y a la clase de peticiones de traspaso (Fig. 5.8). Para las peticiones nuevas, la porción izquierda del diagrama de flujo de la Fig. 5.7 muestra el mecanismo de control de admisión: en principio la petición se rechaza si no existen recursos disponibles para satisfacerla,⁸ de lo contrario la aceptación de la misma depende de que se siga cumpliendo la condición impuesta por el control de admisión. La porción derecha de la Fig. 5.7 describe el esquema adaptativo asociado a clase de peticiones nuevas. De entrada, tanto si se ha definido dicha clase como una BEC como si

⁸Nótese que en el caso monoservicio, independientemente de los recursos necesarios para cursar una sesión, se puede suponer que dicha cantidad es la unidad de recursos sin pérdida de generalidad.



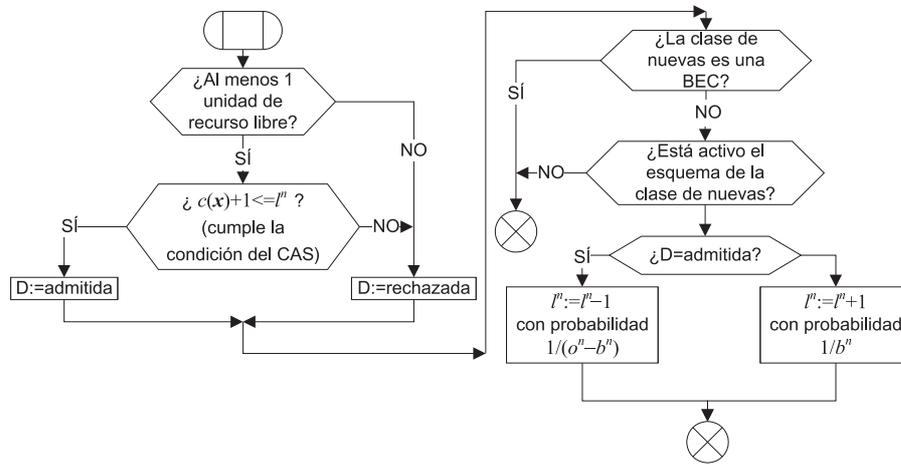


Figura 5.7: Funcionamiento del esquema adaptativo para peticiones nuevas.

el esquema adaptativo se encuentra desactivado, no se realiza modificación alguna de la política. De lo contrario, se procede a llevar a cabo el pertinente ajuste probabilístico. Nótese que al tratarse de la LPC el ajuste para la clase de peticiones nuevas se simplifica significativamente: de hecho sólo consiste en realizar el ajuste directo del parámetro de configuración asociado de la política.

Para las peticiones de traspaso, la porción izquierda del diagrama de flujo de la Fig. 5.8 muestra el mecanismo de control de admisión: al tratarse de la HPC las peticiones de traspaso se admiten siempre, habida cuenta que existan suficientes recursos como para cursar la sesión. La porción derecha de la Fig. 5.8 detalla el esquema adaptativo asociado a la clase de peticiones de traspaso. Los algoritmos de ajuste ejecutados probabilísticamente tras la llegada de una petición de traspaso se detallan en la Fig. 5.9; se tratan de las correspondientes versiones simplificadas de los algoritmos presentados con anterioridad en la Fig. 5.4, debido a que un sistema monoservicio sólo maneja dos clases de peticiones. Como muestra la Fig. 5.9(b) tras el rechazo de una petición de traspaso se lleva a cabo el ajuste directo aumentando el

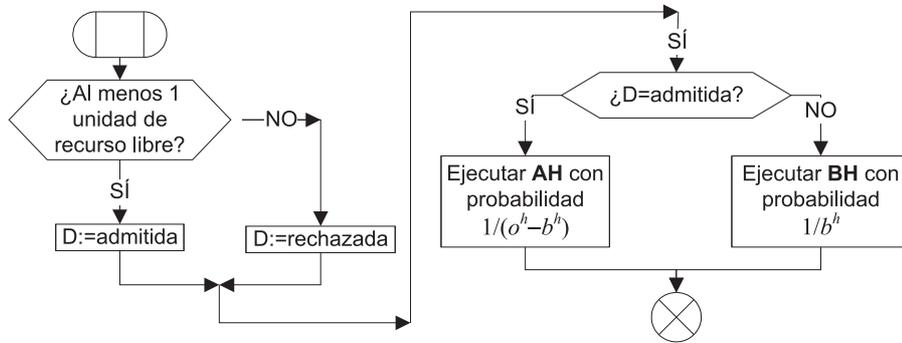


Figura 5.8: Funcionamiento del esquema adaptativo para peticiones de traspaso.

parámetro de la política asociado a dicha clase (aunque recuerdese que esto obedece a motivos de contabilizar el balance entre aceptaciones y rechazos, puesto que no existe control de admisión para los traspasos). A continuación, si es necesario se procede a desactivar el esquema para las peticiones nuevas y se realiza un ajuste indirecto de su parámetro de configuración. Por su lado, la Fig. 5.9(a) muestra la secuencia de acciones tras la aceptación de una petición, que consiste en realizar el ajuste directo, disminuyendo el parámetro de la política para los traspasos, realizar las comprobaciones pertinentes para saber si se requiere un ajuste indirecto sobre el parámetro de las nuevas y a continuación comprobar si deja de ser necesario tener desactivado el esquema adaptativo para las peticiones nuevas, en cuyo caso se vuelve a activar.

5.6 Resultados

Para realizar la evaluación de prestaciones se ha utilizado MöbiusTM [Per06], una herramienta *software* que soporta el modelado mediante «redes de actividad estocástica» —*Stochastic Activity Networks* (SANs) [MMS85]—. MöbiusTM permite simular las SANs que modelan el tipo de sistemas usado en nues-

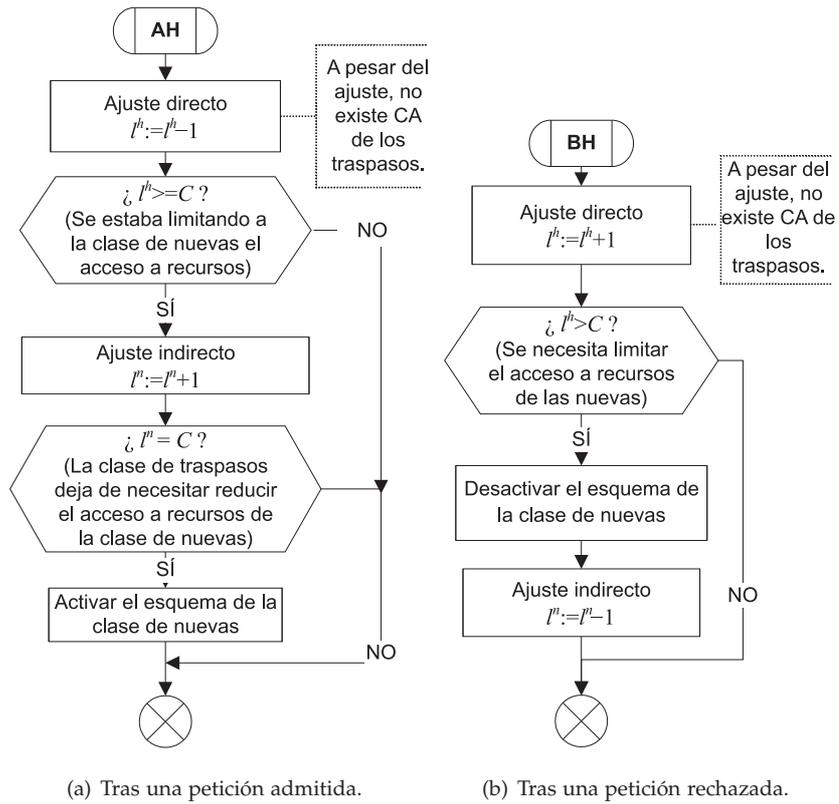
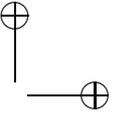
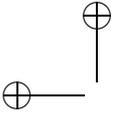


Figura 5.9: Algoritmos de ajuste ejecutados por el esquema para traspasos.

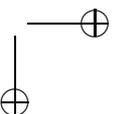
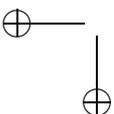


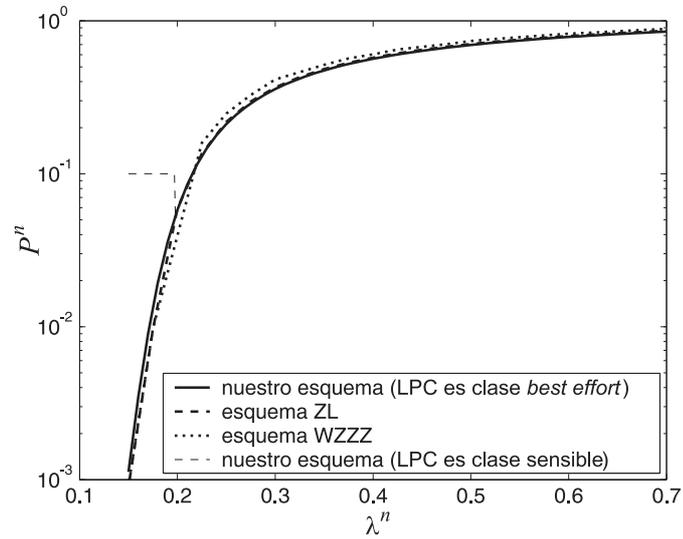
tro estudio, e incluso bajo ciertas condiciones, resolver numéricamente las cadenas de Markov de tiempo continuo asociadas. En particular, para la Sección 5.6.1 se simulan los esquemas ZL y WZZZ (intervalo de confianza de $\pm 5\%$ en torno a la estimación de la media con un nivel de confianza del 95%), mientras que nuestro esquema adaptativo cumple las condiciones específicas para su resolución numérica. En la Sección 5.6.2 los resultados de nuestro esquema adaptativo para la política MGC se obtienen mediante resolución numérica, mientras que los resultados para la política MFGC proceden de simulaciones.

5.6.1 Comparación de esquemas para monoservicio

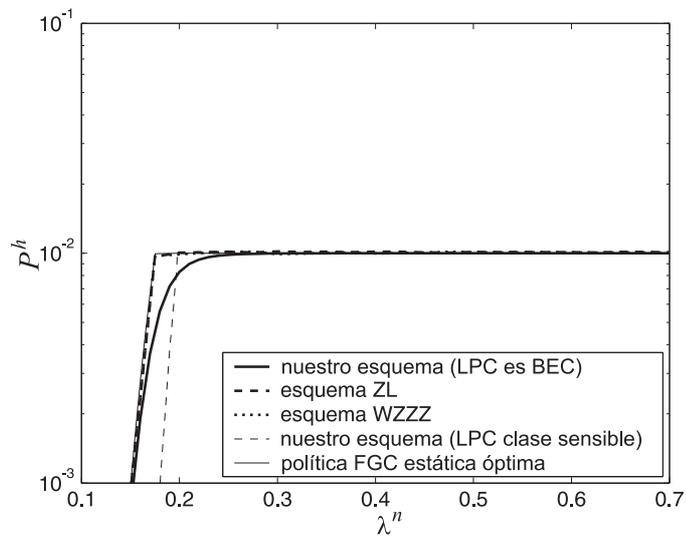
En esta sección se presentan los resultados del estudio comparativo de dos esquemas propuestos con anterioridad para sistemas sistemas monoservicio, ZL y WZZZ, (ver Sección 5.5), respecto a nuestro esquema, tanto en régimen transitorio como estacionario. En ambos regímenes las probabilidades de bloqueo (P^n y P^h) se corresponden con el porcentaje del tiempo en que se ven rechazadas las peticiones que llegan de cada clase respectiva.

Las prestaciones se obtienen para el escenario descrito en la Tabla 5.2 de la página 108. En el caso de nuestro esquema se presentan resultados tanto de la implementación en la que la LPC es una BEC, como resultados exclusivos de nuestro esquema cuando la LPC es una clase sensible, en cuyo caso se asume $B^n = 10\%$. El estudio se centra en el intervalo $\{\lambda : 0.15 \leq \lambda \leq 0.7\}$ (sesiones/segundo), que equivale aproximadamente a $\{\lambda : 10 \leq \lambda \leq 40\}$ (sesiones/minuto) —los valores tomados en [ZL01, WZZZ03]—. Este intervalo permite analizar los esquemas tanto en condiciones de sobrecarga como de subcarga. Por propósitos de comparación, dado que los esquemas ZL y WZZZ se definieron para la política GC, se asume que nuestro esquema ajusta también los parámetros de dicha política.



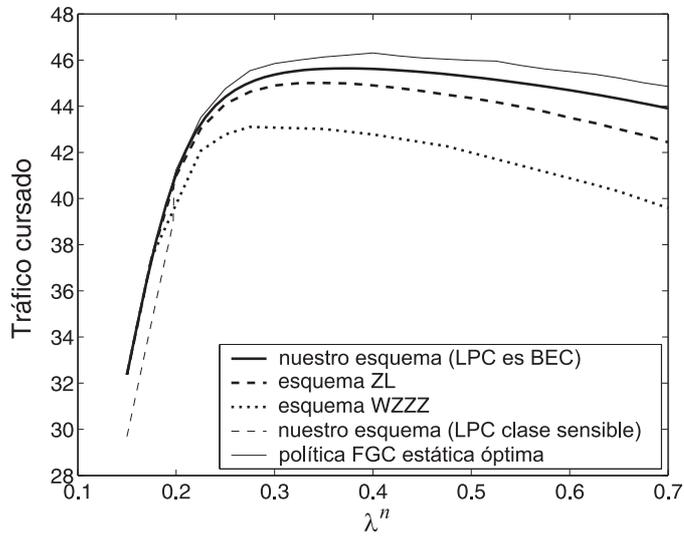


(a) Probabilidad de bloqueo de peticiones de nuevas sesiones.

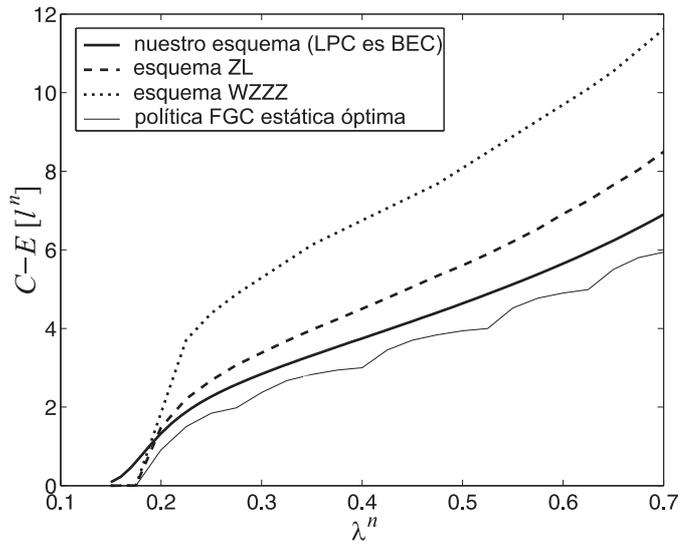


(b) Probabilidad de bloqueo de peticiones de traspaso.

Figura 5.10: Probabilidad de bloqueo con carga estacionaria.



(a) Tráfico cursado.



(b) Número medio de canales de guarda necesarios.

Figura 5.11: Otros parámetros con respecto a la carga estacionaria.

Prestaciones frente a tráfico estacionario

Las Fig. 5.10(a), 5.10(b), 5.11(a) y 5.11(b) muestran la variación de P^n , P^h , el tráfico cursado y el número medio de «canales de guarda» necesarios (equivalente a $C - E[l^n]$) respecto a la tasa de llegada de sesiones nuevas cuando se asume que el tráfico es estacionario. Como se observa, mientras que P^h es similar, P^n es mejor en nuestro esquema. Las Fig. 5.11(a) y 5.11(b) muestran que este mejor comportamiento en la región de carga de interés se debe a una gestión más precisa de los canales de guarda, lo que permite a nuestro esquema cursar más tráfico que el resto. A parte, se ha incluido en las Fig. 5.10(a), 5.10(b) y 5.11(a) los resultados de nuestro esquema cuando la LPC se considera una clase sensible. Como se observa en la Fig. 5.10(a), durante los episodios de subcarga ($\lambda^n < 0.2$) el sistema, para ajustar $P^n = B^n$, tiende a rechazar más peticiones de s^n de las que serían necesarias si se permitiera que $P^n < B^n$ (esta condición seguiría cumpliendo los objetivos), es decir: s^n no se beneficia de la capacidad excedente. La contrapartida positiva de configurar l^n a un valor inferior al necesario es un incremento en la cantidad de recursos a los que tiene acceso s^h , obteniéndose una P^h menor aún, como se muestra en la Fig. 5.10(b). Nótese que si el sistema además, cursara peticiones provenientes de una clase de tráfico elástico, esta clase podría aprovechar el aumento de la capacidad disponible para mejorar sus prestaciones. Además, esta mejora se podría conseguir sin perjudicar el cumplimiento de los objetivos de QoS ni de la clase de peticiones de sesiones nuevas ni de los traspasos. El inconveniente de esta implementación es que el tráfico cursado resulta menor en comparación con la implementación en la que la LPS es una BEC.⁹ De manera significativa, durante los episodios de sobrecarga las dos implementaciones de nuestro esquema funcionan de manera idéntica, (ver Fig. 5.10(a), 5.10(b) y 5.11(a)). Nótese que cuando la LPC es una

⁹En contraste con el tráfico de *streaming*, los requisitos temporales del tráfico elástico son relajados y sus sesiones se pueden ajustar a los recursos disponibles. En consecuencia, parece natural dar prioridad a las clases de tráfico *streaming* (incluida la LPC aunque sea una BEC), y dejar que las clases de tráfico elástico hagan uso de la capacidad excedente.

clase sensible se consigue que $P^n = B^n$ para cierto rango de carga durante el cual $P^h < B^h$. Conforme la carga aumenta, el objetivo de QoS para s^h no se puede cumplir aumentando l^h (ajuste de manera directa) y por tanto, es necesario desactivar el esquema adaptativo de la LPC para poder disminuir l^n (ajuste de manera indirecta), lo que convierte a la LPC en una BEC. En resumen, la capacidad del esquema propuesto para funcionar de dos formas diferentes proporciona una flexibilidad adicional al operador para especificar los objetivos de QoS.

En las Fig. 5.10(b), 5.11(a) y 5.11(b) también se pueden comparar las prestaciones de los diferentes esquemas adaptativos con las prestaciones de la política FGC estática óptima. En la Sección 4.5.1 del capítulo anterior se muestra que la capacidad del sistema obtenida por la política (M)FGC se encuentra cercana a la capacidad lograda por la política óptima. Obsérvese que la evaluación en régimen estacionario se ha llevado a cabo considerando que los procesos de llegada son estacionarios. En este contexto, los parámetros de configuración de la política FGC se han determinado mediante la formulación del problema como una algoritmo de programación no lineal en el que para cada valor de tasa de llegada de sesiones nuevas se realiza la búsqueda del valor de la configuración de la política que maximiza el tráfico sin incumplir los objetivos de QoS. De ahí que se denomine «política FGC estática óptima». Por otro lado, el adjetivo de *estática* procede de que para cada valor de tasa de llegadas, el conjunto de parámetros de configuración óptimos que se determinan es único. Por el contrario, en el caso de los esquemas adaptativos, éstos ignoran, a priori, las tasas de llegada y por lo tanto deben ajustar continuamente los parámetros de configuración de la política GC para hacer cumplir los objetivos de QoS.

Prestaciones frente a tráfico no estacionario

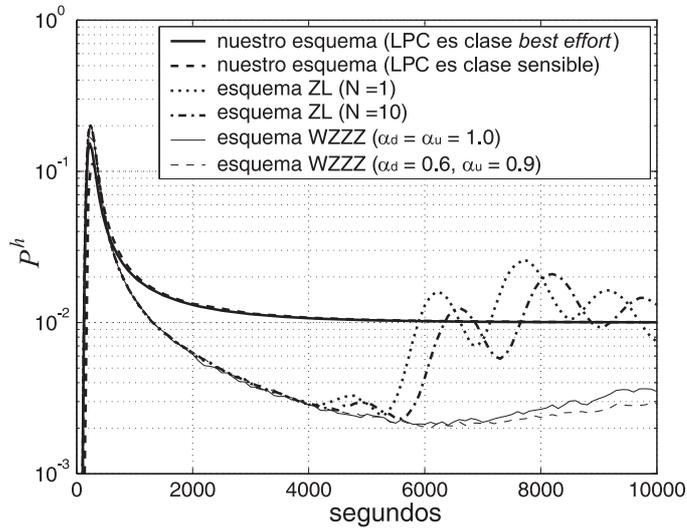
Con el fin de evaluar las prestaciones de cada esquema en régimen transitorio, se muestra primero el comportamiento de los esquemas ante un incremento repentino en el tráfico ofrecido, desde $\lambda^n = 0$ hasta $\lambda^n = 0.333$ (20

sesiones/minuto). Antes de que el incremento repentino se aplique, el sistema se considera en régimen permanente, (en este caso sin sesión alguna en curso).

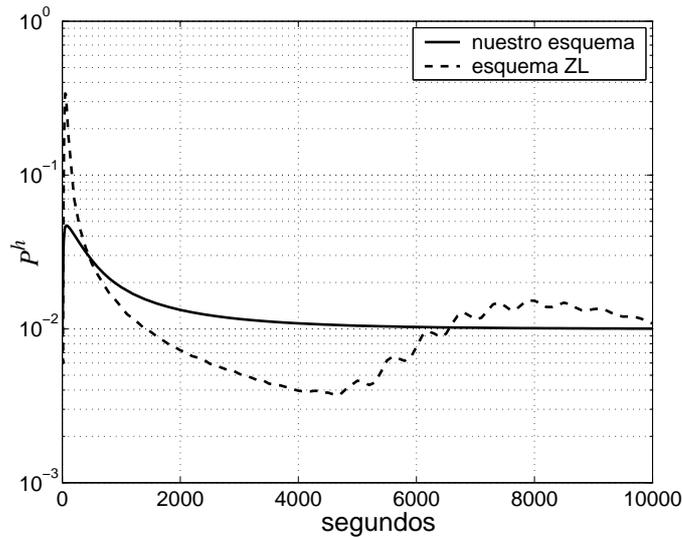
La Fig. 5.12(a) muestra el comportamiento transitorio de la probabilidad de bloqueo de peticiones de traspaso. Como se puede observar, nuestro esquema converge significativamente más rápido (tanto en el caso en que la LPC es una clase sensible como cuando es una BEC). Por el contrario, el esquema ZL muestra un comportamiento más lento además de oscilante alrededor de B^h (tanto con $N = 1$ como con $N = 10$). Así, mientras nuestro esquema requiere sólo $t = 3400$ segundos para reducir P^h a un intervalo del $\pm 10\%$ en torno a su objetivo ($B^h = 0.01$), el esquema ZL necesita $t \approx 30000$ segundos, casi diez veces más tiempo para conseguir lo mismo. Nótese que el esquema con $N = 10$ tarda aún más en converger que en el caso de $N = 1$. Finalmente el esquema WZZZ (con $P_{inc} = 0.2$) provoca que la probabilidad de bloqueo de los traspasos oscile de la misma forma que ZL pero con un tiempo de convergencia excesivamente superior a los anteriores.

Como un sistema inicialmente vacío es improbable, se evalúa ahora un escenario más realista en el que se estudiará el comportamiento transitorio de los esquemas frente a un incremento repentino en la relación λ^h/λ^n , desde 0.2 hasta 0.4, manteniendo $\lambda^n = 0.417$ (25 sesiones/minuto). De nuevo, antes de que se aplique el incremento, el sistema se encuentra en régimen permanente. Como el esquema WZZZ no tiene una velocidad muy competitiva no se tendrán en cuenta sus resultados. La Fig. 5.12(b) muestra el comportamiento transitorio de P^h usando nuestro esquema con la LPC como BEC, frente al esquema ZL con $N = 1$. De nuevo, nuestro esquema supera al esquema ZL en términos de rapidez y estabilidad. Nótese que el periodo de convergencia se suele reducir en situaciones de tráfico elevado debido a que aumenta la frecuencia con que se realizan acciones de ajuste probabilístico por parte del esquema.

El motivo de las inferiores prestaciones de los esquemas ZL y WZZZ en régimen transitorio es consecuencia directa de la ventana temporal —en al-



(a) Incremento repentino de $\lambda^h = 0$ a $\lambda^h = 0.333$.



(b) Incremento repentino de $\lambda^h/\lambda^n = 0.2$ a $\lambda^h/\lambda^n = 0.4$, con $\lambda^n = 0.417$

Figura 5.12: Comportamiento transitorio de los esquemas adaptativos en presencia de un incremento repentino en el tráfico ofrecido.

gunos casos denominada periodo de actualización (*update period*)— durante la que acumulan la información para generar las estimaciones de la QoS percibida (P^h). Hay que tener en cuenta que si el periodo es corto, el esquema se adapta rápidamente a nuevas condiciones de tráfico, pero posiblemente a costa de un empeoramiento de las prestaciones (precisión con la que consiguen el objetivo de QoS). En cambio, si el periodo es largo las prestaciones conseguidas pueden mejorar, pero a costa de una reducción en la velocidad de reacción del esquema; (se debe recordar que un esquema adaptativo debería actuar lo suficientemente rápido como para gestionar adecuadamente la dinámica de una red en funcionamiento real). Dado que las probabilidades de bloqueo objetivo consideradas aquí son del orden de 10^{-2} , el periodo de actualización necesario para que los esquemas ZL y WZZZ consigan buenas prestaciones acaba siendo excesivamente largo, del orden de horas [ZL01]. De ahí que el esquema de adaptación que se propone en este capítulo destaque por su simplicidad, dado que no requiere de intervalos de medida para estimar la QoS percibida por cada flujo, y ni siquiera necesidad propiamente dicha de estimar la QoS experimentada hasta el momento.

Los resultados anteriores asumen que el objetivo de QoS está expresado como fracción irreducible (en este caso $B^h = b^h/o^h = 1/100$) sin embargo, se podría haber escogido cualquier otro múltiplo $k \geq 1$ de la fracción irreducible, es decir, $B^h = kb^h/ko^h$. La Fig. 5.13 muestra (en el mismo escenario no estacionario de la Fig. 5.12(a) cuando la LPC es una BEC), que cualquier valor de k permite que nuestro esquema logre igualar (eventualmente) la P^h percibida a su correspondiente objetivo B^h . Sin embargo como se aprecia, la velocidad de convergencia de P^h decrece a medida que el valor de k aumenta. De la misma manera, la Fig. 5.14 muestra que la velocidad de convergencia de P^n también disminuye a medida que k crece, sin ofrecer una compensación clara a cambio de esta reducción. Es por ello razonable afirmar que el valor de $k = 1$ se confirma como el más adecuado de cara a obtener buenas prestaciones en régimen transitorio.

También hasta el momento se ha considerado un valor de $\Delta l^n = 1$ para la magnitud del ajuste que lleva a cabo nuestro esquema, pero la Fig. 5.15

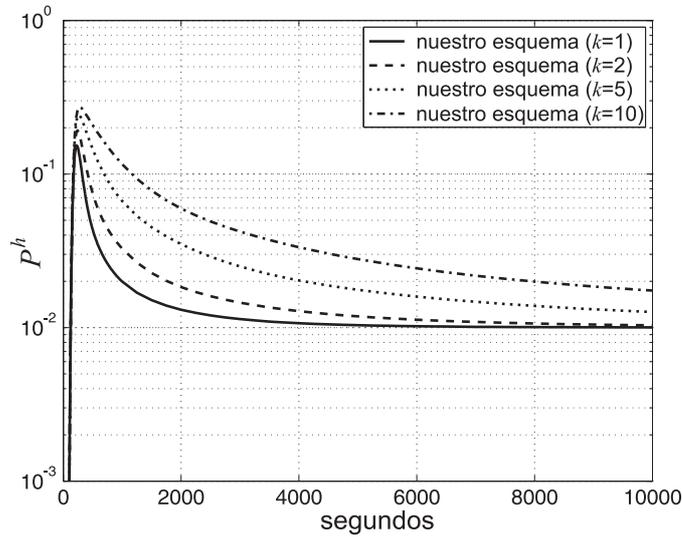


Figura 5.13: Dependencia con k del intervalo de convergencia para P^h .

muestra (también en el mismo escenario no estacionario de la Fig. 5.12(a) cuando la LPC es una BEC), que a medida que Δl^n crece, la velocidad de convergencia de nuestro esquema aumenta también. No obstante, como se observa en la Fig. 5.16, existe un compromiso a tener en cuenta entre la velocidad de convergencia de P^h y el empeoramiento que sufre P^n , aunque valores $2 \leq \Delta l^n \leq 5$ garantizan un incremento significativo de la velocidad de convergencia sin relativo menoscabo del valor experimentado de P^n . Como se aprecia, en el caso de $\Delta l^n = 5$, el esquema propuesto es capaz de adaptarse a las nuevas condiciones de operación en menos de 700 segundos. Dado que en una situación de funcionamiento real no es de esperar que las condiciones cambien de manera tan repentina a la ilustrada en esta sección, es razonable pensar que nuestro esquema gestiona de manera satisfactoria la no estacionariedad de las redes reales.

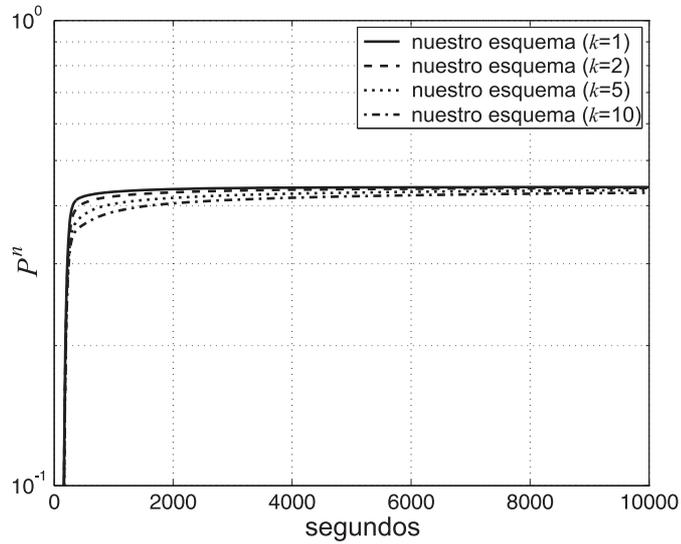


Figura 5.14: Dependencia con k del intervalo de convergencia para P^n .

5.6.2 Evaluación de prestaciones para multiservicio

Dado que los esquemas ZL y WZZZ se diseñaron para la política GC en sistemas monoservicio, esta sección pasa a centrarse exclusivamente en la evaluación de prestaciones de nuestro esquema para sistemas multiservicio considerando tanto el empleo de la política MGC como de la política MFGC.

La evaluación de prestaciones se lleva a cabo para los cinco escenarios diferentes (A–E) definidos en la Tabla 5.3, con los parámetros de QoS B_i expresados como valores porcentuales. Las configuraciones de la Tabla 5.3 se han seleccionado para explorar posibles tendencias en los resultados numéricos, es decir, tomando el escenario A como referencia, el escenario B representa el caso en el que la razón c_1/c_2 es menor, el escenario C aquel en que f_1/f_2 es menor, el escenario D aquel en que B_1/B_2 es menor y finalmente el escenario E el caso en el que B_1 y B_2 son iguales.

La capacidad del sistema cuando se despliegan las políticas MGC y MFGC

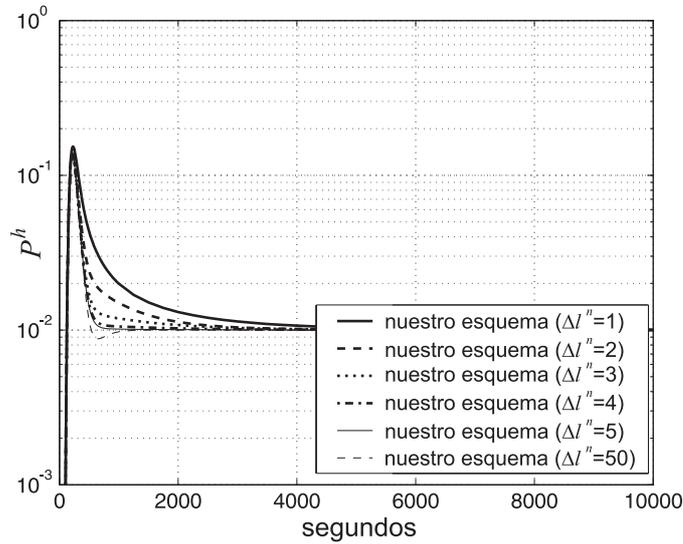


Figura 5.15: Dependencia con Δl^n del intervalo de convergencia para P^h .

para las cinco configuraciones del sistema definidas en la Tabla 5.3, con $C = 10$ son $\lambda_{\text{máx}}^{\text{MGC}} = \{1.89, 0.40, 1.52, 1.97, 1.74\}$ y $\lambda_{\text{máx}}^{\text{MFGC}} = \{2.05, 0.42, 1.65, 2.02, 1.80\}$, respectivamente (consultar Tabla 4.2 de la página 72). Para todos los escenarios definidos en la Tabla 5.3 se asume el siguiente orden de prioridad $s^* = (s_2^h, s_1^h, s_2^n, s_1^n)$. Al igual que en la sección anterior, se evalúan dos implementaciones del esquema adaptativo que se diferencian en el tratamiento de la LPC (s_1^n), una en la que se trata de una clase sensible y una en la que es una BEC.

Prestaciones frente a tráfico estacionario

Para las dos implementaciones del esquema adaptativo en asociación con la política MGC la Tabla 5.4 muestra el cociente P_i/B_i para las cuatro clases en los cinco escenarios. En todos los casos, se ofrece una carga agregada igual a la capacidad del sistema ($\lambda_{\text{máx}}$). Nótese que el ajuste es significativamen-

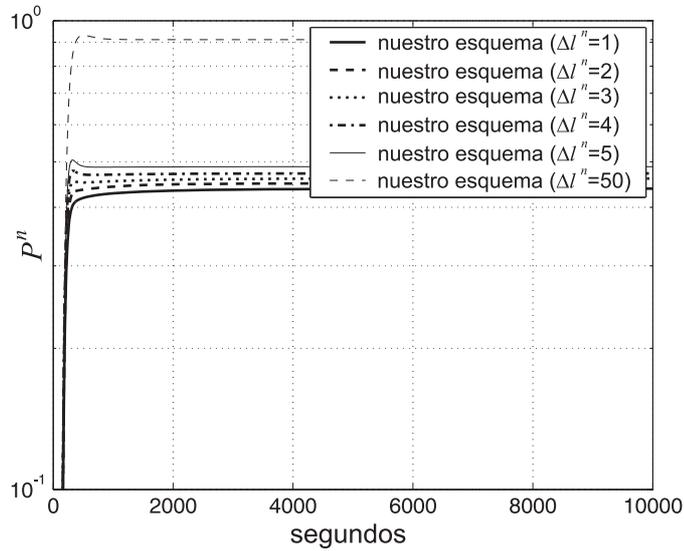


Figura 5.16: Dependencia con Δl^n del intervalo de convergencia para P^n .

te más preciso cuando la LPC es una BEC. Como la política MGC tiene una granularidad relativamente grande, se estudia también el comportamiento de la política MFGC. Para las dos implementaciones del esquema, la Tabla 5.5 muestra las prestaciones del esquema adaptativo cuando se emplean valores de la magnitud de ajuste $\Delta t_i = 0.1$ y $\Delta t_i = 0.01$. Como se observa, a medida que la granularidad del ajuste disminuye, las prestaciones mejoran significativamente.

Las Fig. 5.17(a) y 5.17(b) muestran la variación de P_i con la carga relativa medida respecto a la capacidad $((\lambda - \lambda_{\text{máx}}) / \lambda_{\text{máx}})$ en el escenario C con $C = 10$ unidades de recurso. Como se aprecia, el esquema adaptativo trata de hacer cumplir la igualdad $P_i = B_i$ para las clases sensibles, siempre que esto sea posible, y por lo tanto durante los episodios de subcarga el sistema precisa rechazar más peticiones de las que serían estrictamente necesarias para hacer cumplir los objetivos ($P_i \leq B_i$). No obstante algunas clases (la BEC

Cuadro 5.3: Definición de los escenarios estudiados

	c_1	c_2	f_1	f_2	B_1^n (%)	B_2^n (%)	B_r^h (%)	λ_r^n	λ_r^h	μ_1	μ_2
A	1	2	0.8	0.2	5	1					
B	1	4	0.8	0.2	5	1					
C	1	2	0.2	0.8	5	1	$0.1B_r^n$	$f_r\lambda$	$0.5\lambda_r^n$	1	3
D	1	2	0.8	0.2	1	2					
E	1	2	0.8	0.2	1	1					

y/o la HPC) se benefician de esta capacidad extra. Cuando la LPC es una clase sensible, ver Fig. 5.17(a), no se beneficia de este exceso de capacidad durante los episodios de subcarga y es la primera en ser penalizada durante los episodios de sobrecarga. En contraste, cuando la LPC es una BEC, ver Fig. 5.17(b), se beneficia durante los episodios de subcarga y, al igual que antes, es la primera en sufrir penalización durante los episodios de sobrecarga. Obsérvese que en ambas implementaciones, también se penaliza a la clase s_2^n cuando la disminución del parámetro l_1^n resulta ineficaz ($l_1^n < 0$) para cumplir los objetivos de QoS de las clases de mayor prioridad.

La Fig. 5.18 muestra $E[c(x)]/C$ (el factor de utilización de los recursos), conseguido por una política estática MGC y por el esquema adaptativo para políticas MGC y MFGC. El término «adapt. M(F)GC» hace referencia al esquema adaptativo cuando la LPC es una clase sensible, mientras que «adapt. M(F)GC-BES» se refiere al esquema adaptativo cuando la LPC es una BEC. En la leyenda, el valor numérico que figura a continuación del esquema adaptativo para política MFGC indica la magnitud del ajuste (Δt_i). La configuración de la política MGC estática es la misma para todos los valores de carga y se corresponde con la configuración del escenario D para la que la carga ofrecida máxima coincide con la capacidad. La pobre utilización de los recursos proviene de los estrictos objetivos de QoS que se imponen. Como se observa, la utilización conseguida por el esquema adaptativo es similar a la que obtiene la política estática durante los episodios de subcarga. Durante los episodios

Cuadro 5.4: P_i/B_i del esquema adaptativo para MGC y carga estacionaria igual a la capacidad.

a) LPC es una clase sensible.

P_i/B_i	Escenario				
	A	B	C	D	E
Class 1N	1.004	1.030	1.036	1.841	1.223
Class 2N	0.998	0.992	1.001	0.998	1.007
Class 1H	1.006	0.992	1.002	1.007	0.999
Class 2H	0.848	0.899	0.803	0.988	0.985

b) LPC es una clase *best effort*

P_i/B_i	Escenario				
	A	B	C	D	E
Class 1N	0.938	1.404	0.007	2.348	1.857
Class 2N	1.003	1.065	1.000	1.004	0.999
Class 1H	1.007	1.001	1.007	0.999	0.999
Class 2H	0.993	1.006	0.988	0.989	0.999

Cuadro 5.5: P_i/B_i del esquema adaptativo para MFGC y carga estacionaria igual a la capacidad.

a) $\Delta t_i = 0.1$, LPC es una clase sensible.

P_i/B_i	Escenario				
	A	B	C	D	E
Class 1N	0.915	0.248	0.011	2.682	2.696
Class 2N	0.999	1.002	1.000	1.005	0.991
Class 1H	1.001	0.999	0.977	0.962	0.963
Class 2H	0.461	0.934	0.750	0.520	0.464

b) $\Delta t_i = 0.1$, LPC es una clase *best effort*.

P_i/B_i	Escenario				
	A	B	C	D	E
Class 1N	0.285	0.069	0.005	3.710	0.524
Class 2N	1.005	0.997	1.000	0.999	0.993
Class 1H	1.000	1.006	0.980	1.011	0.987
Class 2H	1.038	0.940	0.746	0.425	0.943

c) $\Delta t_i = 0.01$, LPC es una clase sensible.

P_i/B_i	Escenario				
	A	B	C	D	E
Class 1N	0.229	0.928	0.003	0.664	0.437
Class 2N	1.007	1.001	0.999	1.001	1.005
Class 1H	1.007	0.995	0.972	0.998	0.976
Class 2H	1.063	0.894	0.666	1.024	1.022

d) $\Delta t_i = 0.01$, LPC es una clase *best effort*.

P_i/B_i	Escenario				
	A	B	C	D	E
Class 1N	0.237	0.013	0.004	0.658	0.427
Class 2N	1.002	0.999	0.999	1.001	1.004
Class 1H	1.004	1.002	0.988	0.995	0.995
Class 2H	1.045	1.070	0.667	1.030	1.032

de sobrecarga la utilización del esquema es menor debido a que rechaza peticiones de las clases de menor prioridad para respetar el objetivo de QoS de las clases de mayor prioridad. Nótese que ambas implementaciones del esquema adaptativo se comportan de manera idéntica durante sobrecargas ($(\lambda - \lambda_{\text{máx}})/\lambda_{\text{máx}} > 0$).

Prestaciones frente a tráfico no estacionario

En esta sección se estudia el comportamiento en régimen transitorio del esquema adaptativo con política MGC, después de aplicar un incremento «súbito» de tráfico ofrecido —es decir, con forma de escalón— desde $0.66\lambda_{\text{máx}}$ hasta valores de $0.77\lambda_{\text{máx}}$ (incremento del 10%) y $1.00\lambda_{\text{máx}}$ (incremento del 50%). El sistema está configurado según el escenario A con la LPC como clase sensible. Antes de aplicar el incremento en escalón, el sistema se encuentra en régimen permanente. La resolución numérica del transitorio se lleva a cabo empleando «aleatorización» (*randomization*), también conocida como «método de Jensen» [dSeSG00]. El método consiste en transformar la cadena de Markov de tiempo continuo en su análogo de tiempo discreto y está implementado en Möbius.

La Fig. 5.19 muestra el comportamiento transitorio de las probabilidades de bloqueo. Como se observa, la velocidad de adaptación es relativamente independiente de la magnitud del incremento de tráfico. El periodo de convergencia se mantiene en los dos casos por debajo de los 1000 segundos y como se puede apreciar sólo requiere en torno varios centenares de segundos para acercarse a menos del 10% del valor de QoS objetivo; asimismo, es destacable que no presenta un comportamiento oscilatorio.

5.7 Conclusiones

En este capítulo se ha propuesto un esquema adaptativo de control de admisión de sesiones diseñado para operar bajo condiciones de tráfico no estacio-

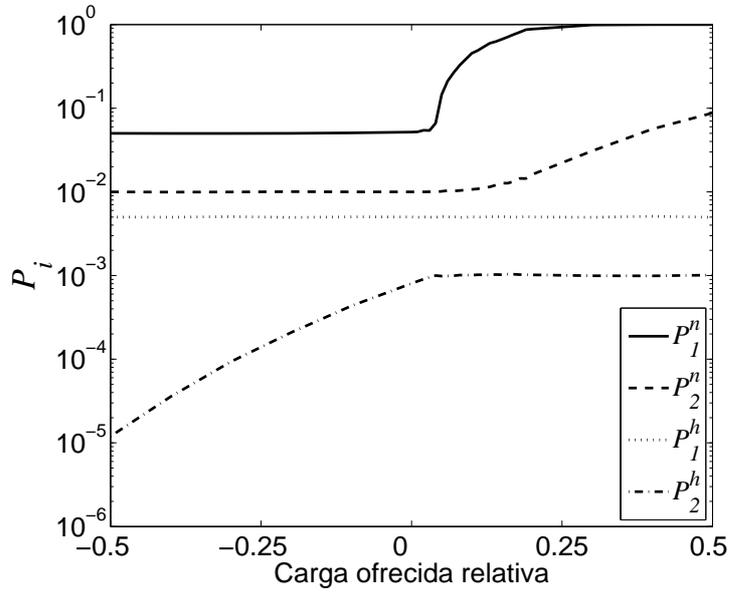
nario en asociación con dos de las políticas más empleadas en sistemas móviles multiservicio (*Multiple Guard Channel* y *Multiple Fractional Guard Channel*). El esquema asegura unos ciertos objetivos de QoS definidos libremente por el operador de antemano y expresados en términos de cotas superiores para las probabilidades de bloqueo que percibe cada clase de petición. La operación del esquema no recurre a intervalos temporales de medida para estimar las prestaciones del sistema sino que se basa en ecuaciones de balance simples que se verifican para cualquier proceso de llegada y distribución de los tiempos de ocupación.

Rasgos relevantes de la propuesta son: su habilidad para manejar tráfico *streaming* multiservicio, su capacidad para realizar un seguimiento y ajuste continuo de la QoS percibida por los usuarios, la inclusión de un mecanismo para establecer un tratamiento diferenciado entre clases durante los episodios de subcarga y sobrecarga y su sencillez de implementación. Se han proporcionado dos implementaciones del esquema: en la primera, la LPC tiene un objetivo de QoS definido, que se debe cumplir siempre se que esto sea posible; en la segunda, la LPC se trata como una clase *best effort* y por lo tanto obtiene una QoS impredecible (que tiende a ser relativamente aceptable durante los episodios de subcarga pero que se degrada significativamente tan pronto como el sistema entra en la región de sobrecarga).

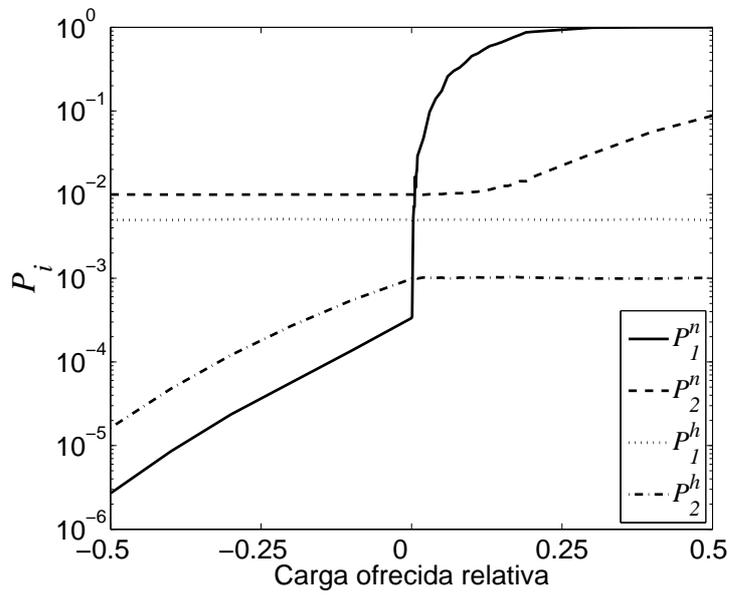
El esquema adaptivo se ha comparado con dos esquemas propuestos con anterioridad en la literatura. Sin embargo, para llevar a cabo la comparación, se ha necesitado particularizar el esquema adaptativo a un sistema monoservicio, puesto que los esquemas anteriores sólo se definieron para este tipo de escenarios. Estos dos esquemas —de manera similar al que se propone en este capítulo— recurren sólo a información acerca del estado de la célula donde se ejecuta el propio esquema para ajustar la configuración de la política de control de admisión. El objetivo del estudio ha sido doble: 1) comparar la capacidad que logran diferentes esquemas adaptativos con la capacidad de la política «casi» óptima; y 2) evaluar el régimen transitorio de los esquemas. El extenso estudio comparativo muestra que el esquema propuesto iguala los objetivos con una precisión excelente a la vez que se consigue tanto un trá-

fico cursado mayor como un periodo de convergencia libre de oscilaciones y de diez a cien veces más corto que en propuestas anteriores. Se proporciona asimismo un mecanismo adicional para reducir la duración del periodo de convergencia, permitiendo establecer un compromiso entre la velocidad a la que las probabilidades de bloqueo percibidas tienden al objetivo y la precisión con que se logra dicho objetivo de QoS.

Finalmente, la evaluación del esquema propuesto sobre varias configuraciones de un sistema multiservicio confirma las características anteriores de precisión y tiempo de convergencia y confirma que el esquema puede manejar satisfactoriamente la no estacionariedad del tráfico que se ofrece a las redes reales.



(a) La LPC es una clase sensible.



(b) La LPS es una clase *best effort*.

Figura 5.17: Variación de P_i con la carga en condiciones estacionarias.

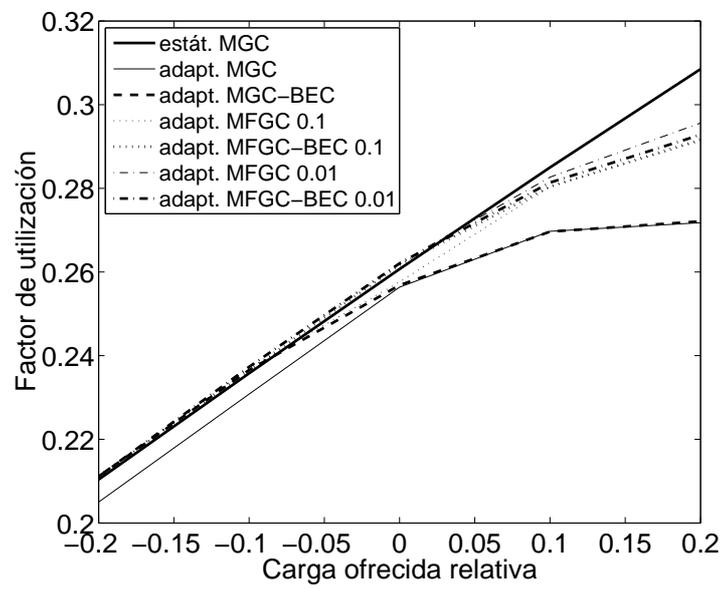
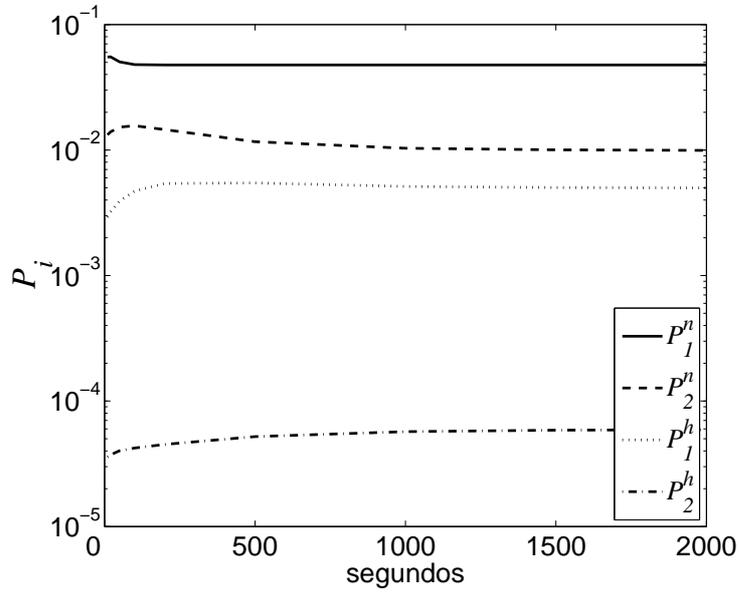
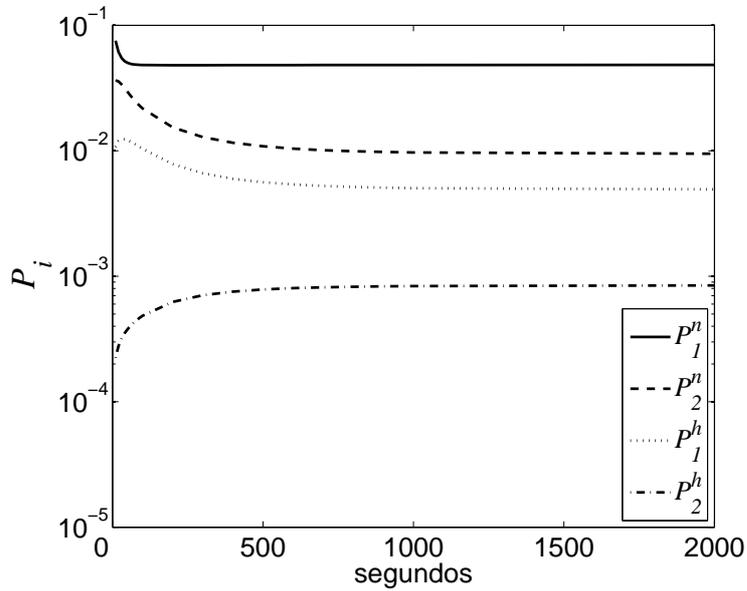


Figura 5.18: Comparativa de factores de utilización de recursos.

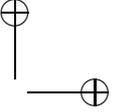
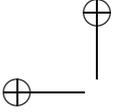


(a) Incremento súbito del 10%.



(b) Incremento súbito del 50%.

Figura 5.19: Comportamiento transitorio de las probabilidades de bloqueo.



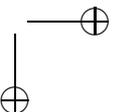
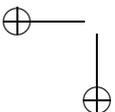
Capítulo 6

Conclusiones

—

La progresión reciente de la telefonía móvil ha dado lugar a un incremento en el volumen de tráfico que deben ser capaces de cursar las redes celulares contemporáneas. Este aumento responde al crecimiento en el número de usuarios, pero también —y cada vez más— a la diversificación que está experimentando la oferta de servicios proporcionados por las redes multiservicio. Esta tendencia acentúa la de por sí inherente escasez del recurso disponible en el medio de transmisión de la interfaz radio (el espectro radioeléctrico) y sigue motivando la adopción en estas redes de los procedimientos pertinentes para la adecuada gestión de los recursos. A este respecto, el control de admisión de sesiones se perfila como un mecanismo de importancia fundamental para la gestión de los recursos radio tanto durante el diseño previo como durante el funcionamiento propiamente dicho de las redes celulares. El control de admisión fue concebido originalmente para redes fijas y aunque muchos de sus supuestos siguen siendo válidos para entornos celulares (el tráfico no es determinista y compiten por cursar su tráfico múltiples servicios sujetos a requisitos de QoS) no es menos cierto que características esenciales de las redes celulares como la escasez y variabilidad de los recursos y sobre todo, la movilidad de los terminales, complican el análisis y síntesis de este

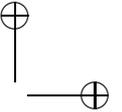
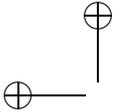
—



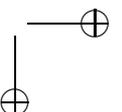
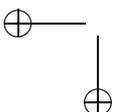
tipo de mecanismos.

En este sentido, esta tesis pretende aportar elementos de juicio significativos que contribuyan a caracterizar y elaborar mecanismos de control de admisión para la gestión de los recursos en las redes móviles celulares multiservicio actuales. Para ello en el Capítulo 3 se ha comenzado por evaluar el impacto que tiene la introducción de la movilidad en las redes fijas; tanto sobre la fase de diseño inicial, como sobre la de operación real de las redes móviles multiservicio. Ello se ha logrado extendiendo apropiadamente a sistemas móviles un método convolucional previo, del que se han tratado de mantener sus propiedades de menor coste computacional y estabilidad numérica. En el Capítulo 4 se ha investigado el diseño de políticas de control de admisión en condiciones de tráfico estacionarias y su aplicación al dimensionado de sistemas móviles multiservicio. Para ello se realiza una comparativa entre las diversas familias de políticas, tanto en términos de la capacidad que puede esperar el operador, como de la sensibilidad de las políticas a las sobrecargas de tráfico o a variaciones en los valores de los parámetros de configuración de las mismas. En dicho capítulo, asimismo se ha propuesto un algoritmo de escalada que reduce en varios órdenes de magnitud el coste computacional asociado a la síntesis de la política óptima. Finalmente, en el Capítulo 5 se trata el diseño de esquemas capaces de adaptar la configuración de las políticas de control de admisión con prestaciones más competitivas (familia *trunk reservation*) a las condiciones no estacionarias del tráfico de los sistemas móviles multiservicio. Además, se compara razonadamente el comportamiento del esquema adaptativo propuesto tanto en régimen transitorio como estacionario con el exhibido por otras contribuciones anteriores para sistemas monoservicio, concluyéndose que la propuesta aporta importantes mejoras sobre sus antecesoras en términos de la velocidad de adaptación a las nuevas condiciones de tráfico, soporte versátil del multiservicio y el estricto cumplimiento que se observa de los objetivos de QoS de las clases de mayor prioridad.

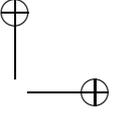
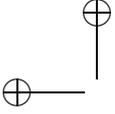
Las líneas de trabajo futuro suscitadas en el marco del desarrollo de esta tesis están relacionadas con el ubícuo panorama de operación multiservicio



y multiacceso que comienza a distinguirse para las redes móviles más allá de la tercera generación (con las redes celulares y las redes inalámbricas colaborando juntas para cursar la comunicación). Será importante evaluar cuánta información añadida es capaz de mejorar la actuación del control de admisión en estos sistemas y —sobre todo— concretar la naturaleza de dicha información (predicción de los traspasos entrantes o salientes a la célula, historiales de tráfico coordinados entre las distintas redes que se intercambian tráfico móvil, etc.). Otra cuestión interesante (que se plantea ya en las redes actuales) tiene que ver con las características que debería tener el control de admisión óptimo en sistemas en los que conviven clases de tráfico *streaming* con clases de tráfico elásticas. Asimismo resulta relevante estudiar el impacto que tiene sobre el control de admisión la redefinición de «recurso» que introducen los sistemas basados en acceso múltiple por división de código (CDMA), donde la capacidad disponible depende tanto de la interferencia que se experimenta en la célula bajo estudio como de la generada por las células cercanas. Finalmente y a modo de resumen, puesto que las redes celulares futuras serán profundamente heterogéneas es necesario vislumbrar hoy cuáles serán los mecanismos de control de admisión necesarios para la adecuada gestión de los recursos radio de los sistemas del mañana.



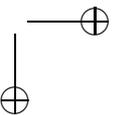
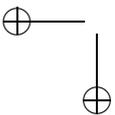
Apéndices



Apéndice A

Notación, variables y parámetros más utilizados

—	R	número de servicios diferentes	—
	C	número de unidades de recurso del sistema	
	r	servicio.	
	i	clase de peticiones.	
	c_r	número de unidades de recurso para cursar una petición	
	d_i	número de unidades de recurso para cursar una petición	
	x_r	número de sesiones en progreso de un servicio	
	$c(x)$	número de unidades de recurso ocupadas del sistema	
	a_r	tráfico ofrecido por terminal	
	M_r	número de terminales	
	f_r	penetración del servicio	
	$1/\mu_i$	tiempo medio de ocupación de los recursos por la conexión	
	$1/\mu_r^s$	tiempo medio de duración de la sesión	
	$1/\mu_r^d$	tiempo medio de residencia en una célula	
	λ_r'	intensidad de tráfico por terminal libre	
	λ_r^n	tasa de llegada de peticiones de establecimiento de nueva sesión	



Apéndice A. Notación, variables y parámetros más utilizados

λ_r^h	tasa de llegadas de peticiones de traspaso
$\lambda_{\text{máx}}$	capacidad del sistema
P_r^n	probabilidad de que una sesión nueva no sea admitida
P_r^h	probabilidad de que una sesión traspasada no sea admitida
B_r^n	cota superior objetivo para P_r^n
B_r^h	cota superior objetivo para P_r^h
l_r^n l_r^h	parámetros de configuración de la política <i>Multiple Guard Channel</i>
t_r^n t_r^h	parámetros de configuración de la política <i>Multiple Fractional Guard Channel</i>

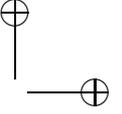
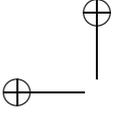
Apéndice B

Abreviaturas y acrónimos

1G 2G 3G 4G	Generaciones de telefonía móvil celular
AG	Algoritmo Genético
AMPS	<i>Advanced Mobile Service</i>
ATM	<i>Asynchronous Transfer Mode</i>
BEC	<i>Best Effort Class</i>
BPP	Binomial-Poisson-Pascal
CA	Control de Admisión
CAS	Control de Admisión de Sesiones
CC	<i>Coordinate Convex</i>
CDG	<i>CDMA Development Group</i>
CDMA	<i>Code Division Multiple Access</i>
CP	<i>Complete Partitioning</i>
CS	<i>Complete Sharing</i>
CSD	<i>Circuit Switched Data</i>
CT	Congestión de Tráfico
CTMC	<i>Continuous Time Markov Chain</i>
D-AMPS	<i>Digital AMPS</i>
DCA	<i>Dynamic Channel Allocation</i>

EDGE	<i>Enhanced Data rates for Global Evolution</i>
ETSI	<i>European Telecommunications Standards Institute</i>
FCA	<i>Fixed Channel Allocation</i>
FDD	<i>Frequency Division Duplex</i>
FDMA	<i>Frequency Division Multiple Access</i>
FGC	<i>Fractional Guard Channel</i>
FL	<i>Fractional Limit</i>
GC	<i>Guard Channel</i>
GM	<i>Guaranteed Minimum</i>
GoS	<i>Grade of Service</i>
GPRS	<i>General Packet Radio Service</i>
GSM	<i>Groupe Spécial Mobile/Global System for Mobility</i>
HCO	<i>Algoritmo de optimización de la política MFGC de [HUCPOG03a]</i>
HPC	<i>Highest Priority Class</i>
HSCSD	<i>High-Speed Circuit-Switched Data</i>
IL	<i>Integer Limit</i>
IMT-2000	<i>International Mobile Telecommunications</i>
IMT-DS	<i>IMTS Direct Spread</i>
IMT-TC	<i>IMT Time Code)</i>
IPv6	<i>Internet Protocol version 6</i>
ITU	<i>International Union of Telecommunications</i>
LMDS	<i>Local Multipoint Distribution Service</i>
LPC	<i>Lowest Priority Class</i>
MAP	<i>Markovian Arrival Process</i>
MFGC	<i>Multiple Fractional Guard Channel</i>
MGC	<i>Multiple Guard Channel</i>
MMDS	<i>Multichannel Multipoint Distribution Service</i>
MMPP	<i>Markov-Modulated Poisson Process</i>
N-AMPS	<i>Narrowband AMPS</i>
PASTA	<i>Poisson Arrivals See Time Averages</i>
PB	<i>Probabilidad de Bloqueo</i>
PCS	<i>Personal Communications Services</i>

PDC	<i>Personal Digital Cellular</i>
PF	Probabilidad de Fracaso
PFS	<i>Product Form Solution</i>
PMC	Algoritmo de optimización de la política MFGC de [PMCG05]
PP	<i>Partitioning Policy</i>
PP	Probabilidad de Pérdidas
QoS	<i>Quality of Service</i>
RDSI-BA	Red Digital de Servicios Integrados de Banda Ancha
RS	<i>Randomized Stationary</i>
SAN	<i>Stochastic Activity Networks</i>
SL	Servicio Limitante
TC	Tráfico Coursado
TDD	<i>Time Division Duplex</i>
TDMA	<i>Time Division Multiple Access</i>
TD-SCDMA	<i>Time Division Synchronized CDMA</i>
TO	Tráfico Ofrecido
TP	<i>Threshold Policy</i>
TR	<i>Trunk Reservation</i>
UL	<i>Upper Limit</i>
VP	<i>Virtual Partitioning</i>
W-CDMA	<i>Wideband CDMA</i>
WLAN	<i>Wireless Local Area Network</i>
WZZZ	Esquema adaptativo para la política GC propuesto en [WZZZ03]
ZL	Esquema adaptativo para la política GC propuesto en [ZL01]



Apéndice C

Publicaciones

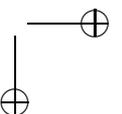
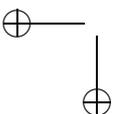
C.1 Relacionadas con la tesis

C.1.1 Capítulo de libro

1. Jorge Martínez-Bauset, Vicent Pla, David García-Roger, M.^aJose Domenech-Benloch, and Jose Manuel Gimenez-Guzman. *Designing admission control policies to minimize blocking/forced-termination*. In *Modeling and Simulation of Wireless Networks, Analysis, Evaluation and Enhancement of QoS for Wireless Multimedia*, Nova Science Publishers. Accepted for publication in July 2006.

C.1.2 Revista

1. David García and Jorge Martínez. *Gestión de recursos en redes 3G: uso de programación evolutiva para la evaluación comparada de políticas*. *Revista Gerencia de Tecnología Informática (GTI)*, 2(5):26–33, May 2004.



2. David García, Jorge Martínez, and Vicent Pla. *Admission control policies in multiservice cellular networks: Optimum configuration and sensitivity*. "Mobile and Wireless Systems", Lecture Note in Computer Science, **3427**:121–135, Springer-Verlag, Berlin, Feb. 2005.
3. David Garcia-Roger, M.^aJose Domenech-Benlloch, Jorge Martinez-Bauset, and Vicent Pla. *Hierarchical admission control in mobile cellular networks using adaptive bandwidth reservation*. "Wireless and Mobility", Lecture Note in Computer Science, **3883**:130–144, Springer-Verlag, Berlin, May 2006.
4. David Garcia-Roger, M.^aJose Domenech-Benlloch, Jorge Martinez-Bauset, and Vicent Pla. *Esquema adaptativo de reserva para el control de admisión jerárquico en redes móviles celulares*. Revista del IEEE América Latina, Dec. 2006. Accepted for publication.

C.1.3 Congreso

Internacional

1. Jorge Martínez and David García. *Impact of link failures in wireless multi-service access networks with tree topology*. In Proceedings of the 8th European Conference on Networks & Optical Communications (NOC), pp. 221–227, Jul. 2003.
2. David García, Jorge Martínez and Vicent Pla. *Comparative evaluation of admission control policies in cellular multiservice networks*. In Proceedings of the The 16th International Conference on Wireless Communications (Wireless), pp. 517–531, July 2004.
3. David Garcia, Jorge Martínez and Vicent Pla. *Optimality and sensitivity study of admission control policies for multimedia wireless networks*. In Proceedings of the IEEE 6th Workshop on Multimedia Signal Processing Conference (MMSP), pp. 458–461, Sep. 2004.

4. David Garcia-Roger, M.^aJose Domenech-Benlloch, Jorge Martinez-Bauset, and Vicent Pla. *Adaptive admission control scheme for multiservice mobile cellular networks*. In Proceedings of the EuroNGI First Conference on Next Generation Internet: Traffic Engineering (NGI), pp. 288–295, Apr. 2005.
5. David Garcia-Roger, M.^aJose Domenech-Benlloch, Jorge Martinez-Bauset, and Vicent Pla. *Comparative evaluation of adaptive trunk reservation schemes for mobile cellular networks*. In Proceedings of the 3rd International Working Conference on Performance Modelling and Evaluation of Heterogeneous Networks (HET-NETs), pp. 25/1–10, July 2005.
6. David Garcia-Roger, Jorge Martinez-Bauset, and Vicent Pla. *Efficient algorithm to determine the optimal configuration of admission control policies in multiservice mobile wireless networks*. In Proceedings of the 3rd International Working Conference on Performance Modelling and Evaluation of Heterogeneous Networks (HET-NETs), pp. 32/1–9, July 2005.
7. David Garcia-Roger, M.^aJose Domenech-Benlloch, Jorge Martinez-Bauset, and Vicent Pla. *Adaptive trunk reservation policies in multiservice mobile wireless networks*. In Proceedings of the 8th International Conference on Management of Multimedia Networks and Services, Lecture Note in Computer Science, **3754**:47–58, Springer-Verlag, Berlin, Oct. 2005.
8. David Garcia-Roger, M.^aJose Domenech-Benlloch, Jorge Martinez-Bauset, and Vicent Pla. *Adaptive admission control in mobile cellular networks with streaming and elastic traffic*. In Proceedings of the 20th International Teletraffic Congress. Accepted on Feb. 2007.

Nacional

1. David García, Jorge Martínez, y Vicent Pla. *Optimización y sensibilidad frente a sobrecargas de políticas de control de admisión para redes celulares multiservicio*. En Actas de las XIV Jornadas Telecom I+D, noviembre 2004.

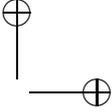
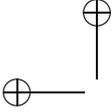
2. David García Roger, M.^a José Doménech Benlloch, Jorge Martínez Bauset, y Vicent Pla. *Esquema adaptativo de reserva para redes móviles celulares*. En Actas de las V Jornadas de Ingeniería Telemática (JITEL), pp. 25–32, Sep. 2005.
3. David García Roger, M.^a José Doménech Benlloch, Jorge Martínez Bauset, y Vicent Pla. *Esquema adaptativo para el control de admisión en redes celulares multiservicio*. En Actas de las XV Jornadas Telecom I+D, noviembre 2005.

C.2 Otras publicaciones

C.2.1 Congreso

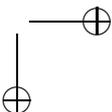
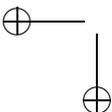
Internacional

1. David García and Luis Guijarro. *A comparative analysis of e-Business reference models*. In Proceedings of the IEEE Youth Forum in Computer Science and Engineering (YUFORIC), pp. 25–32, Nov. 2001.



Bibliografía

- [Aei78] Joseph M. Aein, *A multi-user-class, blocked-calls-cleared, demand access model*, IEEE Transactions on Communications **26** (1978), no. 3, 378–385.
- [Ahm05] Mohamed H. Ahmed, *Call admission control in wireless networks: a comprehensive survey*, IEEE Communications Surveys & Tutorials **7** (2005), no. 1, 49–68.
- [AJK01] Eitan Altman, Tania Jimenez, y Ger Koole, *On optimal call admission control in a resource-sharing system*, IEEE Transactions on Communications **49** (2001), no. 9, 1659–1668.
- [AL02] Attahiru S. Alfa y Wei Li, *PCS networks with correlated arrival process and retrial phenomenon*, IEEE Transactions on Wireless **1** (2002), no. 4, 630–637.
- [AMDCL⁺99] Marco Ajmone Marsan, Giovanni De Carolis, Emilio Leonardi, Renato Lo Cigno, y Michela Meo, *How many cells should be considered to accurately predict the performance of cellular networks?*, Proceedings of the European Wireless'99 and 4th ITG Mobile Communications (Munich, Germany), octubre 1999.
- [Bar01] Novella Bartolini, *Handoff and optimal channel assignment in wireless networks*, Mobile Networks and Applications (MONET) **6** (2001), no. 6, 511–524.
- [BC02] Novella Bartolini y Imrich Chlamtac, *Call admission control in wireless multimedia networks*, Proceedings of the 13th IEEE International Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC) (Lisboa, Portugal), vol. 1, septiembre 2002, pp. 285–289.
- 
- 



- [BF01] Cory C. Beard y Victor S. Frost, *Prioritized resource allocation for stressed networks*, IEEE/ACM Transactions on Networking **9** (2001), no. 5, 618–633.
- [BJ00] Francisco Barceló y Javier Jordán, *Channel holding time distribution in public telephony systems (PAMR and PCS)*, IEEE Transactions on Vehicular Technology **49** (2000), no. 5, 1615–1625.
- [BM98] Sem C. Borst y Debasis Mitra, *Virtual partitioning for robust resource sharing: computational techniques for heterogeneous traffic*, IEEE Journal on Selected Areas in Communications **16** (1998), no. 5, 668 – 678.
- [BS97] Subir K. Biswas y Bhaskar Sengupta, *Call admissibility for multirate traffic in wireless ATM networks*, Proceedings of IEEE INFOCOM, vol. 2, 1997, pp. 649–657.
- [BS99] Francisco Barcelo y J. I. Sanchez, *Probability distribution of the inter-arrival time to cellular telephony channels*, Proceedings of IEEE 49th Vehicular Technology Conference, 1999, pp. 762–766.
- [CB00] Ming-Hsing Chiu y Mostafa A. Bassiouni, *Predictive schemes for handoff prioritization in cellular networks based on mobile positioning*, IEEE Journal on Selected Areas in Communications **18** (2000), no. 3, 510–522.
- [CC97] Chi-chao Chao y Wai Chen, *Connection admission control for mobile multiple-class personal communications networks*, IEEE Journal on Selected Areas in Communications **15** (1997), no. 8, 1618–1626.
- [CG01] Vicente Casares-Giner, *Integration of dispatch and interconnect traffic in a land mobile trunking system. waiting time distributions*, Telecommunication Systems **16** (2001), no. 3,4, 539–554, Previously presented at 4th INFORMS Telecommunications Conference, Boca Ratón (Florida) March 8–11, 1998.
- [CH96] Vicente Casares y Jack M. Holtzman, *Dispatch versus interconnect traffic. A comparative analysis in a land mobile trunking*

- system, Proceedings of the 46th VTC, abril 1996, pp. 242–246, see for more details WINLAB TR-118, Rutgers University, NJ, May 1996.
- [CL95] Edward Chlebus y Wieslaw Ludwin, *Is handoff traffic really poissonian?*, Proceedings of ICUPC'95, 1995, pp. 348–353.
- [CLW95] Gagan L. Choudhury, Kin K. Leung, y Ward Whitt, *Efficiently providing multiple grades of service with protection against overloads in shared resources*, AT&T Technical Journal **74** (1995), no. 4, 50–63.
- [Coo81] Robert B. Cooper, *Introduction to queuing theory: Second edition*, North Holland (Elsevier), 1981.
- [CPVAOG04] Felipe A. Cruz-Perez, Jose L. Vazquez-Avila, y Lauro Ortigoza-Guerrero, *Recurrent formulas for the multiple fractional channel reservation strategy in multi-service mobile cellular networks*, IEEE Communications Letters **8** (2004), no. 10, 629–631.
- [CS02] Sunghyun Choi y Kang G. Shin, *Adaptive bandwidth reservation and admission control in QoS-sensitive cellular networks*, IEEE Transactions on Parallel and Distributed Systems **13** (2002), no. 9, 882–897.
- [CSC94] Chung-Ju Chang, Tian-Tsair Su, y Yueh-Yiing Chiang, *Analysis of a cutoff priority cellular radio system with finite queueing and renegingdropping*, IEEE/ACM Transactions on Networking **2** (1994), no. 2, 166–175.
- [DBGGMBCG05] M.^a José Doménech-Benlloch, José Manuel Giménez-Guzmán, Jorge Martínez-Bauset, y Vicente Casares-Giner, *Efficient and accurate methodology for solving multiserver retrial systems*, IEE Electronics Letters **41** (2005), no. 17, 967–969.
- [DJ92] John N. Daigle y Nikhil Jain, *A queueing system with two arrival streams and reserved servers with application to cellular telephone*, Proceedings of INFOCOM'92, IEEE, mayo 1992, pp. 9C.2.1–9.C.2.7.

- [DRFG99a] Enrico Del Re, Romano Fantacci, y Giovanni Giambene, *Different queuing policies for handover request in low earth orbit mobile satellite systems*, IEEE Transactions on Vehicular Technology **48** (1999), no. 2, 448–458.
- [DRFG99b] ———, *Handover queuing strategies with dynamic and fixed channel allocation techniques in low earth orbit mobile satellite systems*, IEEE Transactions on Communications **47** (1999), no. 1, 89–102.
- [DS02] Daryl J. Daley y Leslie D. Servi, *Loss probabilities of hand-in traffic under various protocols. I. models and algebraic results*, Telecommunication Systems **19** (2002), no. 2, 209–226.
- [DS04] ———, *Loss probabilities of hand-in traffic under various protocols: II. model comparisons*, Performance Evaluation **55** (2004), no. 3-4, 231–249.
- [dSeSG00] Edmundo de Souza e Silva y H. Richard Gail, *Transient solutions for markov chains*, p. 43–81, Kluwer Academic Publishers, 2000.
- [DTL03] Selvamuthu Dharmaraja, Kishor S. Trivedi, y Dimitris Logothetis, *Performance modeling of wireless networks with generally distributed handoff interarrival times*, Computer Communications **26** (2003), 1747–1755.
- [EAYH01a] El-Sayed M. El-Alfy, Yu-Dong Yao, y Harry Heffes, *Adaptive resource allocation with prioritized handoff in cellular mobile networks under QoS provisioning*, Proceedings of IEEE the 54th Vehicular Technology Conference (VTC Fall), 2001, pp. 2113–2117.
- [EAYH01b] ———, *Autonomous call admission control with prioritized handoff in cellular networks*, Proceedings of IEEE ICC, 2001, pp. 1386 – 1390.
- [EAYH01c] ———, *A learning approach for call admission control with prioritized handoff in mobile multimedia networks*, Proceedings of IEEE the 53rd Vehicular Technology Conference (VTC Spring), 2001, pp. 972 – 976.

- [EAYH01d] ———, *A model-based q-learning scheme for wireless channel allocation with prioritized handoff*, Proceedings of IEEE GLOBECOM, 2001, pp. 3668 – 3672.
- [EE99] Jamie S. Evans y David Everitt, *Effective bandwidth-based admission control for multiservice CDMA cellular networks*, IEEE Transactions on Vehicular Technology **48** (1999), no. 1, 36–46.
- [EM93] Anwar I. Elwalid y Debasis Mitra, *Effective bandwidth of general Markovian traffic sources and admission control of high speed networks*, IEEE/ACM Transactions on Networking **1** (1993), no. 3, 329–343.
- [ET99] Howard G. Ebersman y Ozan K. Tonguz, *Handoff ordering using signal prediction priority queuing in personal communication systems*, IEEE Transactions on Vehicular Technology **48** (1999), no. 1, 20–35.
- [Fan00] Romano Fantacci, *Performance evaluation of prioritized handoff schemes in mobile cellular networks*, IEEE Transactions on Vehicular Technology **49** (2000), no. 2, 485–493.
- [GAN91] Roch Guerin, Hamid Ahmadi, y Mahmoud Naghshineh, *Equivalent capacity and its application to bandwidth allocation in high-speed networks*, IEEE Journal on Selected Areas in Communications **9** (1991), no. 7, 968–981.
- [GB06] Majid Ghaderi y Raouf Boutaba, *Call admission control in mobile cellular networks: a comprehensive survey*, Wireless Communications and Mobile Computing **6** (2006), no. 1, 69–93.
- [GJ03] Eva Gustafsson y Annika Jonsson, *Always best connected*, IEEE Wireless Communications **10** (2003), 49–55.
- [GMP04] David García, Jorge Martínez, y Vicent Pla, *Comparative evaluation of admission control policies in cellular multiservice networks*, Proceedings of the The 16th International Conference on Wireless Communications (Wireless) (Calgary, Canada), julio 2004, pp. 517–531.

- [GMP05] ———, *Admission control policies in multiservice cellular networks: Optimum configuration and sensitivity*, Lecture Notes in Computer Science **3427** (2005), 121–135.
- [GRDBMBP05] David Garcia-Roger, M.^a Jose Domenech-Benlloch, Jorge Martinez-Bauset, y Vicent Pla, *Adaptive admission control scheme for multiservice mobile cellular networks*, Proceedings of the 1st EuroNGI Conference on Next Generation Internet Networks Traffic Engineering (NGI) (Roma, Italy), abril 2005, pp. 288–295.
- [Gué87] Roch A. Guérin, *Channel occupancy time distribution in a cellular radio system*, IEEE Transactions on Vehicular Technology **35** (1987), 89–99.
- [Gué88] ———, *Queueing-blocking system with two arrival streams and guard channel*, IEEE Transactions on Communications **36** (1988), no. 2, 153–163.
- [HBBPGMIA00] Alberto Heras-Brandín, Purificación Bartolomé-Pascual, Daniel Gómez-Mateo, y Jesus Izquierdo-Arce, *A multiservice dimensioning procedure for 3G CDMA*, Proceedings of the First International Conference on 3G Mobile Communication Technologies (IEE 3G) (London, UK), marzo 2000, pp. 406–410.
- [HKK04] Lei Huang, Sunil Kumar, y C.-C. Jay Kuo, *Adaptive resource allocation for multimedia QoS management in wireless networks*, IEEE Transactions on Vehicular Technology **53** (2004), no. 2, 547–558.
- [HR86] Daehyong Hong y Stephen S. Rappaport, *Traffic model and performance analysis for cellular mobile radio telephone systems with prioritized and nonprioritized handoff procedures*, IEEE Transactions on Vehicular Technology **VT-35** (1986), no. 3, 77–92.
- [HSSK01] Hirotoshi Hidaka, Kazuyoshi Saitoh, Noriteru Shinagawa, y Takehiko Kobayashi, *Teletraffic characteristics of cellular communication for different types of vehicle motion*, IEICE Transactions on Communications **E84-B** (2001), no. 3, 558–565.

- [HSSK02] ———, *Self-similarity in cell dwell time caused by terminal motion and its effects on teletraffic of cellular communication networks*, IEICE Transactions on Fundamentals of Electronics, Communications and Computer Sciences **E85-A** (2002), no. 7, 1445–1453.
- [HUCPOG03a] Heraclio Heredia-Ureta, Felipe A. Cruz-Pérez, y Lauro Ortigoza-Guerrero, *Capacity optimization in multiservice mobile wireless networks with multiple fractional channel reservation*, IEEE Transactions on Vehicular Technology **52** (2003), no. 6, 1519 – 1539.
- [HUCPOG03b] ———, *Multiple fractional channel reservation for multi-service cellular networks*, Proceedings of IEEE ICC, 2003, pp. 964 – 968.
- [HUCPOG03c] ———, *Multiple fractional channel reservation for optimum system capacity in multi-service cellular networks*, IEE Electronics Letters **39** (2003), no. 1, 133–134.
- [IS97] Villy B. Iversen y Sergey N. Stepanov, *The usage of convolution algorithm with truncation for estimation of individual blocking probabilities in circuit-switched telecommunication networks*, Proceedings of the 15th International Teletraffic Congress, (ITC 15) (Washington, DC, USA), vol. 2, junio 1997, pp. 1327–1336.
- [Ive87] Villy B. Iversen, *The exact evaluation of multi-service loss systems with access control*, Proceedings of the Teleteknik and Seventh Nordic Teletraffic Seminar (NTS-7) (Lund, (Sweden)), vol. 31, agosto 1987, pp. 56–61.
- [Ive02] ———, *Teletraffic engineering handbook*, ITU-D SG 2 and ITC, 2002, <http://www.tele.dtu.dk/teletraffic/>.
- [Jab96] Bijan Jabbari, *Teletraffic aspects of evolving and next-generation networks*, IEEE Personal Communications **3** (1996), no. 6, 4–9.
- [JJ01] Wha Sook Jeon y Dong Guen Jeong, *Call admission control for mobile multimedia communications with traffic asymmetry between uplink and downlink*, IEEE Transactions on Vehicular Technology **50** (2001), no. 1, 59–66.

- [JL96] Chris Jedrzycki y Victor C. M. Leung, *Probability distributions of channel holding time in cellular telephony systems*, Proceedings of VTC'96, mayo 1996, pp. 247–251.
- [JV94] Scott Jordan y Pravin P. Varaiya, *Control of multiple service, multiple resource communication networks*, IEEE Transactions on Communications **42** (1994), no. 11, 2979 – 2988.
- [KA99] Mohana D. Kulavaratharajah y A. Hamid Aghvami, *Tele-traffic performance evaluation of microcellular personal communication networks (PCN's) with prioritized handoff procedures*, IEEE Transactions on Vehicular Technology **48** (1999), no. 1, 137–152.
- [Kau81] Joseph M. Kaufman, *Blocking in a shared resource environment*, IEEE Transactions on Communications **29** (1981), no. 10, 1474–1481.
- [KFZK04] Insoo Koo, Anders Furuskar, Jens Zander, y Kiseon Kim, *Erlang capacity of multiaccess systems with service-based access selection*, IEEE Communication Letters **8** (2004), no. 11, 662–664.
- [KN96] Irene Katzela y Mahmoud Naghshineh, *Channel assignment schemes for cellular mobile telecommunication systems: A comprehensive survey*, IEEE Personal Communications (1996), 10–31.
- [KZ97] Farooq Khan y Djamal Zeghlache, *Effect of cell residence time distribution on the performance of cellular mobile networks*, Proceedings of VTC'97, IEEE, 1997, pp. 949 – 953.
- [LA95] Chin-Tau Lea y Anwar Alyatama, *Bandwidth quantization and states reduction in the broadband ISDN*, IEEE/ACM Transactions on Networking **3** (1995), no. 3, 352–360.
- [LAN97] David A. Levine, Ian F. Akyildiz, y Mahmoud Naghshineh, *A resource estimation and call admission algorithm for wireless multimedia networks using the shadow cluster concept*, IEEE/ACM Transactions on Networking **5** (1997), no. 1, 1–12.

- [LC97] Yi-Bing Lin y Imrich Chlamtac, *Effects of Erlang call holding times on PCS call completion*, Proceedings of INFOCOM'97, IEEE, 1997.
- [Lee91] W. C. Y. Lee, *Smaller cells for greater performance*, IEEE Communications Magazine **29** (1991), no. 11, 19–23.
- [LJCP03] Wei Kuang Lai, Yu-Jyr Jin, Hsin Wei Chen, y Chieh Ying Pan, *Channel assignment for initial and handoff calls to improve the call-completion probability*, IEEE Transactions on Vehicular Technology **52** (2003), no. 4, 876–890.
- [LLC98] Bo Li, Chuang Lin, y Samuel T. Chanson, *Analysis of a hybrid cutoff priority scheme for multiple classes of traffic in multimedia wireless networks*, Wireless Networks Journal (WINET) **4** (1998), no. 4, 279–290.
- [LMN94] Yi-Bing Lin, Seshadri Mohan, y Anthony Noerpel, *Queueing priority channel assignment strategies for PCS handoff and initial access*, IEEE Transactions on Vehicular Technology **43** (1994), no. 3, 704–712.
- [McM91] David McMillan, *Traffic modelling and analysis for cellular mobile networks*, Proceedings of ITC-13 on Teletraffic and Data traffic in a Period of Change (Copenhague) (A. Jensen y V.B. Iversen, eds.), IAC, Elsevier Science, junio 1991, pp. 627–632.
- [McM95] ———, *Delay analysis of a cellular mobile priority queuing system*, IEEE/ACM Transactions on Networking **3** (1995), no. 3, 310–319.
- [MDCL⁺01] Marco Ajmone Marsan, Giovanni M. De Carolis, Emilio Leonardi, Renato Lo Cigno, y Michela Meo, *Efficient estimation of call blocking probabilities in cellular mobile telephony networks with customer retrials*, IEEE Journal on Selected Areas in Communications **19** (2001), no. 2, 332–346.
- [MHY05] Yufeng Ma, Xiulin Hu, y Shi Yimei, *Intelligent call admission control using fuzzy logic in wireless networks*, Proceedings of the American Control Conference (ACC) (Portland, USA), junio 2005, pp. 3981–3985.

- [Mic94] Zbigniew Michalewicz, *Genetic algorithms + data structures = evolution programs*, Springer-Verlag, 1994.
- [MK00] Werner Mohr y Walter Konhäuser, *Access network evolution beyond third generation mobile communications*, IEEE Communications Magazine **38** (2000), 122–133.
- [MMS85] John F. Meyer, Ali Movaghar, y William H. Sanders, *Stochastic activity networks: Structure, behavior, and application*, Proc. Int. W. Timed Petri Nets (Torino, Italy), julio 1985, pp. 106–115.
- [MS01] Kenneth Mitchell y Khosrow Sohraby, *An analysis of the effects of mobility on bandwidth allocation strategies in multi-class cellular wireless networks*, Proceedings of IEEE INFOCOM, vol. 2, 2001, pp. 1075–1084.
- [Nel95] Randolph Nelson, *Probability, stochastic processes and queueing theory*, Springer-Verlag, 1995.
- [NHA05] Dusit Niyato, Ekram Hossain, y Attahiru S. Alfa, *Performance analysis and adaptive call admission control in cellular mobile networks with time-varying traffic*, Proceedings of the IEEE International Conference on Communications (ICC) (Seoul, Korea), vol. 5, mayo 2005, pp. 3324 – 3328.
- [NS96] Mahmoud Naghshineh y Mischa Schwartz, *Distributed call admission control in mobile wireless networks*, IEEE Journal on Selected Areas in Communications **14** (1996), no. 4, 711–717.
- [OR01] Philip V. Orlik y Stephen S. Rappaport, *On the handoff arrival process in cellular communications*, Wireless Networks Journal (WINET) **7** (2001), no. 2, 147–157.
- [PC01] Vicent Pla y Vicente Casares, *Delay-loss analysis of channel assignment schemes in mobile cellular with handoff priority and hysteresis control*, Proceedings of 14th ITC Specialist Seminar on Access Networks and Systems, 2001, pp. 221–230.
- [PCG02] Vicent Pla y Vicente Casares-Giner, *Effect of the handoff area sojourn time distribution on the performance of cellular networks*, Proceedings of IEEE MWCN, septiembre 2002, pp. 401–405.

- [PCG03] ———, *Optimal admission control policies in multiservice cellular networks*, Proceedings of the International Network Optimization Conference (INOC 2003) (Paris, France), octubre 2003, pp. 466–471.
- [PCG04] ———, *Análisis y diseño de políticas de control de admisión en redes celulares multiservicio*, Revista IEEE América Latina 2 (2004), no. 1, 10–19.
- [Per06] Performability Engineering Research Group (PERFORM), *MöbiusTM. model-based environment for validation of system reliability, availability, security, and performance. user manual*, 2006.
- [PG85] Edward C. Posner y Roch Guérin, *Traffic policies in cellular radio that minimize blocking of handoff calls*, Proceedings of ITC 11, 1985.
- [PGGMCG04] Vicent Pla, José Manuel Giménez-Guzmán, Jorge Martínez, y Vicente Casares-Giner, *Optimal admission control using handover prediction in mobile cellular networks*, Proceedings of the 2nd International Working Conference on Performance Modelling and Evaluation of Heterogeneous Networks (HET-NETs'04), julio 2004, pp. 44/1–10.
- [PMCG05] Vicent Pla, Jorge Martínez, y Vicente Casares-Giner, *Algorithmic computation of optimal capacity in multiservice mobile wireless networks*, IEICE Transactions on Communications E88-B (2005), no. 2, 797–799.
- [QLdM04] Claudia Quevedo-Lodi y J. Roberto B. de Marca, *Performance of a dynamic multiple class admission control strategy for wireless systems*, Proceedings of the IEEE 60th Vehicular Technology Conference (VTC-Fall) (Los Angeles, USA), septiembre 2004, pp. 1083–1087.
- [RK03] Adam Rosenberg y Sid Kemp, *CDMA capacity and quality optimization*, McGraw-Hill, 2003.
- [Ros70] Sheldon M. Ross, *Applied probability models with optimization applications*, Holden-Day, 1970.
- [Ros95] Keith W. Ross, *Multiservice loss models for broadband telecommunication networks*, Springer Verlag, 1995.

- [RSAK99] Parameswaran Ramanathan, Krishna M. Sivalingam, Prathima Agrawal, y Shaline Kishore, *Dynamic resource allocation schemes during handoff for mobile multimedia wireless networks*, IEEE Journal on Selected Areas in Communications **17** (1999), no. 7, 1270–1283.
- [RT89] Keith W. Ross y Danny H. K. Tsang, *The stochastic knapsack problem*, IEEE Transactions on Communications **37** (1989), no. 7, 740–747.
- [RT00] Myuran Rajaratnam y Fambirai Takawira, *Nonclassical traffic modeling and performance analysis of cellular mobile networks with and without channel reservation*, IEEE Transactions on Vehicular Technology **49** (2000), no. 3, 817–834.
- [RT01] ———, *Handoff traffic characterization in cellular networks under nonclassical arrivals and service time distributions*, IEEE Transactions on Vehicular Technology **50** (2001), no. 4, 954–970.
- [RTN97] Ramachandran Ramjee, Don Towsley, y Ramesh Nagarajan, *On optimal call admission control in cellular networks*, Wireless Networks Journal (WINET) **3** (1997), no. 1, 29–41.
- [Sal04] Apostolis K. Salkintzis, *Interworking techniques and architectures for WLAN/3G integration toward 4G mobile data networks*, IEEE Wireless Communications **11** (2004), 50–61.
- [Sat94] Hiroshi Sato, *Teletraffic technologies in ATM networks*, Artech House Publishers, 1994.
- [SB98] Richard S. Sutton y Andrew G. Barto, *Reinforcement learning*, The MIT press, Cambridge, Massachusetts, 1998, <http://www.cs.ualberta.ca/~sutton/book/ebook/the-book.html>.
- [Sch90] Rudolf Schehrer, *On a cut-off priority queueing system with hysteresis and unlimited waiting room*, Computer Networks and ISDN Systems **20** (1990), 45–56.
- [Sch03] Mathias Schweigel, *The cell residence time in rectangular cells and its exponential approximation*, Proceedings of ITC 18, 2003.

- [SHNK00] Mahmoud R. Sherif, Ibrahim W. Habib, Mahmoud Nags-hineh, y Parviz Kermani, *Adaptive allocation of resources and call admission control for wireless ATM using genetic algo-rithms*, IEEE Journal on Selected Areas in Communications **18** (2000), no. 2, 106–117.
- [SK04] Wee-Seng Soh y Hyong S. Kim, *Dynamic bandwidth reser-vation in cellular networks using road topology based mobility prediction*, Proceedings of the Twenty-third Annual Joint Conference of the IEEE Computer and Communications Societies (INFOCOM) (Hong Kong, China), vol. 4, marzo 2004, pp. 2766–2777.
- [SS96] Moshe Sidi y David Starobinski, *New call blocking versus hand-off blocking in cellular networks*, Proceedings of the Fif-teenth Annual Joint Conference of the IEEE Computer So-cieties. (INFOCOM) (San Francisco, CA, USA), vol. 1, mar-zo 1996, pp. 35–42.
- [SS97] ———, *New call blocking versus handoff blocking in cellu-lar networks*, Wireless Networks Journal (WINET) **3** (1997), no. 1, 15–27.
- [SW04] Haw-Yun Shin y Jean-Lien C. Wu, *The study of dynamic multi-channel scheme with channel de-allocation in wireless net-works*, Computer Networks **45** (2004), 463–482.
- [TGM97] Phuoc Tran-Gia y Michel Mandjes, *Modeling of customer re-trial phenomenon*, IEEE Journal on Selected Areas in Com-munications **15** (1997), no. 8, 1406–1414.
- [TJ91] Sirin Tekinay y Bijan Jabbari, *Handover and channel assign-ment in mobile cellular networks*, IEEE Communications Ma-gazine (1991), 42–46.
- [TJ92] ———, *A measurement-based prioritization scheme for hando-vers in mobile cellular networks*, IEEE Journal on Selected Areas in Communications **10** (1992), no. 8, 1343–1350.
- [TRV98] Nishith D. Tripathi, Jeffrey H. Reed, y Hugh F. VanLan-dingham, *Handoff in cellular systems*, IEEE Personal Com-munications (1998), 26–37.

- [VC00] András G. Valkó y Andrew T. Campbell, *An efficiency limit of cellular mobile systems*, *Computer Communications* **23** (2000), no. 5-6, 441–451.
- [vDT03] Erik A. van Doorn y An T. K. Ta, *Proofs for some conjectures of Rajaratnam and Takawira on the peakedness of handoff traffic*, *IEEE Transactions on Vehicular Technology* **52** (2003), no. 4, 953–957.
- [VLLX02] Johan De Vriendt, Philippe Lainé, Christophe Lerouge, y Xiaofeng Xu, *Mobile network evolution: A revolution on the move*, *IEEE Communications Magazine* **40** (2002), 104–111.
- [Wes02] Krzysztof Wesołowski, *Mobile communication systems*, John Wiley & Sons, 2002.
- [WLR⁺04] Yaya Wei, Chuang Lina, Fengyuan Rena, Raad Raad, y Eryk Dutkiewicz, *Dynamic handoff scheme in differentiated QoS wireless multimedia networks*, *Computer Communications* **27** (2004), no. 10, 1001–1011.
- [WWL02] Si Wu, K.Y. Michael Wong, y Bo Li, *A dynamic call admission policy with precision QoS guarantee using stochastic control for mobile wireless networks*, *IEEE/ACM Transactions on Networking* **10** (2002), no. 2, 257–271.
- [WZA03] Jingao Wang, Qing-An Zeng, y Dharma P. Agrawal, *Performance analysis of a preemptive and priority reservation handoff scheme for integrated service-based wireless mobile networks*, *IEEE Transactions on Mobile Computing* **2** (2003), no. 1, 65–75.
- [WZZZ03] Xue-Ping Wang, Jian-Li Zheng, Wei Zeng, y Gen-Du Zhang, *A probability-based adaptive algorithm for call admission control in wireless network*, *Proceedings of the International Conference on Computer Networks and Mobile Computing (ICCNMC'03)* (Shanghai, China), octubre 2003, pp. 197–204.
- [XCW00] Yang Xiao, Philip Chen, y Yan Wang, *A near optimal call admission control with genetic algorithm for multimedia services in wireless/mobile networks*, *Proceedings of the National Aerospace and Electronics Conference (NAECON)*, 2000, pp. 787–792.

- [XCW01] ———, *Optimal admission control for multi-class of wireless adaptive multimedia services*, IEICE Transactions on Communications **E84-B** (2001), no. 4, 795–804.
- [XT03] Ariton E. Khafa y Ozan K. Tonguz, *Does mixed lognormal channel holding time affect the handover performance of guard channel scheme ?*, Proceedings of IEEE GLOBECOM, 2003, pp. 3452–3456.
- [XT04] ———, *Dynamic priority queueing of handover calls in wireless networks: An analytical framework*, IEEE Journal on Selected Areas in Communications **22** (2004), no. 5, 904 – 916.
- [YK02] Oliver Yu y Shashank Khanvilkar, *Dynamic adaptive QoS provisioning over GPRS wireless mobile links*, Proceedings of the IEEE International Conference on Communications (ICC 2002) (New York, NY, (USA)), vol. 2, abril 2002, pp. 1100–1104.
- [YL97] Oliver T. W. Yu y Victor C. M. Leung, *Adaptive resource allocation for prioritized call admission over an ATM-based wireless PCN*, IEEE Journal on Selected Areas in Communications **15** (1997), no. 7, 1208–1225.
- [YL02] Fei Yu y Victor Leung, *Mobility-based predictive call admission control and bandwidth reservation in wireless cellular networks*, Computer Networks **38** (2002), no. 5, 577–589.
- [YMW⁺04] Jianxin Yao, Jon W. Mark, Tung Chong Wong, Yong Huat Chew, Kin Mun Lye, y Kee-Chaing Chua, *Virtual partitioning resource allocation for multiclass traffic in cellular systems with QoS constraints*, IEEE Transactions on Vehicular Technology **53** (2004), no. 3, 847 – 864.
- [YR97] Aylin Yener y Christopher Rose, *Genetic algorithms applied to cellular call admission: local policies*, IEEE Transactions on Vehicular Technology **46** (1997), no. 1, 72–79.
- [ZA02] Qing-An Zeng y Dharma P. Agrawal, *Modeling and efficient handling of handoffs in integrated wireless mobile networks*, IEEE Transactions on Vehicular Technology **51** (2002), no. 6, 1469 – 1478.

- [ZC99] Helen Zeng y Imrich Chlamtac, *Handoff traffic distribution in cellular networks*, Proceedings of IEEE Wireless Communications and Networking Conference (WCNC), 1999, pp. 413–417.
- [ZD97] Mahmood M. Zonoozi y Prem Dassanayake, *User mobility modeling and characterization of mobility patterns*, IEEE Journal on Selected Areas in Communications **15** (1997), no. 7, 1239–1252.
- [ZJB00] Mao Zeng, Annamalai Annamalai Jr., y Vijay K. Bhargava, *A harmonization of global third-generation mobile systems*, IEEE Communications Magazine **38** (2000), 94–104.
- [ZL01] Yi Zhang y Derong Liu, *An adaptive algorithm for call admission control in wireless networks*, Proceedings of the IEEE Global Communications Conference (GLOBECOM'01) (San Antonio, TX, (USA)), noviembre 2001, pp. 3628–3632.