

UNIVERSIDAD DE GRANADA

DEPARTAMENTO DE

TEORÍA DE LA SEÑAL, TELEMÁTICA Y COMUNICACIONES



**MECANISMOS DE MEJORA DE CALIDAD DE
EXPERIENCIA SOBRE REDES 802.11**

JORGE NAVARRO ORTIZ

TESIS DOCTORAL

Granada, España. Julio, 2010

Editor: Editorial de la Universidad de Granada
Autor: Jorge Navarro Ortiz
D.L.: GR 3441-2010
ISBN: 978-84-693-5188-8

D. Juan Manuel López Soler

Profesor Titular de Ingeniería Telemática del
Departamento de Teoría de la Señal, Telemática y Comunicaciones de la
Universidad de Granada

CERTIFICA:

Que la presente memoria, titulada *Mecanismos de Mejora de Calidad de Experiencia sobre Redes 802.11*, ha sido realizada por D. Jorge Navarro Ortiz bajo su dirección en el Departamento de Teoría de la Señal, Telemática y Comunicaciones de la Universidad de Granada. Esta memoria constituye la Tesis que D. Jorge Navarro Ortiz presenta para optar al grado de Doctor con Mención Europea por la Universidad de Granada.

Granada, Julio de 2010

Fdo.:

Dr. D. Juan Manuel López Soler
Director de la Tesis

The darkest hour is just before the dawn.

– Thomas Fuller

A Mercedes

Abstract

Since Guglielmo Marconi first developed a wireless telegraph system at the end of the XIX century, wireless communications have noticeably evolved. This progress has been influenced by voice communications –e.g. cellular networks–, but the Internet’s growth has promoted the usage of other services. These are not only circumscribed to web browsing and e-mail, but also new applications have appeared with requirements that were not considered in the initial design of the Internet, such as video broadcasting, audio streaming or voice over IP.

Similarly to the Internet, the IEEE 802.11 standard was not conceived to support Quality of Service but it is based on a *best effort* scheme. For that reason, the 802.11e amendment introduced new medium access schemes that added QoS support to these environments. Even though there exists a number of solutions based on these standards, none of them is able to simultaneously and efficiently handle a heterogeneous traffic mix (e.g. including real-time and TCP traffic), being this the main objective of this thesis.

For the proposed solutions, it has been considered as initial hypothesis that the incorporation of Quality of Experience (QoE) in the design of radio resource management algorithms will introduce a significant improvement on the system performance. For this reason, the first contribution has been the definition of a methodology based on the end-user’s perception. This methodology is based on the usage of utility functions, which map objective and measurable statistics (e.g. data rate, delay and losses) into the subjective quality seen by the user. These functions do not only permit to evaluate the system performance in terms of QoE, but also to take into account the user’s opinion for taking decisions in the different developed solutions. In addition, the usage of the perceived quality –using the MOS scale– permits to perform a comparison between services with different characteristics, which could not be done using the typical indicators (e.g. VoIP performance is influenced by delay, jitter and losses whereas FTP efficiency depends mainly on data rate).

The second contribution consists in a packet scheduling discipline which is able to

x | Abstract

handle different types of services, such as real-time or TCP-based applications. For this purpose, an existing scheduler from the ATM switching literature, the *Monolithic Shaper-Scheduler*, has been adapted for wireless networks using HCCA –coordinated medium access mechanism with parameterized QoS support–. This scheduler has a proper behavior for these services individually, thanks to its main characteristics (guaranteed throughput and bounded delay).

The third contribution proposed a solution for scenarios with a heterogeneous traffic mix. To achieve this goal, a general framework for resource sharing, designated the *Territory Method*, has been designed. It splits the wireless capacity into territories for the different *traffic classes*. The size of these territories is adjusted based on the QoE –estimated in real-time using the previously defined utility functions–, in addition to the definition of a dedicated capacity for guaranteeing a minimum performance –thus avoiding the starvation of that service type–.

Based on this general framework, the fourth contribution consists in the definition of several policies that the network administrator may use to customize the criteria for sharing the transmission opportunities. Four basic policies have been designed for resource sharing: preemptive, non-preemptive, MOS-based fairness and revenue improvement.

The performance of these solutions has been evaluated and analyzed through extensive simulations. This analysis demonstrated that, in the considered scenarios, the proposed solutions outperform some reference schedulers from the HCCA literature, in terms of experienced quality, revenue and fairness between services.

The last contribution of this work is an enhancement of the TCP protocol for wireless networks. It is based on the removal of redundant functionalities already available in the MAC layer when using an HCCA scheduler. For this purpose, a performance enhancing proxy was used to *split* the TCP connection (available in the current literature and known as *split TCP*) in two parts: wired and wireless. The evaluation, performed in some relevant scenarios, shows that there is a significant throughput gain for those flows using this solution.

Resumen

Desde que a finales del siglo XIX Guglielmo Marconi desarrollara el primer sistema telegráfico sin cables, las comunicaciones inalámbricas han evolucionado de forma notable. Este progreso ha sido propiciado fundamentalmente por las comunicaciones de voz –como es el caso de las comunicaciones móviles–, pero el desarrollo de Internet ha fomentado el uso de otros servicios. Éstos no sólo se circunscriben a la navegación web y al correo electrónico, sino que han aparecido nuevas aplicaciones con unos requisitos que no fueron contemplados en la concepción inicial de la Red, como pueden ser la difusión de vídeo, el *streaming* de audio o la voz sobre IP.

Al igual que Internet, el estándar IEEE 802.11 no fue concebido para soportar requisitos de Calidad de Servicio, sino que se basa en un esquema de *mejor esfuerzo*. Por este motivo, la norma 802.11e introdujo nuevos esquemas de acceso al medio que añaden soporte de QoS en este tipo de entornos. Si bien existen múltiples soluciones basadas en estas normas, ninguna es capaz de soportar simultánea y eficazmente un tráfico heterogéneo (e.g. una mezcla de servicios de tiempo real y de tráfico elástico), siendo éste el objetivo fundamental de la tesis.

Para las soluciones propuestas, se ha considerado como hipótesis de partida que la incorporación de la Calidad de Experiencia (QoE) en el diseño de algoritmos de gestión de recursos introducirá una mejora significativa en el rendimiento del sistema. Por ello, la primera contribución ha sido la definición de una metodología basada en la percepción del usuario final. Ésta se basa en el uso de funciones de utilidad que mapean estadísticas objetivas y medibles de la red (e.g. velocidad de transmisión, retardo y pérdidas) en la calidad subjetiva que apreciaría un usuario. Estas funciones permiten no sólo evaluar el rendimiento del sistema en términos de QoE, sino que la opinión del usuario también se puede tener en cuenta para la toma de decisiones en las diferentes soluciones desarrolladas. Además, la calidad percibida –usando la escala MOS– permite la comparación entre diferentes servicios con características distintas, algo que no se podría realizar con los indicadores habituales (e.g. el rendimiento de VoIP se ve influido por el retardo, *jitter* y pérdidas mientras que el de FTP depende fundamentalmente de la

velocidad de transmisión).

La segunda contribución consiste en una disciplina de planificación de paquetes capaz de contemplar servicios de diferentes tipos, como aquellos de tiempo real o los basados en TCP. Para ello, se ha adaptado un planificador existente en la literatura sobre conmutadores ATM, el *Monolithic Shaper-Scheduler*, para el entorno de redes IEEE 802.11 con HCCA –mecanismo coordinado de acceso al medio con soporte parametrizado de QoS–. Sus características (tasa de transmisión garantizada y retardo acotado) hacen que tenga un comportamiento adecuado para estos servicios individualmente.

La tercera contribución propone una solución para escenarios con tráfico heterogéneo, i.e. con mezcla de tráficos. Para ello se ha diseñado una plataforma general para la gestión de recursos, denominada el *Método de los Territorios*, que divide la capacidad del medio inalámbrico en territorios para las diferentes *clases de tráfico*. El tamaño de estos territorios se establece a partir de la QoE –estimada en tiempo real a partir de las funciones de utilidad previamente definidas–, además de incluir una capacidad dedicada para garantizar un rendimiento mínimo y evitar la inanición de algún tipo de servicio.

Basándose en esta plataforma general, la cuarta contribución consiste en la definición de políticas para que el administrador de red personalice los criterios a utilizar para compartir las oportunidades de transmisión. Así, se han diseñado cuatro políticas básicas para el reparto de los recursos: con derecho preferente, sin derecho preferente, justo en términos de QoE y para aumentar el beneficio.

El rendimiento de estas soluciones ha sido evaluado y analizado a través de los resultados obtenidos en diferentes campañas de simulación, comparándose con algunos planificadores de referencia existentes en la literatura sobre HCCA. Este análisis demuestra que, en los escenarios considerados, las soluciones propuestas funcionan mejor en términos de calidad experimentada, beneficio económico y justicia entre servicios.

La última contribución de este trabajo es una mejora del protocolo TCP para redes WLAN. Ésta se basa en la eliminación de funcionalidades redundantes con aquellas realizadas por la capa MAC cuando se utiliza un planificador de paquetes basado en

HCCA. Para poder realizar esta simplificación, ha sido necesario utilizar un *proxy* de rendimiento mejorado existente en la literatura, creando una conexión TCP *partida* (*split TCP*). La evaluación realizada en algunos escenarios relevantes muestra que se mejora significativamente la velocidad de transmisión de los flujos que utilizan esta solución.

Publicaciones

Los siguientes artículos incluyen algunas de las principales contribuciones del presente trabajo:

- **Jorge Navarro-Ortiz**, Pablo Ameigeiras, Juan J. Ramos-Munoz, and Juan M. Lopez-Soler. An 802.11e HCCA scheduler with an end-to-end quality aware territory method. *Computer Communications*, 32(11):1281–1297, 2009.
- **Jorge Navarro-Ortiz**, Juan M. Lopez-Soler, and Giovanni Stea. Quality of experience based resource sharing in iee 802.11e hcca. In *Proceedings of the 16th European Wireless conference (EW 2010)*, pages 1–8, April 2010.

Además, algunas figuras y resultados relevantes aparecen en las siguientes publicaciones:

- Pablo Ameigeiras, Juan J. Ramos-Munoz, **Jorge Navarro-Ortiz**, Preben Mogensen, and Juan M. Lopez-Soler. QoE oriented cross-layer design of a resource allocation algorithm in beyond 3G systems. *Computer Communications*, 33(5):571–582, October 2009.
- Pablo Ameigeiras, Juan J. Ramos-Munoz, **Jorge Navarro-Ortiz**, Juan Ramiro-Moreno, and Juan M. Lopez-Soler. QoE evaluation of scheduling algorithms for NRT services in LTE. In *Proceedings of the ICT Mobile and Wireless Communications Summit (ICT-MobileSummit 2009)*, June 2009.

Por ultimo, se espera que el trabajo desarrollado permita publicar –en revistas y/o conferencias internacionales de reconocido prestigio– las contribuciones en los siguientes temas:

- Extensión del artículo “*Quality of Experience Based Resource Sharing in IEEE 802.11e HCCA*” para su publicación en una revista internacional.

xvi | Publicaciones

- Diseño y evaluación de la propuesta para la mejora del rendimiento de TCP en redes IEEE 802.11 (véase el Capítulo 7).
- Instrumentos para ayuda a la simulación –en concreto para la herramienta Network Simulator 2–, como puede ser la generación *online* de estadísticas de Calidad de Experiencia (que permitan tanto la evaluación del rendimiento como la toma de decisiones) o herramientas para el postprocesado de los resultados.

Reconocimientos

El trabajo realizado en esta tesis ha sido parcialmente financiado por el Proyecto TIN2009-13992-C02-02 del Ministerio de Ciencia e Innovación de España.

Agradecimientos

Quiero expresar mi gratitud a las personas que, de una u otra manera, me han ayudado, aguantado, comprendido, animado, escuchado... durante el desarrollo de esta Tesis.

GRACIAS a Juan Manuel López Soler, por su dedicación en la dirección de ésta y por sus ánimos durante los últimos años. GRACIAS.

GRACIAS a mis compañeros y amigos del grupo de investigación en Señales, Telemática y Comunicaciones de la Universidad de Granada, por su apoyo en todos los aspectos. Y en particular a Pablo y Juanjo, por su inestimable ayuda con –entre otras cosas– las funciones de utilidad.

GRACIAS al grupo de investigación *Computer Network Group* del *Dipartimento di Ingegneria dell'Informazione* de la *Università di Pisa*, en especial al profesor Stea. También a Claudio Ciconetti, al profesor Lenizini y al resto de profesores, investigadores y doctorandos, por su cálida acogida y por sus constantes enseñanzas, sin las que la mención europea de esta tesis no hubiera sido posible.

GRACIAS a mi familia y amigos, por estar ahí cuando les he necesitado. GRACIAS a mis padres, Francisco y Josefa, que me quieren sin reservas. Ellos me han enseñado más que nadie con su ejemplo. GRACIAS a mis hermanos, Francisco y Antonio, por alegrarse de mis éxitos y entristecerse con mis derrotas.

GRACIAS a Mercedes, por animarme, por aguantarme, por quererme, por ser como eres. GRACIAS.

Acrónimos

AC	Categoría de Acceso (<i>Access Category</i>)
AP	Punto de Acceso (<i>Access Point</i>)
BPSK	Modulación Binaria por Desplazamiento de Fase (<i>Binary Phase-Shift Keying</i>)
CCK	Modulación de Código Complementario (<i>Complementary Code Keying</i>)
CFP	Período Libre de Contención (<i>Contention Free Period</i>)
CP	Período de Contención (<i>Contention Period</i>)
CS/CCA	Procedimiento del nivel físico de 802.11 para realizar la Comprobación de Portadora / Evaluación de Canal Libre (<i>Carrier Sense / Clear Channel Assesment</i>)
CT	Clases de Tráfico
DBPSK	Modulación Binaria por Desplazamiento Diferencial de Fase (<i>Differential Binary Phase-Shift Keying</i>)
DQPSK	Modulación en Cuadratura por Desplazamiento Diferencial de Fase (<i>Differential Quadrature Phase-Shift Keying</i>)
DSSS	Espectro Expandido por Secuencia Directa (<i>Direct Sequence Spread Spectrum</i>)
EDCA	Acceso Mejorado al Canal de forma Distribuida (<i>Enhanced Distributed Channel Access</i>)
FDDI	Interfaz de Datos Distribuidos por Fibra (<i>Fiber Distributed Data Interface</i>)
FEC	Corrección de Errores hacia Delante (<i>Forward Error Correction</i>)
FHSS	Espectro Expandido por Salto de Frecuencias (<i>Frequency Hopping Spread Spectrum</i>)
FTP	Protocolo de Transferencia de Ficheros (<i>File Transfer Protocol</i>)
GFSK	Modulación por Desplazamiento de Frecuencias con filtro Gaussiano (<i>Gaussian Frequency-Shift Keying</i>)
HCCA	Acceso al Canal Controlado por HCF (<i>HCF Controlled Channel Access</i>)
HCF	Función de Coordinación Híbrida (<i>Hybrid Coordination Function</i>)
HSPA	Acceso de Paquetes de Alta Velocidad (<i>High Speed Packet Access</i>) extiende y mejora el rendimiento de las redes 3G basadas en WCDMA.
IEEE	Instituto de Ingenieros Eléctricos y Electrónicos (<i>Institute of Electrical and</i>

xxii | Acrónimos

	<i>Electronics Engineers)</i>
IP	Protocolo de Internet (<i>Internet Protocol</i>)
IR	Infrarrojo
ISM	Banda de frecuencias para usos Industriales, Científicos y Médicos (<i>Industrial, Science and Medical</i>)
ITU-T	Unión Internacional de Telecomunicación, sector de estandarización en Telecomunicación (<i>International Telecommunication Union, Telecommunication standardization sector</i>). También denominada UIT-T en español.
LR	Familia de servidores Tasa–Retardo (<i>Latency–Rate servers</i>)
LTE	Long Term Evolution es un nuevo estándar de la norma 3GPP, definida como evolución de 3GPP UMTS.
MOS	Nota Media de Opinión (<i>Mean Opinion Score</i>)
MSS	Planificador–Conformador Monolítico (<i>Monolithic Shaper–Scheduler</i>)
MSS	Tamaño Máximo de Segmento en TCP (<i>Maximum Segment Size</i>)
OFDM	Multiplexación por División en Frecuencias Ortogonales (<i>Orthogonal Frequency Division Modulation</i>)
OSI	Interconexión de Sistemas Abiertos (<i>Open System Interconnection</i>)
PBCC	Codificación Convolutiva Binaria de Paquetes (<i>Packet Binary Convolutional Coding</i>)
PDU	Unidad de Datos de Protocolo (<i>Protocol Data Unit</i>)
PEP	<i>Proxy</i> de rendimiento mejorado (<i>Performance Enhanced Proxy</i>)
PESQ	Evaluación de la Calidad Vocal por Percepción (<i>Perceptual Evaluation of Speech Quality</i>)
PLCP	Protocolo de Convergencia de la Capa Física (<i>Physical Layer Convergence Protocol</i>)
PMD	Capa Dependiente del Medio Físico (<i>Physical Medium Dependent</i>)
PPM	Modulación por Posición de Pulsos (<i>Pulse-Position Modulation</i>)
QAM	Modulación en Amplitud en Cuadratura (<i>Quadrature Amplitude Modulation</i>)
QAP	Punto de Acceso con soporte de QoS (<i>QoS Access Point</i>)
QoE	Calidad de Experiencia (<i>Quality of Experience</i>)
QoS	Calidad de Servicio (<i>Quality of Service</i>)
QPSK	Modulación en Cuadratura por Desplazamiento de Fase (<i>Quadrature Phase-Shift Keying</i>)
QSTA	Estación inalámbrica con soporte de QoS (<i>QoS Station</i>)

RRM	Gestión de Recursos Radio (<i>Radio Resource Management</i>)
RTT	Tiempo de Ida y Vuelta (<i>Round-Trip Time</i>)
SDU	Unidad de Datos de Servicio (<i>Service Data Unit</i>)
SNR	Relación Señal – Ruido (<i>Signal to Noise Ratio</i>)
STA	Estación inalámbrica (<i>Station</i>)
TCP	Protocolo de Control de Transporte (<i>Transport Control Protocol</i>)
UDP	Protocolo de Datagramas de Usuario (<i>User Datagram Protocol</i>)
UMTS	Sistema Universal de Telecomunicaciones Móviles (<i>Universal Mobile Telecommunications System</i>) es una de las tecnologías móviles de tercera generación del 3GPP.
U-NII	Banda de frecuencias para Infraestructura de Información Nacional sin Licencia (<i>Unlicensed National Information Infrastructure</i>)
VoIP	Voz sobre IP (<i>Voice over IP</i>)
WLAN	Red Inalámbrica de Área Local (<i>Wireless Local Area Network</i>)

Contenido

1. Introducción.....	1
1.1 Motivación	1
1.2 Objetivos	3
1.3 Contribuciones de la tesis.....	4
1.4 Organización del documento.....	6
2. Redes WLAN	9
2.1 Estándar IEEE 802.11	10
2.2 Arquitectura	14
2.2.1 Capa física	17
2.2.2 Capa de Control de Acceso al Medio (MAC).....	22
2.3 Calidad de Servicio en redes WLAN: estándar IEEE 802.11e	30
2.3.1 Acceso al canal de forma distribuida: EDCA	30
2.3.2 Acceso al canal de forma coordinada: HCCA	34
2.3.3 Requisitos de los planificadores para HCCA.....	35
2.3.4 Planificador de referencia en HCCA	36
2.4 Trabajos relacionados.....	37
2.5 Conclusiones	40
3. Calidad de Experiencia (QoE).....	41
3.1 Clases de Tráfico.....	43
3.2 Funciones de utilidad	45
3.2.1 Función de utilidad para tráfico VoIP.....	46
3.2.2 Función de utilidad para tráfico <i>streaming</i> de vídeo.....	47
3.2.3 Función de utilidad para tráfico <i>streaming</i> de vídeo sobre TCP.....	48

xxvi | Contenido

3.2.4	Función de utilidad para tráfico web.....	52
3.2.5	Función de utilidad para tráfico FTP	54
3.3	Metodología para la evaluación de rendimiento.....	54
3.4	Conclusiones	57
4.	Una variante del Planificador–Conformador Monolítico para HCCA	59
4.1	Funcionamiento del planificador MSS.....	60
4.2	Adaptación para HCCA.....	66
4.3	Evaluación de rendimiento	68
4.3.1	Condiciones experimentales.....	69
4.3.2	Resultados	75
4.4	Conclusiones	90
5.	Gestión de recursos para servicios heterogéneos: Método de los Territorios.....	91
5.1	Funcionamiento del Método de los Territorios	93
5.2	Evaluación de rendimiento	100
5.2.1	Condiciones experimentales.....	100
5.2.2	Resultados	102
5.3	Conclusiones	113
6.	Políticas de gestión de recursos basadas en Calidad de Experiencia.....	115
6.1	Reparto con derecho preferente.....	116
6.2	Reparto sin derecho preferente.....	116
6.3	Reparto justo en términos de Calidad de Experiencia.....	118
6.4	Reparto para aumento de beneficio	118
6.5	Evaluación de rendimiento	120
6.5.1	Condiciones experimentales.....	121
6.5.2	Resultados	121
6.6	Comparativa entre las diferentes aproximaciones	134

6.7	Conclusiones	139
7.	TCP mejorado sobre HCCA.....	141
7.1	Una variante de TCP partido para HCCA.....	142
7.2	Evaluación de rendimiento.....	147
7.2.1	Condiciones experimentales	147
7.2.2	Resultados.....	148
7.3	Conclusiones	160
8.	Conclusiones	163
9.	Conclusions	169
	Apéndice A	173
A.	Thesis summary	173
A.1	Introduction and problem statement.....	173
A.1.1	Dissertation's goals.....	175
A.2	IEEE 802.11e and HCCA	176
A.2.1	Related works	179
A.3	Quality of Experience	181
A.4	Radio resource management strategies for HCCA	188
A.4.1	A variant of the Monolithic Shaper-Scheduler for HCCA	188
A.4.2	Resource management for heterogeneous traffic mix based on QoE: the Territory Method	193
A.4.3	Resource sharing policies for HCCA.....	197
A.4.4	Main results of the proposed solutions for resource allocation	200
A.5	Improved TCP over HCCA.....	205
A.5.1	Main results for improved TCP over HCCA	209
A.6	Conclusions	212

xxviii | Contenido

B. Retardo extra introducido por la capa MAC de IEEE 802.11e	213
C. Comprobación de la implementación de la variante de MSS para HCCA	215
D. Selección de parámetros para el Método de los Territorios	221
Bibliografía	235

Índice de figuras

Figura 2.1. Arquitectura de protocolos en el estándar IEEE 802.11.....	11
Figura 2.2. Elementos de una red 802.11.....	15
Figura 2.3. Topología de una red independiente o <i>ad hoc</i>	16
Figura 2.4. Topología de una red de infraestructura en una BSA.....	16
Figura 2.5. Red de infraestructura con varias BSAs.....	17
Figura 2.6. Formato de trama MAC.	23
Figura 2.7. CSMA/CA en DCF.	25
Figura 2.8. Esquema de acceso RTS / CTS.	26
Figura 2.9. Ciclos de PCF y DCF.	28
Figura 2.10. Categorías de Acceso en el mecanismo de acceso EDCA.	31
Figura 2.11. Ejemplo de prioridad estadística utilizando EDCA.....	33
Figura 2.12. Intervalo de baliza de IEEE 802.11e.	35
Figura 3.1. Ejemplo de función de utilidad para tráfico <i>VoIP</i> (entorno interior y código G.711).	47
Figura 3.2. Ejemplo de función de utilidad para tráfico <i>streaming</i> de vídeo (RM, 15 fps, SBR 1Mbps).	49
Figura 3.3. Evolución artificial de MOS para streaming de vídeo sobre TCP.	51

xxx | Índice de figuras

Figura 3.4. Función de utilidad para tráfico web en función del tiempo de descarga.	53
Figura 3.5. Función de utilidad para tráfico <i>web</i> en función de la tasa binaria.....	53
Figura 3.6. Función de utilidad para tráfico FTP.....	55
Figura 4.1. Conformador dual-leaky-bucket.....	61
Figura 4.2. Actividades de <i>test de conformidad virtual</i> y <i>encolado de flujo</i> del planificador MSS.....	63
Figura 4.3. Actividades de desencolado de flujo y selección de flujo en el planificador MSS.....	64
Figura 4.4. Escenario heterogéneo utilizado en la experimentación.....	70
Figura 4.5. Modelo de transición de estados para una conversación.....	72
Figura 4.6. Calidad experimentada para los diferentes servicios usando el planificador MSS con carga de tráfico media.....	76
Figura 4.7. Justicia entre servicios del planificador MSS con carga de tráfico media.	76
Figura 4.8. Calidad experimentada para los diferentes servicios usando el planificador MSS con carga de tráfico alta.....	77
Figura 4.9. Justicia entre servicios del planificador MSS con carga de tráfico alta...	77
Figura 4.10. Calidad experimentada para los diferentes servicios usando el planificador MSS con carga de tráfico muy alta.	79
Figura 4.11. Justicia entre servicios del planificador MSS con carga de tráfico muy alta.	79
Figura 4.12. Calidad experimentada para los diferentes servicios usando el planificador de referencia con carga de tráfico media.....	82

Figura 4.13. Justicia entre servicios del planificador de referencia con carga de tráfico media.....	82
Figura 4.14. Calidad experimentada para los diferentes servicios usando el planificador de referencia con carga de tráfico alta.....	83
Figura 4.15. Justicia entre servicios del planificador de referencia con carga de tráfico alta.....	83
Figura 4.16. Calidad experimentada para los diferentes servicios usando el planificador WTTP con carga de tráfico media.	85
Figura 4.17. Justicia entre servicios del planificador WTTP con carga de tráfico media.	85
Figura 4.18. Calidad experimentada para los diferentes servicios usando el planificador WTTP con carga de tráfico alta.	86
Figura 4.19. Justicia entre servicios del planificador WTTP con carga de tráfico alta.	86
Figura 4.20. Comparación de la calidad experimentada en FTP para los planificadores MSS y FHCF.	88
Figura 4.21. Comparación de la calidad experimentada en VoIP para los planificadores MSS y FHCF.	89
Figura 4.22. Comparación del rendimiento (tasa binaria) en FTP para los planificadores MSS y FHCF.	89
Figura 5.1. Ejemplo de definición de territorios para una clase de tráfico genérica..	94
Figura 5.2. Ejemplo de la relación entre la calidad subjetiva y la evolución del territorio para una clase de tráfico genérica.	95
Figura 5.3. Diagrama de flujo del Método de los Territorios.	99
Figura 5.4. Calidad experimentada para los diferentes servicios utilizando el Método de los Territorios con carga de tráfico media.	104

xxxii | Índice de figuras

Figura 5.5. Justicia entre servicios utilizando el Método de los Territorios con carga de tráfico media.	104
Figura 5.6. Distribución de recursos utilizando el Método de los Territorios con carga de tráfico media.	104
Figura 5.7. Calidad experimentada para los diferentes servicios utilizando el Método de los Territorios con carga de tráfico alta.	106
Figura 5.8. Justicia entre servicios utilizando el Método de los Territorios con carga de tráfico alta.	106
Figura 5.9. Distribución de recursos utilizando el Método de los Territorios con carga de tráfico alta.	106
Figura 5.10. Calidad experimentada para los diferentes servicios utilizando el Método de los Territorios con carga de tráfico muy alta.	108
Figura 5.11. Justicia entre servicios utilizando el Método de los Territorios con carga de tráfico muy alta.	108
Figura 5.12. Distribución de recursos utilizando el Método de los Territorios con carga de tráfico muy alta.	108
Figura 5.13. Perfil de tráfico VoIP durante el experimento con carga dinámica.	110
Figura 5.14. Evolución temporal de los resultados utilizando el Método de los Territorios para VoIP.	111
Figura 5.15. Evolución temporal de los resultados utilizando el Método de los Territorios para <i>streaming</i> de vídeo.	111
Figura 5.16. Evolución temporal de los resultados utilizando el Método de los Territorios para vídeo sobre TCP.	111
Figura 5.17. Evolución temporal de los resultados utilizando el Método de los Territorios para tráfico web.	111
Figura 5.18. Evolución temporal de los resultados utilizando el Método de los Territorios para FTP.	112

Figura 6.1. Diagrama de flujo del Método de los Territorios utilizando la política de <i>reparto con derecho preferente</i>	117
Figura 6.2. Diagrama de flujo de la política de <i>reparto sin derecho preferente</i>	118
Figura 6.3. Diagrama de flujo de la política de <i>reparto justo en términos de Calidad de Experiencia</i>	119
Figura 6.4. Diagrama de flujo de la política de <i>reparto para aumento de beneficio</i>	120
Figura 6.5. Resultados usando la política de <i>reparto con derecho preferente</i> para <i>carga media</i>	122
Figura 6.6. Resultados usando la política de <i>reparto con derecho preferente</i> para <i>carga alta</i>	123
Figura 6.7. Resultados usando la política de <i>reparto con derecho preferente</i> para <i>carga muy alta</i>	124
Figura 6.8. Resultados usando la política de <i>reparto sin derecho preferente</i> para <i>carga media</i>	126
Figura 6.9. Resultados usando la política de <i>reparto sin derecho preferente</i> para <i>carga alta</i>	127
Figura 6.10. Resultados usando la política de <i>reparto sin derecho preferente</i> para <i>carga muy alta</i>	128
Figura 6.11. Resultados usando la política de <i>reparto justo en términos de Calidad de Experiencia</i> para <i>carga media</i>	129
Figura 6.12. Resultados usando la política de <i>reparto justo en términos de Calidad de Experiencia</i> para <i>carga alta</i>	130
Figura 6.13. Resultados usando la política de <i>reparto justo en términos de Calidad de Experiencia</i> para <i>carga muy alta</i>	131
Figura 6.14. Resultados usando la política de <i>aumento de beneficio</i> para <i>carga media</i>	132

xxxiv | Índice de figuras

Figura 6.15. Resultados usando la política de <i>aumento de beneficio</i> para <i>carga alta</i>	133
Figura 6.16. Resultados usando la política de <i>aumento de beneficio</i> para <i>carga muy alta</i>	134
Figura 6.17. Comparativa entre las diferentes aproximaciones para la gestión de recursos en el caso de <i>carga media</i>	136
Figura 6.18. Comparativa entre las diferentes aproximaciones para la gestión de recursos en el caso de <i>carga alta</i>	137
Figura 6.19. Comparativa entre las diferentes aproximaciones para la gestión de recursos en el caso de <i>carga muy alta</i>	139
Figura 7.1. Escenario con una conexión <i>TCP convencional</i>	143
Figura 7.2. Escenario con una conexión <i>TCP partida</i>	144
Figura 7.3. Escenario con una conexión <i>TCP partido & UDP</i>	147
Figura 7.4. Resultados para 1 descarga FTP y un retardo en la red cableada de 2 ms.	150
Figura 7.5. Resultados para 1 descarga FTP y un retardo en la red cableada de 50 ms.	150
Figura 7.6. Resultados para 1 descarga FTP y un retardo en la red cableada de 50 ms.	150
Figura 7.7. Resultados para 1 descarga FTP y un retardo en la red cableada de 500 ms.	150
Figura 7.8. Resultados para 1 descarga FTP, 3 Mbps de tráfico UDP y un retardo (cable) de 2 ms.	152
Figura 7.9. Resultados para 1 descarga FTP, 3 Mbps de tráfico UDP y un retardo (cable) de 50 ms.	152

Figura 7.10. Resultados para 1 descarga FTP, 3 Mbps de tráfico UDP y un retardo (cable) de 100 ms.	152
Figura 7.11. Resultados para 1 descarga FTP, 3 Mbps de tráfico UDP y un retardo (cable) de 500 ms.	152
Figura 7.12. Resultados para varios usuarios FTP, un retardo (cable) de 2 ms.	154
Figura 7.13. Resultados para varios usuarios FTP, un retardo (cable) de 50 ms.	154
Figura 7.14. Resultados para varios usuarios FTP, un retardo (cable) de 100 ms. ...	154
Figura 7.15. Resultados para varios usuarios FTP, un retardo (cable) de 500 ms. ...	154
Figura 7.16. Ganancia para varios usuarios FTP.	155
Figura 7.17. Resultados para varios usuarios FTP, un retardo (cable) de 2 ms y 10% de errores.	156
Figura 7.18. Resultados para varios usuarios FTP, un retardo (cable) de 50 ms y 10% de errores.	156
Figura 7.19. Resultados para varios usuarios FTP, un retardo (cable) de 100 ms y 10% de errores.	156
Figura 7.20. Resultados para varios usuarios FTP, un retardo (cable) de 500 ms y 10% de errores.	156
Figura 7.21. Resultados para varios usuarios FTP, un retardo (cable) de 500 ms y 20% de errores.	157
Figura 7.22. Resultados para varios usuarios FTP, un retardo (cable) de 500 ms y 20% de errores.	157
Figura 7.23. Resultados para varios usuarios FTP, un retardo (cable) de 500 ms y 20% de errores.	157
Figura 7.24. Resultados para varios usuarios FTP, un retardo (cable) de 500 ms y 20% de errores.	157

xxxvi | Índice de figuras

Figura 7.25. Resultados para varios usuarios FTP, un retardo (cable) de 2 ms y 30% de errores.	158
Figura 7.26. Resultados para varios usuarios FTP, un retardo (cable) de 50 ms y 30% de errores.	158
Figura 7.27. Resultados para varios usuarios FTP, un retardo (cable) de 100 ms y 30% de errores.	158
Figura 7.28. Resultados para varios usuarios FTP, un retardo (cable) de 500 ms y 30% de errores.	158
Figura 7.29. Ganancia para varios usuarios FTP y 10% de errores.	159
Figura 7.30. Ganancia para varios usuarios FTP y 20% de errores.	159
Figura 7.31. Ganancia para varios usuarios FTP y 30% de errores.	159
Figure A.1. IEEE 802.11e beacon interval.	177
Figure A.2. Artificial MOS evolution for TCP video streaming.	186
Figure A.3. Flowchart of the flow dequeue activity and the flow selection procedure.	190
Figure A.4. Flowchart of the test for virtual compliance and the flow enqueue activity.	191
Figure A.5. Example of territories definition for a generic traffic class.	194
Figure A.6. Example of the relationship between subjective quality and the territory evolution for a generic traffic class.	194
Figure A.7. Flowchart of the territory method using the <i>preemptive</i> policy.	197
Figure A.8. Flowchart for the <i>non-preemptive</i> policy.	199
Figure A.9. Flowchart for the <i>MOS-driven fairness</i> policy.	199

Figure A.10. Flowchart for the <i>revenue improvement</i> policy.	200
Figure A.11. Comparison among the different RRM approaches in the case with <i>medium load</i>	202
Figure A.12. Comparison among the different RRM approaches in the case with <i>high load</i>	203
Figure A.13. Comparison among the different RRM approaches in the case with <i>high load</i>	205
Figure A.14. Scenario with a <i>normal TCP</i> connection.	207
Figure A.15. Scenario with a <i>split TCP</i> connection.	207
Figure A.16. Scenario with a <i>split TCP & UDP</i> connection.	209
Figure A.17. Results for 1 FTP download and one-way delay of 2 ms (wired link).	210
Figure A.18. Results for 1 FTP download with 3 Mbps of UDP traffic and one-way delay of 2 ms (wired link).	210
Figure A.19. Gain for several FTP users and 0% frame error rate.	211
Figure A.20. Gain for several FTP users and 10% frame error rate.	211
Figure A.21. Gain for several FTP users and 20% frame error rate.	212
Figure A.22. Gain for several FTP users and 30% frame error rate.	212
Figura C.1. Retardo máximo del planificador MSS con tráfico VoIP.	217
Figura C.2. Retardo máximo del planificador MSS con tráfico mixto (VoIP y FTP).	218
Figura C.3. Diferencia mínima entre el servicio prestado y el servicio requerido por HCCA del planificador MSS con tráfico VoIP.	219

xxxviii | Índice de figuras

Figura C.4. Diferencia mínima entre el servicio prestado y el servicio requerido por HCCA para tráfico VoIP del planificador MSS con tráfico mixto (VoIP y FTP).....	219
Figura C.5. Diferencia mínima entre el servicio prestado y el servicio requerido por HCCA para tráfico FTP del planificador MSS con tráfico mixto (VoIP y FTP).....	220
Figura D.1. Evolución temporal del tráfico VoIP durante el experimento con carga dinámica.	222
Figura D.2. Evolución temporal de MOS instantáneo, MOS promediado y territorio para VoIP (<i>CT conversacional</i>) utilizando PFM=0.2 s y PAT=1 s.....	223
Figura D.3. Evolución temporal de MOS instantáneo, MOS promediado y territorio para <i>streaming</i> de vídeo (<i>CT streaming</i>) PFM=0.2 s y PAT=1 s.	223
Figura D.4. Evolución temporal de MOS instantáneo, MOS promediado y territorio para <i>streaming</i> de vídeo sobre TCP (<i>CT TCP de tiempo rea</i>) utilizando PFM=0.2 s y PAT=1 s.....	223
Figura D.5. Evolución temporal de MOS instantáneo, MOS promediado y territorio para navegación web (<i>CT interactiva</i>) utilizando PFM=0.2 s y PAT=1 s.	223
Figura D.6. Evolución temporal de MOS instantáneo, MOS promediado y territorio para FTP (<i>CT de fondo</i>) utilizando PFM=0.2 s y PAT=1 s.	224
Figura D.7. Evolución temporal de MOS instantáneo, MOS promediado y territorio para VoIP (<i>CT conversacional</i>) utilizando PFM=1 s y PAT=5 s.....	224
Figura D.8. Evolución temporal de MOS instantáneo, MOS promediado y territorio para <i>streaming</i> de vídeo (<i>CT streaming</i>) PFM=1 s y PAT=5 s.	225
Figura D.9. Evolución temporal de MOS instantáneo, MOS promediado y territorio para <i>streaming</i> de vídeo sobre TCP (<i>CT TCP de tiempo rea</i>) utilizando PFM=1 s y PAT=5 s.....	225

Figura D.10. Evolución temporal de MOS instantáneo, MOS promediado y territorio para navegación web (<i>CT interactiva</i>) utilizando PFM=1 s y PAT=5 s.	225
Figura D.11. Evolución temporal de MOS instantáneo, MOS promediado y territorio para FTP (<i>CT de fondo</i>) utilizando PFM=1 s y PAT=5 s.	225
Figura D.12. Evolución temporal de MOS instantáneo, MOS promediado y territorio para VoIP (<i>CT conversacional</i>) utilizando PFM=5 s y PAT=25 s.	226
Figura D.13. Evolución temporal de MOS instantáneo, MOS promediado y territorio para <i>streaming</i> de vídeo (<i>CT streaming</i>) PFM=5 s y PAT=25 s.	226
Figura D.14. Evolución temporal de MOS instantáneo, MOS promediado y territorio para <i>streaming</i> de vídeo sobre TCP (<i>CT TCP de tiempo rea</i>) utilizando PFM=5 s y PAT=25 s.	227
Figura D.15. Evolución temporal de MOS instantáneo, MOS promediado y territorio para navegación web (<i>CT interactiva</i>) utilizando PFM=5 s y PAT=25 s.	227
Figura D.16. Evolución temporal de MOS instantáneo, MOS promediado y territorio para FTP (<i>CT de fondo</i>) utilizando PFM=5 s y PAT=25 s.	227
Figura D.17. Evolución temporal de MOS instantáneo, MOS promediado y territorio para VoIP (<i>CT conversacional</i>) utilizando $\alpha_i = 0.1$	229
Figura D.18. Evolución temporal de MOS instantáneo, MOS promediado y territorio para <i>streaming</i> de vídeo (<i>CT streaming</i>) utilizando $\alpha_i = 0.1$	229
Figura D.19. Evolución temporal de MOS instantáneo, MOS promediado y territorio para <i>streaming</i> de vídeo sobre TCP (<i>CT TCP de tiempo rea</i>) utilizando $\alpha_i = 0.1$	229
Figura D.20. Evolución temporal de MOS instantáneo, MOS promediado y territorio para navegación web (<i>CT interactiva</i>) utilizando $\alpha_i = 0.1$	229
Figura D.21. Evolución temporal de MOS instantáneo, MOS promediado y territorio para FTP (<i>CT de fondo</i>) utilizando $\alpha_i = 0.1$	230

xl | Índice de figuras

Figura D.22. Evolución temporal de MOS instantáneo, MOS promediado y territorio para VoIP (<i>CT conversacional</i>) utilizando $\alpha_i = 0.5$	230
Figura D.23. Evolución temporal de MOS instantáneo, MOS promediado y territorio para <i>streaming</i> de vídeo (<i>CT streaming</i>) utilizando $\alpha_i = 0.5$	230
Figura D.24. Evolución temporal de MOS instantáneo, MOS promediado y territorio para <i>streaming</i> de vídeo sobre TCP (<i>CT TCP de tiempo rea</i>) utilizando $\alpha_i = 0.5$	231
Figura D.25. Evolución temporal de MOS instantáneo, MOS promediado y territorio para navegación web (<i>CT interactiva</i>) utilizando $\alpha_i = 0.5$	231
Figura D.26. Evolución temporal de MOS instantáneo, MOS promediado y territorio para FTP (CT de fondo) utilizando $\alpha_i = 0.5$	231
Figura D.27. Evolución temporal de MOS instantáneo, MOS promediado y territorio para VoIP (<i>CT conversacional</i>) utilizando $\alpha_i = 0.9$	232
Figura D.28. Evolución temporal de MOS instantáneo, MOS promediado y territorio para <i>streaming</i> de vídeo (<i>CT streaming</i>) utilizando $\alpha_i = 0.9$	232
Figura D.29. Evolución temporal de MOS instantáneo, MOS promediado y territorio para <i>streaming</i> de vídeo sobre TCP (<i>CT TCP de tiempo rea</i>) utilizando $\alpha_i = 0.9$	232
Figura D.30. Evolución temporal de MOS instantáneo, MOS promediado y territorio para navegación web (<i>CT interactiva</i>) utilizando $\alpha_i = 0.9$	232
Figura D.31. Evolución temporal de MOS instantáneo, MOS promediado y territorio para FTP (CT de fondo) utilizando $\alpha_i = 0.9$	233

Índice de tablas

Tabla 2.1. Resumen de los estándares y correcciones IEEE 802.11.....	13
Tabla 2.2. Resumen de los estándares y correcciones IEEE 802.11 (cont.).....	14
Tabla 2.3. Parámetros temporales de los niveles físicos de IEEE 802.11-2007.....	29
Tabla 2.4. Parámetros por defecto para EDCA utilizando la capa física DSSS de 802.11b.....	32
Tabla 3.1. Escala MOS.....	43
Tabla 3.2. Parámetros del Modelo E para entornos de interior con fuentes usando el códec G.711.....	46
Tabla 3.3. Definición de los parámetros a_i para streaming de vídeo.....	49
Tabla 4.1. Parámetros de la capa física inalámbrica.....	70
Table A.1. E-Model parameters for indoor with G.711 codec.....	184
Table A.2. a_i definitions for video streaming MOS.....	185

Capítulo 1

Introducción

1.1 Motivación

Desde que James Clerk Maxwell describiese por primera vez las ondas electromagnéticas de manera formal hace siglo y medio, las comunicaciones inalámbricas –sin cables– han evolucionado de forma notable, especialmente en las últimas décadas. El término *inalámbrico* apareció por primera vez cuando, en 1896, Guglielmo Marconi desarrolló el primer sistema telegráfico sin cables (más tarde, en 1909, compartiría el premio Nobel de Física con Karl Ferdinand Braun por su contribución a las radiocomunicaciones).

La aplicación más importante que ha fomentado el desarrollo de las tecnologías inalámbricas es la comunicación por voz, con una demanda creciente en el mercado de masas: la moda de los radioaficionados, el mercado de los *walkie-talkies*, la interconexión entre sistemas radio y las redes de telefonía y la increíble proliferación de teléfonos móviles conectados mediante sistemas celulares.

Sin embargo, las comunicaciones inalámbricas no se limitan al servicio de voz sino

2 | Introducción

que también incluyen la transmisión de datos, principalmente gracias al desarrollo de Internet. Desde sus orígenes, Internet ha crecido vertiginosamente tanto en número de usuarios como en prestaciones y servicios ofrecidos. Éstos no sólo se circunscriben a la navegación web y al uso de correo electrónico, sino que han aparecido nuevas aplicaciones con unos requisitos no contemplados en la concepción inicial de la Red, como pueden ser la difusión de vídeo (e.g. IPTV, videoconferencia), el *streaming*¹ de audio (e.g. música bajo demanda) o nuevamente las comunicaciones de voz (VoIP). Este crecimiento en las comunicaciones de datos ha influido de forma decisiva en el desarrollo de las tecnologías inalámbricas.

En la actualidad existen diferentes tecnologías inalámbricas que permiten la transmisión de datos de alta velocidad, e.g. Ultra-Wide Band, WiMax, HSPA o LTE. En los entornos de área local, las redes IEEE 802.11 [1] (tecnología también conocida como WLAN –red inalámbrica de área local– ó Wi-Fi²) son las más extendidas, permitiendo que usuarios equipados con un PC portátil, una PDA o incluso un teléfono móvil se puedan conectar de forma inalámbrica a Internet (o a otras redes de datos como intranets corporativas). Estas redes están desplegadas en diferentes escenarios como oficinas, *hotspots* en aeropuertos, hoteles, cafeterías, en el hogar, etcétera.

Al igual que Internet, el estándar IEEE 802.11 no fue concebido para soportar requisitos de Calidad de Servicio (QoS, *Quality of Service*) como los requeridos por los servicios multimedia, sino que se basa en un esquema de *mejor esfuerzo*. El estándar original incluyó dos mecanismos de acceso al medio, DCF (*Distributed Coordination Function*) –modo distribuido de implementación obligatoria– y PCF (*Point Coordination Function*) –esquema opcional de acceso coordinado–. PCF fue propuesto para soportar servicios con requisitos estrictos de tiempo, pero no puede garantizar tasas de transmisión

¹ El término *streaming* apareció en Abril de 1995 con el lanzamiento de RealAudio 1.0. Antes de su aparición, la reproducción de contenido multimedia a través de Internet conllevaba necesariamente la descarga del archivo (normalmente de gran tamaño) al disco duro local.

² Wi-Fi™ es una marca que utilizan los fabricantes como sello de los productos certificados por la *Wi-Fi Alliance* basados en los estándares IEEE 802.11.

o retardos máximos ya que el tiempo de transmisión de una estación móvil no está acotado.

Si bien existen múltiples soluciones basadas en la norma 802.11e [2] –e.g. [3] [4] [5] [6] [7] [8] [9] [10] [11] [12] [13]– para transportar tráfico de tiempo real en este tipo de redes con las garantías requeridas, ninguna de ellas –hasta donde llega el conocimiento del autor– es capaz de soportar simultánea y eficazmente tanto servicios de tiempo real como de tráfico elástico³ (e.g. aplicaciones basadas en TCP), mezcla habitual en Internet. Éste será uno de los objetivos fundamentales del trabajo desarrollado en esta tesis.

Por otro lado, los trabajos de diferentes autores [14] [15] [16] [17] [18] [19] hacen pensar que la incorporación de la Calidad de Experiencia (QoE, *Quality of Experience*) – grado de satisfacción de un usuario final basado en su percepción subjetiva– en el diseño de algoritmos y procedimientos para la gestión de los recursos introducirá una mejora significativa en el rendimiento del sistema. Esta hipótesis inicial será la base de algunas de las principales contribuciones desarrolladas.

1.2 Objetivos

El objetivo general de la tesis es proporcionar nuevos mecanismos que tengan como objeto mejorar la QoE experimentada por el usuario final en un entorno inalámbrico de área local –basado en redes IEEE 802.11–. En concreto, los mecanismos a diseñar deben considerar escenarios cuyo tráfico ha sido generado por una mezcla de servicios con diferentes requisitos, como aplicaciones de tiempo real y de tráfico elástico.

³ El tráfico elástico es aquel que puede ajustarse a cambios en el retardo y rendimiento a través de un conjunto de redes, satisfaciendo las necesidades de sus aplicaciones. Este es el tráfico tradicional generado por servicios que utilizan TCP como protocolo de transporte.

4 | Introducción

Los objetivos particulares de esta tesis se pueden desglosar en los siguientes puntos:

- La provisión de medidas de calidad subjetiva de los servicios más relevantes en las redes inalámbricas, que sirvan para evaluar su rendimiento y como criterio para su diseño.
- El desarrollo de mecanismos que permitan a una red inalámbrica garantizar requisitos estrictos de calidad de servicio.
- El diseño de algoritmos y procedimientos de gestión de recursos para mejorar la calidad experimentada por los usuarios en un entorno con tráfico heterogéneo.
- La mejora del rendimiento de las aplicaciones basadas en TCP para entornos inalámbricos, consiguiendo un mejor aprovechamiento de la capacidad del medio físico.

1.3 Contribuciones de la tesis

Del cumplimiento de los objetivos anteriores se han generado las siguientes contribuciones, algunas de ellas parcialmente difundidas en las publicaciones científicas que se detallan en cada punto:

- Definición de una métrica común, la Nota Media de Opinión (MOS, *Mean Opinion Score*), para evaluar el rendimiento y tomar decisiones respecto a la asignación de recursos en escenarios con mezcla de tráficos. Ésta es la metodología –basada en la QoE– considerada en las propuestas desarrolladas en el presente trabajo, siendo utilizada en [20] y [21].
- Obtención de funciones de utilidad que mapean métricas objetivas del rendimiento de los servicios –e.g. tasa binaria y retardo– en métricas subjetivas –percepción del usuario–. Algunas de estas funciones han sido extraídas de la literatura científica existente [22] [23] [24], mientras que otras han sido obtenidas durante el desarrollo del trabajo. En concreto, la función de utilidad para la

1.3. Contribuciones de la tesis | 5

navegación web fue deducida en [25] y [16], mientras que la utilizada para tráfico de vídeo sobre TCP aparece en [21].

- Diseño y evaluación de de un planificador de paquetes que garantiza un ancho de banda mínimo y un retardo máximo acotado para entornos inalámbricos a través del mecanismo HCCA. Este planificador –variante del *Monolithic Shaper-Scheduler*– fue desarrollado en [20] y se utiliza también en [21].
- Diseño y evaluación de una arquitectura general, el *Método de los Territorios*, para la compartición de recursos con decisiones basadas en QoE y que garantiza el cumplimiento de requisitos estrictos de tiempo, a la vez que permite el soporte de diferentes tipos de servicios (clases de tráfico). Este mecanismo fue propuesto inicialmente en [20] y mejorado en [21].
- Diseño y evaluación de políticas, aplicables a la arquitectura anterior, con diferentes criterios que permiten al administrador gestionar su red de acuerdo a sus intereses. En concreto, estas políticas permiten:
 - Priorizar flujos según su tipo de servicio, de forma que el tráfico de tiempo real no vea su rendimiento influenciado por clases de tráfico menos prioritarias.
 - Asignar recursos en función de las necesidades actuales del sistema, sin considerar prioridades entre las diferentes aplicaciones.
 - Igualar la calidad experimentada por los usuarios de diferentes servicios, de forma que la gestión sea justa en términos de QoE.
 - Mejorar los ingresos obtenidos por el gestor de la red, utilizando el beneficio obtenido como criterio de diferenciación entre servicios.

Estas políticas fueron diseñadas y evaluadas en [21], comprobándose la flexibilidad y el comportamiento adecuado de la plataforma utilizada.

- Diseño de una modificación del protocolo TCP para redes inalámbricas provistas de HCCA, permitiendo la mejora del rendimiento de las aplicaciones que utilizan este protocolo para su transporte. Esta mejora se fundamenta en la eliminación de

6 | Introducción

funcionalidades redundantes entre este protocolo de transporte y el nivel MAC –control de acceso al medio– de las redes IEEE 802.11e.

Algunas de estas contribuciones han formado parte de los resultados del Proyecto QUARTET (referencia TIN2009-13992-C02-02) del Ministerio de Ciencia e Innovación de España.

1.4 Organización del documento

Esta tesis se estructura en 9 capítulos. El Capítulo 1, *Introducción*, es el capítulo actual en el que se recoge el contexto y la motivación de la investigación documentada en esta tesis, sus principales contribuciones y se presenta la estructura del documento.

El Capítulo 2, *Redes WLAN*, expone de forma detallada el funcionamiento de las redes basadas en el estándar IEEE 802.11, incluyendo los elementos integrantes, las diferentes topologías y la funcionalidad por capas de su arquitectura de protocolos. En este último aspecto se hará especial énfasis, ya que la capa MAC proporciona los servicios que permitirán garantizar el cumplimiento de requisitos de QoS. Este capítulo incluye también una revisión del estado del arte de los aspectos y mecanismos relacionados con el tema de la tesis.

En el Capítulo 3, *Calidad de Experiencia (QoE)*, se definen los conceptos de QoS y QoE, se propone una métrica común para comparar el rendimiento de diferentes servicios y se describe una metodología de trabajo basada en medidas de percepción subjetiva. Para modelar y obtener el grado de satisfacción de los usuarios se utilizan funciones de utilidad que establecen una correspondencia entre algunos indicadores objetivos y medibles en la red –como la velocidad de transmisión o el retardo– y la opinión del usuario. Algunas de estas funciones de utilidad han sido extraídas de la bibliografía existente –para voz sobre IP (VoIP), difusión de vídeo y descarga de ficheros– mientras que otras han sido elaboradas durante el desarrollo del presente trabajo y son contribuciones originales –para navegación web y vídeo sobre TCP–.

El Capítulo 4, *Una variante del Planificador–Conformador Monolítico para HCCA*,

1.4. Organización del documento | 7

expone el diseño y evaluación de una propuesta basada en el uso de un planificador de paquetes incorporado al punto de acceso. Esta solución cumple los requisitos impuestos por el estándar IEEE 802.11e y permite gestionar tanto tráfico de tasa constante (CBR) como de tasa variable (VBR) de forma individual.

Para mejorar la eficacia en el soporte de mezclas de tráfico se ha diseñado la propuesta incluida en el Capítulo 5, *Gestión de recursos para servicios heterogéneos: Método de los Territorios*. Esta arquitectura está concebida para priorizar flujos en función de la Clase de Tráfico a la que pertenecen, a la vez que administra los recursos según la calidad experimentada por sus usuarios permitiendo la coexistencia de diferentes tipos de servicios.

El Capítulo 6, *Políticas de gestión de recursos basadas en Calidad de Experiencia*, expone diferentes políticas –basadas en el Método de los Territorios– que permiten al administrador gestionar su red de acuerdo a sus intereses particulares. Así, se han diseñado esquemas basados en prioridades, sin prioridades, en la justicia entre servicios –respecto a la calidad experimentada– y en el aumento de los ingresos.

El Capítulo 7, *TCP mejorado sobre HCCA*, incluye una modificación del protocolo TCP en el medio inalámbrico a través de la eliminación de funcionalidades redundantes –ya proporcionadas por los mecanismos utilizados en el acceso al medio inalámbrico–. Su objetivo es ampliar el ancho de banda efectivo que pueden utilizar las aplicaciones basadas en este protocolo de transporte.

Por último, en los Capítulos 8 y 9 se exponen las conclusiones finales de la tesis, en español y en inglés respectivamente, tal y como exige la normativa vigente de la Universidad de Granada sobre *Tesis con Mención Europea*. A continuación de las conclusiones se adjunta un resumen en inglés de las principales contribuciones de la tesis (Apéndice A), se incluyen los apéndices referidos en los distintos capítulos y finalmente se identifica la bibliografía utilizada.

Capítulo 2

Redes WLAN

En el momento actual, la familia de protocolos de acceso al medio inalámbrico dominante para entornos de interior es IEEE 802.11, también conocida como WLAN (*Wireless Local Area Network*). La primera versión del estándar fue publicada en 1997 [1]. Conforme pasó el tiempo, mientras que la capa MAC original se mantuvo intacta, la tecnología continuó su evolución mediante nuevas especificaciones de la capa física con mayores velocidades de transmisión, apareciendo variantes al estándar original (802.11a, b, g, y recientemente 802.11n). WLAN ganó una inmensa popularidad debido a su bajo coste y fácil despliegue, siendo en la actualidad una de las tecnologías inalámbricas con mayor penetración.

Sin embargo, el estándar IEEE 802.11 no fue diseñado para soportar servicios con las garantías necesarias de Calidad de Servicio. Para solventar esta falta de soporte de QoS, la corrección 802.11e [2] introdujo nuevos mecanismos de acceso al canal soportando, por un lado, la diferenciación del tráfico mediante prioridades relativas y, por otro lado, requisitos estrictos de tiempo real.

En el presente trabajo se descartaron las soluciones basadas en la priorización de tráfico (mecanismo de acceso distribuido, véase la Sección 2.3.1), ya que no garantizan el cumplimiento de requisitos tales como la acotación del retardo máximo de los paquetes o

garantías de tasa mínima de transmisión. No obstante, el mecanismo de acceso centralizado (véase la Sección 2.3.2) sí permite este tipo de garantías siendo, por tanto, el mecanismo elegido. La literatura actual incluye numerosas soluciones basadas en éste último, pero, como se comentará más adelante, ninguna soporta de forma eficiente y simultánea los diferentes tipos de servicio existentes en las redes actuales, i.e. servicios con requisitos estrictos de tiempo (e.g. VoIP, *streaming* de vídeo/audio) y servicios de tráfico elástico (e.g. basados en TCP, como navegación web y descarga de ficheros).

Este capítulo, por tanto, tiene como objetivo fijar el contexto y los antecedentes relacionados con el soporte de QoS en redes WLAN con tráfico heterogéneo.

Para ello, el capítulo se ha organizado en las siguientes secciones. La Sección 2.1 presenta los fundamentos del estándar original IEEE 802.11-1997, tecnología que define el entorno para el desarrollo de este trabajo. En la Sección 2.3 se describen las mejoras introducidas por la variante 802.11e para soportar Calidad de Servicio (actualmente agrupada en el estándar IEEE 802.11-2007 [26]). La Sección 2.4 incluye un análisis de las principales referencias bibliográficas relacionadas con la problemática abordada. Por último, la Sección 2.5 resume las conclusiones extraídas.

2.1 Estándar IEEE 802.11

El estándar IEEE 802.11 [1] incluye la especificación de la capa física del modelo de referencia OSI (*Open System Interconnection*) y de la subcapa de Control de Acceso al Medio (MAC, *Medium Access Control*). El nivel físico realiza la transmisión de bits a través del medio inalámbrico, encargándose de la modulación y codificación, sintonización de canales, etcétera. Además, interactúa con la capa MAC (parte inferior del nivel de enlace), que gestiona el direccionamiento y el acceso múltiple a un medio compartido. MAC también incluye las funcionalidades de fragmentación, encriptación, sincronización, gestión de potencia y tránsito entre redes. El funcionamiento de estas capas está descrito en las Secciones 2.2.1 y 2.2.2, respectivamente.

El Control de Enlace Lógico (LLC, *Logical Link Control*) utilizado es el descrito por el estándar IEEE 802.2, actuando de interfaz entre la subcapa MAC y la capa de red. LLC

es responsable de la multiplexación de protocolos transmitidos sobre el nivel MAC, así como de proporcionar control de errores y de flujo.

Al utilizar la subcapa LLC definida por IEEE 802.2, esta arquitectura (representada en la Figura 2.1) proporciona:

- a) Un interfaz transparente a las capas superiores: las estaciones se pueden mover dentro de una red WLAN y transitar entre diferentes redes inalámbricas, y siguen pareciendo estacionarias para la subcapa LLC y capas superiores.
- b) Compatibilidad con la arquitectura de protocolos TCP/IP, que se ejecutan sobre IEEE 802.11 de forma similar que sobre las redes cableadas Ethernet.

En 1997, IEEE proporcionó tres tipos de capa física para las redes inalámbricas de área local (WLAN) a través del estándar 802.11. Una capa utiliza la transmisión mediante infrarrojos en banda base (IR, *InfraRed baseband*). Las otras dos capas están basadas en la transmisión de ondas de radiofrecuencia en la banda ISM (*Industrial, Scientific and Medical*) a 2.4GHz, en concreto mediante espectro expandido por salto de frecuencias (FHSS, *Frequency Hopping Spread Spectrum*) y espectro expandido por secuencia directa (DSSS, *Direct Sequence Spread Spectrum*). Todas estas opciones proporcionan tasas binarias de 1 y 2 Mbps. El estándar original se conoce como IEEE 802.11 *legacy* o IEEE 802.11-1997, versión revisada en 1999 (IEEE 802.11-1999). En ese mismo año, IEEE definió dos extensiones de mayor velocidad: 802.11b en la banda de 2.4GHz con tasas de hasta 11 Mbps, basada en la tecnología DSSS y 802.11a en la banda U-NII

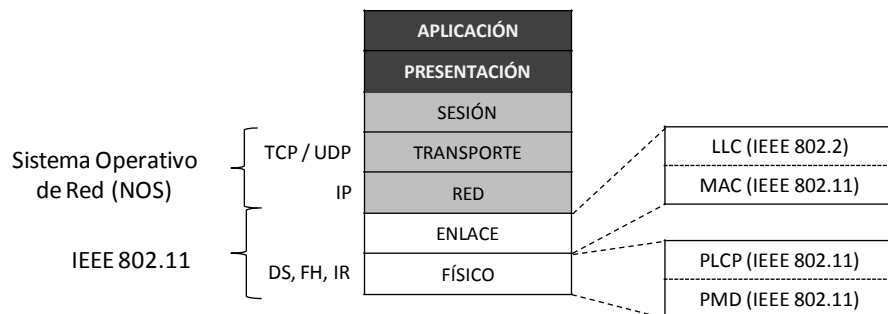


Figura 2.1. Arquitectura de protocolos en el estándar IEEE 802.11.

(*Unlicensed National Information Infrastructure*) de 5GHz con tasas de hasta 54Mbps, basadas en multiplexación por división en frecuencias ortogonales (OFDM, *Orthogonal Frequency Division Multiplexing*). En 2003 se especificó la extensión 802.11g que mejoraba las tasas de 802.11b llegando hasta 54Mbps, gracias a la adopción de OFDM (al igual que 802.11a) en la banda de 2.4GHz. Más tarde, en 2007, se aprobó el agrupamiento de muchas de estas correcciones (*amendments*) con la versión de 1999 del estándar, creándose así el estándar IEEE 802.11-2007. En concreto, las correcciones incluidas fueron 802.11a, b, d, e, g, h, i y j.

Más recientemente, en 2009, se definió la modificación del estándar 802.11n con el objetivo de mejorar las velocidades alcanzadas por el estándar IEEE 802.11-2007. Este nuevo estándar propone, entre otras mejoras, la inclusión de múltiples antenas tanto en emisión como en recepción, i.e. antenas MIMO (*Multiple-Input, Multiple-Output*), y el aumento del ancho de banda de los canales de 20 a 40 MHz. Utilizando antenas MIMO 4 x 4 : 4 (4 antenas transmisoras, 4 antenas receptoras y 4 flujos de datos) con canales de 40 MHz se pueden alcanzar tasas físicas de hasta 600 Mbps.

Respecto a la capa MAC, 802.11e (2005) fue el primer estándar complementario para mejorar el soporte de calidad de servicio de las redes IEEE 802.11. Este estándar será descrito detalladamente en la Sección 2.3, ya que es la norma sobre la que se desarrolla esta tesis.

Otro conjunto de estándares están relacionados con la transición entre puntos de acceso pertenecientes a una misma WLAN. 802.11F define un protocolo entre puntos de acceso para permitir el tránsito de estaciones entre puntos de acceso de diferentes fabricantes. Este estándar fue aprobado en 2003 pero fue retirado en 2006, quedando como prácticas recomendadas (de ahí que se escriba su letra en mayúscula). Otros estándares sobre la transición entre puntos de acceso son 802.11k (2008) y 802.11r (2008). El primero es una corrección al estándar 802.11-2007 para la gestión de recursos radio (*RRM, Radio Resource Management*), permitiendo distribuir el tráfico entre puntos de acceso de forma inteligente (anteriormente las estaciones escogían el punto de acceso basándose exclusivamente en su SNR (Relación Señal-Ruido)). El segundo, permite aumentar la velocidad en las transiciones entre puntos de acceso gracias a la eliminación de algunas etapas (derivación de claves PTK (*pairwise transient key*) y control de admisión para QoS) y a un mecanismo de pre-reserva antes de la transición.

2.1. Estándar IEEE 802.11 | 13

Respecto a la seguridad, 802.11i mejora los mecanismos de seguridad y autenticación para la capa MAC de 802.11, reemplazando el esquema de seguridad original (WEP, *Wired Equivalent Privacy*) por los protocolos WPA (*Wi-Fi Protected Access*, implementación de un subconjunto de funcionalidades de 802.11i) y WPA2 (implementación completa de 802.11i).

Las Tablas 2.1 y 2.2 muestran un resumen de los estándares y correcciones IEEE 802.11 [27].

Estándar o corrección	Descripción
IEEE 802.11	El estándar WLAN original con tasas físicas de 1 y 2Mbps, en la banda RF de 2.4GHz (FHSS y DSSS) e infrarrojo (IR) (1997). Todas las especificaciones técnicas listadas debajo son correcciones (<i>amendments</i>) a este estándar, salvo las prácticas recomendadas 802.11F y 802.11T.
IEEE 802.11a	54Mbps, 5 GHz (1999)
IEEE 802.11b	Mejoras a 802.11 para proporcionar 5.5 y 11Mbps (1999)
IEEE 802.11c	Procedimientos de operación como puente; incluido en el estándar 802.1D (2001)
IEEE 802.11d	Extensiones para transición internacional (2001)
IEEE 802.11e	Mejoras de Calidad de Servicio (QoS) (2005)
IEEE 802.11F	Protocolo entre Puntos de Acceso (2003, retirado en 2006)
IEEE 802.11g	54Mbps, 2.4 GHz (compatible con 802.11b) (2003)
IEEE 802.11h	Gestión de espectro 802.11a (5GHz) para compatibilidad europea (2004)
IEEE 802.11i	Mejoras de seguridad (2004)
IEEE 802.11j	Extensiones para Japón (2004)
IEEE 802.11-2007	Versión nueva del estándar que incluye las mejoras a, b, d, e, g, h, i y j (2007)
IEEE 802.11k	Mejoras en la gestión de recursos radio (2008)
IEEE 802.11n	Aumento de la tasa binaria utilizando antenas MIMO (2009)
IEEE 802.11p	Acceso inalámbrico para entornos vehiculares (WAVE – <i>Wireless Access for the Vehicular Environment</i>) (previsto para Junio de 2010)
IEEE 802.11r	Transición rápida (2008)
IEEE 802.11s	Redes malladas (<i>mesh networking</i>), conjunto de servicios extendidos (ESS, <i>Extended Service Set</i>) (previsto para Septiembre de 2010)

Tabla 2.1. Resumen de los estándares y correcciones IEEE 802.11.

IEEE 802.11T	Predicción del rendimiento inalámbrico (WPP, <i>Wireless Performance Prediction</i>) – métodos de test y métricas (cancelado)
IEEE 802.11u	Interoperabilidad con redes no-802 (e.g. redes celulares) (previsto para Septiembre de 2010)
IEEE 802.11v	Gestión de red inalámbrica (previsto para Junio de 2010)
IEEE 802.11w	Seguridad para las tramas de gestión (2009)
IEEE 802.11y	Operación en E.E.U.U. (3650 – 3700MHz) (2008)
IEEE 802.11z	Extensiones para el establecimiento de enlace directo (DLS, <i>Direct Link Setup</i>) (Agosto de 2007 – Diciembre de 2011)
IEEE 802.11aa	Flujos de transporte de vídeo (Marzo de 2008 – Octubre de 2011)
IEEE 802.11mb	Mantenimiento del estándar; se convertirá en 802.11-2011 (publicación prevista para Junio de 2011)
IEEE 802.11ac	Tasa binaria muy alta (< 6GHz); mejoras potenciales sobre 802.11n: mejor esquema de modulación; canales más anchos (80 ó 160 MHz), MIMO multiusuario (Septiembre de 2008 – Diciembre de 2012)
IEEE 802.11ad	Tasa binaria muy alta (60GHz) (Diciembre de 2008 – Diciembre de 2012)
IEEE 802.11ae	Gestión de Calidad de Servicio (QoS) (Septiembre de 2009 – Junio de 2012)
IEEE 802.11af	Espectro no utilizado en difusión de TV (Septiembre de 2009 – Diciembre de 2012)

Tabla 2.2. Resumen de los estándares y correcciones IEEE 802.11 (cont.).

2.2 Arquitectura

Desde un punto de vista topológico, una red IEEE 802.11 se puede descomponer en cuatro tipos de componentes fundamentales (véase la Figura 2.2):

- **Estación** (STA, *Station*): dispositivo con interfaz de red inalámbrica, origen y/o destino de las transferencias de datos.
- **Punto de acceso** (AP, *Access Point*): estación especial que permite la conexión a la red cableada por parte de las estaciones asociadas.



Figura 2.2. Elementos de una red 802.11.

- **Medio inalámbrico** (WM, *Wireless Medium*): medio en el que se realizan las comunicaciones.
- **Sistema distribuido** (DS, *Distributed System*): red utilizada para la comunicación entre puntos de acceso.

Las estaciones (incluyendo los APs) utilizan los servicios definidos por el estándar para poder comunicarse entre ellas. Éstos se agrupan en dos tipos de conjuntos de servicios, que permiten definir las topologías posibles en las redes WLAN.

Por un lado, el **conjunto de servicios básicos** (**BSS**, *Basic Service Set*) se define como el conjunto de servicios que permiten que un grupo de estaciones se puedan comunicar entre ellas directamente. Estas comunicaciones tienen lugar en el área de servicio básico (**BSA**, *Basic Service Area*). La utilización del BSS da lugar a dos topologías básicas de redes inalámbricas: las redes independientes o *ad hoc*, y las redes de infraestructura.

Las **redes independientes** están formadas por estaciones que se comunican directamente, por lo que han de estar dentro de su área de cobertura (véase la Figura 2.3). Suelen ser de tamaño pequeño y de duración corta (e.g. para compartir datos en una reunión), por lo que también se les denomina redes **ad hoc**.

Las **redes de infraestructura** utilizan un punto de acceso para sus comunicaciones, incluso entre estaciones que estén mutuamente dentro de su rango de cobertura (véase la Figura 2.4). De esta forma, las comunicaciones entre dos estaciones se realizan en dos saltos: primero entre la estación origen y el AP, y después entre el AP y la estación destino. En este tipo de redes, las estaciones deben *asociarse* al punto de acceso. El proceso de asociación siempre lo inicia la estación, y el AP aceptará o rechazará la solicitud. No hay ningún límite teórico acerca del número de estaciones que se pueden asociar a un mismo AP, si bien existirá un límite práctico debido a los recursos disponibles.

Las redes de infraestructura permiten aumentar la cobertura del área de servicio, ya que se pueden comunicar estaciones que no son accesibles directamente (siempre que cada una esté en el área de cobertura del AP). Además, las estaciones pueden entrar en modo de ahorro de energía, ya que el AP almacenará sus paquetes para enviarlos cuando

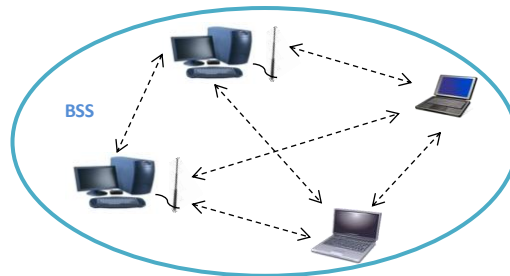


Figura 2.3. Topología de una red independiente o *ad hoc*.

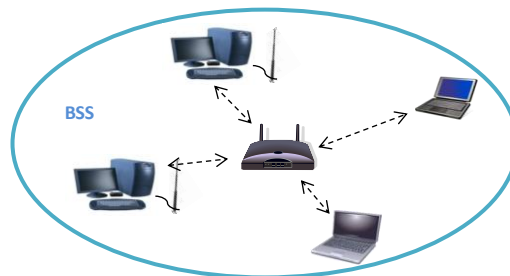


Figura 2.4. Topología de una red de infraestructura en una BSA.

salgan de dicho modo. Sin embargo, consumen más recursos radio ya que se duplica el número de tramas transmitidas por el doble salto comentado.

Por otro lado, el segundo grupo de servicios se denomina *conjunto de servicios extendidos* (**ESS**, *Extended Service Set*). Éstos permiten crear redes con una cobertura mayor, dado que permite enlazar diferentes puntos de acceso. Los APs se comunicarán a través de una red, constituyéndose así el sistema distribuido. Esta red suele ser de tipo Ethernet (IEEE 802.3), si bien el estándar sólo define los servicios requeridos y no la tecnología a utilizar, pudiendo ser incluso una red WLAN. En este caso, al realizarse las comunicaciones a través de puntos de acceso, la topología siempre será de red de infraestructura (véase la Figura 2.5).

2.2.1 Capa física

El borrador inicial del estándar IEEE 802.11 especificó tres implementaciones diferentes para la capa física: espectro expandido por salto de frecuencias (FHSS), espectro expandido por secuencia directa (DSSS) e infrarrojo (IR).

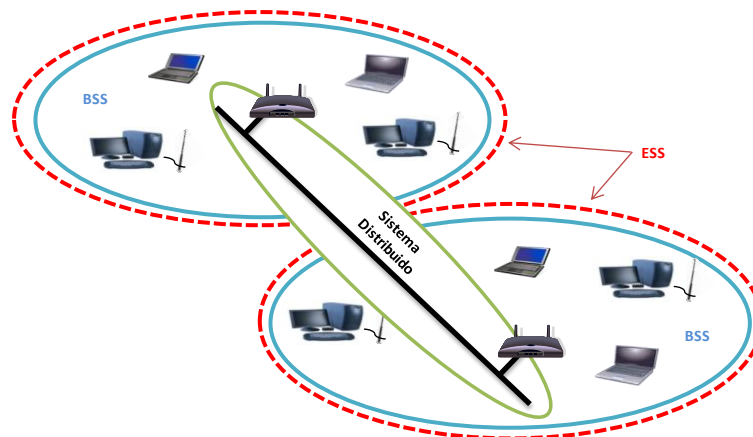


Figura 2.5. Red de infraestructura con varias BSAs.

El primer tipo de capa física, FHSS, utiliza la banda ISM (*Industrial, Scientific and Medical*) en 2.4GHz (2.4000 – 2.4835 GHz). En Europa se especifica un máximo de 79 canales (pueden variar dependiendo del país, definiéndose en 802.11 y 802.11d), estando el primero centrado en 2.402GHz (canal 2) y los consecutivos separados entre sí 1MHz. Se establecen tres conjuntos de secuencias de salto, con 26 secuencias por conjunto, estableciéndose un mínimo de 2.5 saltos por segundo. Gracias al uso de diferentes secuencias, varias BSSs pueden coexistir en la misma área geográfica. La tasa de acceso básico de 1Mbps utiliza la modulación GFSK (Modulación por Desplazamiento de Frecuencias con filtro Gaussiano) de dos niveles. La tasa de acceso mejorado de 2 Mbps utiliza GFSK de 4 niveles, i.e. se codifican dos bits por símbolo a transmitir.

DSSS también utiliza la banda ISM en 2.4GHz, con las modulaciones DBPSK (desplazamiento de fase binaria diferencial) y DQPSK (desplazamiento de fase en cuadratura diferencial) para las tasas de acceso básico y mejorado respectivamente. En este caso, el rango de frecuencias disponibles (2.4000 – 2.4835GHz) se divide en 13 canales (en Europa, 11 en E.E.U.U. y 14 en Japón) con un ancho de banda de 22MHz (lóbulo principal). La separación entre las frecuencias centrales de los canales es de 5MHz, estando el primer canal centrado en 2.412GHz. El ensanchamiento se lleva a cabo mediante una secuencia de expansión Baker de 11 *chips* (símbolos del código pseudoaleatorio utilizado para la expansión) generados con un reloj de 11MHz. Al existir solapamiento entre canales adyacentes, las BSSs que se solapen geográficamente deberían utilizar canales cuyas frecuencias centrales tuviesen una separación de al menos 22MHz, i.e. 5 canales. Este requisito hace que sólo sea posible utilizar tres frecuencias entre BSSs adyacentes o solapadas si no se quiere tener interferencias entre ellas (se recomienda el uso de los canales 1, 6 y 11 en E.E.U.U y los canales 1, 5, 9 y 13 en Europa, con un pequeño solapamiento en este último caso).

La especificación de IR identifica el rango de longitudes de onda entre 850 y 950nm. La banda IR está diseñada únicamente para su uso en interiores y permite tanto transmisiones directas como no directas, i.e. por medio de reflexiones. La codificación del acceso básico a 1Mbps se realiza mediante una Modulación de Posición de Pulsos (PPM) en la que se mapean 4 bits de datos en 16 bits codificados para la transmisión (16-PPM). El acceso mejorado se realiza usando 2 bits de datos por cada 4 bits codificados (4-PPM).

La corrección **802.11a** opera en la banda U-NII en 5GHz. Los canales utilizados dependen de las regulaciones específicas del país, pero en Europa (según el Anexo J de IEEE 802.11-2007) son los canales 36, 40, 44, 48, 52, 56, 60, 64, 100, 104, 108, 112, 116, 120, 124, 128, 132, 136 y 140, definiéndose éstos como los centrados en las frecuencias $5\text{GHz} + x \cdot 5\text{MHz}$, con x igual al número de canal. Debido al uso de la banda de 5GHz, el rango de cobertura es inferior al de las variantes en 2.4GHz. Sin embargo, dado que esta banda suele estar menos utilizada, las interferencias son menores por lo que la cobertura real suele ser ligeramente inferior. Este estándar mejora la velocidad de transmisión (hasta 54Mbps) del nivel físico gracias a la utilización de OFDM, dividiendo el canal en 52 subportadoras con un ancho de banda de 0.3125 MHz (20MHz / 64). De éstas, 48 son utilizadas para transmitir datos y 4 son subportadoras piloto. Cada subportadora puede utilizar las modulaciones BPSK, QPSK, 16-QAM o 64-QAM con una duración de símbolo de 4 microsegundos (incluyendo un intervalo de guarda de 0.8 microsegundos). Según la modulación utilizada y la redundancia introducida (FEC, *Forward Error Correction*) la tasa binaria se puede reducir a 48, 36, 24, 18, 12, 9 y 6Mbps, permitiéndose así la adaptación a las condiciones variables del medio inalámbrico.

La variante **802.11b** aumenta la velocidad de transmisión respecto al estándar original hasta 11Mbps utilizando la misma banda de 2.4GHz. La principal novedad es la incorporación de CCK (*Complementary Code Keying*) como técnica de modulación para las nuevas tasas binarias de 5.5 y 11Mbps (codificando 4 y 8 bits en símbolos de 8 chips, respectivamente), continuando con DSSS para las velocidades inferiores de 1 y 2 Mbps (compatible con dispositivos 802.11). Este estándar también permite el uso de PBCC (*Packet Binary Convolutional Coding*) como opcional.

El tercer estándar complementario sobre modulación fue **802.11g**, también empleado en la banda de 2.4 GHz y compatible con 802.11b. En este caso, las nuevas tasas binarias (6, 9, 12, 18, 24, 36, 48 y 54Mbps) utilizan OFDM (de forma similar a 802.11a), manteniéndose los modos anteriores (CCK para velocidades de 5.5 y 11Mbps, y DBPSK/DQPSK + DSSS para 1 y 2Mbps).

IEEE 802.11n se basa en los estándares anteriores añadiendo antenas MIMO (Múltiple Entrada Múltiple Salida, *Multiple-Input Multiple-Output*) y canales de 40 MHz en la capa física, y la agregación de tramas en la capa MAC.

MIMO es una tecnología que utiliza múltiples antenas para enviar más información que usando una antena. Para ello, utiliza multiplexación por división espacial (SDM, *Spatial Division Multiplexing*), que consiste en la transmisión a través de las diferentes antenas de varios flujos de datos independientes usando el mismo canal. De esta forma, MIMO SDM puede mejorar las velocidades de transmisión de forma significativa conforme aumenta el número de flujos. Cada flujo requiere de una antena discreta tanto en el emisor como en el receptor así como su circuitería de RF, aumentando así su coste de implementación.

Por otro lado, el uso de canales de 40 MHz duplica el ancho de banda de las capas físicas 802.11 anteriores, permitiendo duplicar la tasa binaria en esta capa. Se puede utilizar tanto en la banda de 5 GHz como en la de 2.4GHz.

Además, el transmisor y el receptor utilizan técnicas de precodificación y postcodificación para optimizar la capacidad del enlace MIMO. La precodificación incluye la creación de haces direccionales (*spatial beamforming*) y la codificación espacial, mejorándose la calidad de la señal recibida y aumentando la tasa de transmisión al usar diversidad espacial, respectivamente.

El número de flujos de datos simultáneos está limitado por el mínimo número de antenas a ambos lados del enlace. La notación $a \times b : c$ ayuda a identificar cuáles son las características de transmisión. El primer número (a) es el número máximo de antenas en el transmisor. El segundo número (b) es el número máximo de antenas en el receptor. El tercer número (c) es el número máximo de flujos *espaciales* de datos que se pueden utilizar. El estándar 802.11n permite hasta $4 \times 4 : 4$, siendo algunas configuraciones típicas $2 \times 2 : 2$, $2 \times 3 : 2$, $3 \times 3 : 2$ y últimamente $3 \times 3 : 3$. Con la configuración $4 \times 4 : 4$ y canales de 40 MHz se pueden conseguir velocidades teóricas de 600 Mbps. Dependiendo de las condiciones radio se utilizará uno u otro esquema de modulación y codificación (MCS, *Modulation and Coding Scheme*), variándose el tipo de modulación (BPSK, QPSK, 16-QAM o 64-QAM) y la cantidad de redundancia (FEC) empleada.

La agregación de tramas también mejora el rendimiento, definiéndose dos tipos: agregación de MSDU (A-MSDU, *Aggregated MAC Service Data Unit*) y agregación de MPDUs (A-MPDU, *Aggregated MAC Packet Data Unit*). Esta agregación permite empaquetar múltiples MSDUs o MPDUs en múltiples tramas, reduciéndose la sobrecarga

de cabeceras y aumentando la tasa binaria en el plano de usuario. La agregación A-MPDU requiere la Confirmación de Bloques (*Block Acknowledgment*) introducida en 802.11e y optimizada en este estándar.

Por último, este estándar tiene mecanismos para garantizar la compatibilidad con los dispositivos que implementen las variantes anteriores (802.11a, b y g).

Subcapas del nivel físico

Como se mostró en la Figura 2.1, IEEE 802.11 y sus variantes dividen la capa física en dos subcapas: PMD (capa Dependiente del Medio Físico, *Physical Medium Dependent*) y PLCP (Protocolo de Convergencia de la Capa Física, *Physical Layer Convergence Protocol*). Estas capas son diferentes para cada implementación del nivel físico (al contrario que la capa MAC que es común para los diferentes niveles físicos). En el nivel inferior, PMD define las características y los métodos de transmisión y recepción (esquemas de modulación y demodulación) a través del medio inalámbrico entre dos o más estaciones. En el nivel superior, PLCP adapta las capacidades de la capa PMD a los servicios requeridos a la capa física, i.e. define la metodología para mapear las MPDUs en un formato de tramas adecuado para enviar y recibir datos de usuario e información de gestión a través del medio inalámbrico entre dos o más estaciones. Para ello, PLCP añade un preámbulo (dependiente del nivel físico utilizado), compuesto por un campo de sincronización (*sync*) y un campo de delimitación de comienzo de trama (SFD, *Start Frame Delimiter*), y una cabecera a la MPDU. La cabecera incluye un campo de señal (*signal*, determina el tipo de modulación utilizada en 802.11 y 802.11b), un campo de servicio (información sobre el tipo de transmisión en 802.11b y g, valiendo 0 en 802.11), un campo de longitud (*length*, número de bytes de la MPDU) y un campo de comprobación (FCS, *Frame Check Sequence*) que protege tanto a los datos como a los tres campos anteriores mediante un código de redundancia cíclico (CRC-16). Esta estructura de PDU (*PLCP Packet Data Unit*, i.e. mensaje de nivel PLCP) permite la transferencia asíncrona de la MPDU, teniendo la estación receptora que sincronizarse a cada trama entrante.

De esta forma, la capa física proporciona servicios a la capa MAC de la estación a través de un SAP (Punto de Acceso a Servicio, *Service Access Point*), PHY-SAP (SAP de la capa física), mediante un conjunto de primitivas. Este SAP define el interfaz entre la

capa MAC y la capa PLCP. También se define un conjunto de primitivas entre las capas PMD y PLCP, llamado PMD-SAP (SAP de la capa PMD).

Las operaciones principales de las diferentes capas físicas son muy similares, siendo éstas:

- Comprobación de portadora / evaluación de canal libre (CS/CCA, *Carrier Sense / Clear Channel Assessment*)
- Transmisión (Tx)
- Recepción (Rx)

El procedimiento CS/CCA se utiliza para determinar el estado del medio inalámbrico, y se ejecuta sólo cuando la estación no está transmitiendo o recibiendo un paquete. Se utiliza para dos propósitos específicos: detectar el comienzo de una señal de la red (CS) y determinar si el canal está inactivo antes de transmitir un paquete (CCA).

La transmisión (Tx) se utiliza para enviar bytes individuales de la trama de datos. Este procedimiento lo solicita el mecanismo CS/CCA tras recibir una primitiva de petición de transmisión de la capa MAC. Tras ello, la capa MAC ejecuta el protocolo CSMA/CA (véase la Sección 2.2.2) antes de ejecutar el proceso de transmisión.

Por último, la recepción (Rx) se utiliza para recibir bytes individuales de la trama de datos. Cuando CS/CCA detecta un preámbulo de sincronismo válido seguido por una cabecera PLCP, ejecuta este procedimiento (sólo se “recibe” la trama MAC, no el preámbulo ni la cabecera PLCP). Para más información sobre las primitivas y su funcionamiento se puede consultar el estándar [26].

2.2.2 Capa de Control de Acceso al Medio (MAC)

La capa MAC es única para los diferentes niveles físicos existentes en IEEE 802.11, siendo responsable de los procedimientos relativos a la asignación del canal,

direccionamiento de PDUs, formato de tramas, control de errores, fragmentación y reensamblado.

En el estándar 802.11, esta capa define dos funciones para el acceso al medio: la Función de Coordinación Distribuida (DCF, *Distributed Coordination Function*) y la Función de Coordinación Puntual (PCF, *Point Coordination Function*). La primera es de implementación obligatoria mientras que la segunda es optativa. DCF es un mecanismo basado en contención, i.e. las estaciones compiten por el uso del medio inalámbrico, y permite la transmisión asíncrona de paquetes. Este modo puede utilizarse en cualquier tipo de topología, i.e. en redes *ad hoc* y en redes de infraestructura. Por su parte, PCF implementa un modo de transmisión sin contención y proporciona una transmisión síncrona a través de un acceso basado en sondeo (*polling*). La utilización de PCF requiere de un coordinador (PC, *Point Coordinator*) que gestione el uso del medio compartido, normalmente implementado en el punto de acceso. Por ello, PCF sólo se utiliza en redes de infraestructura. En este caso se alternan Períodos de Contención (CP, *Contention Period*) y Períodos Libres de Contención (CFP, *Contention Free Period*), usándose los primeros para el modo PCF y las segundas para el modo DCF.

Para desempeñar sus funciones, la capa MAC define tres tipos diferentes de trama: de gestión, de control y de datos. Las tramas de gestión se utilizan para asociar/desasociar estaciones con el AP, temporización/sincronización, y autenticación. Las tramas de control se usan para el intercambio de mensajes durante el período CP, para sus confirmaciones y para terminar el período CFP. Por último, las tramas de datos se usan para transmitir datos en cualquier período, y pueden combinarse con el envío de mensajes de sondeo (*polling*) y confirmaciones durante el período CFP.

El formato de estas tramas está ilustrado en la Figura 2.6. Nótese que el cuerpo de la trama (MSDU) es un campo de longitud variable consistente en los datos y en información usada para la encriptación si se implementan los protocolos WEP y/o

Bytes: 2	2	6	6	6	2	6	2	0 – 2324	4
Frame Control	Duration / ID	Address 1	Address 2	Address 3	Sequence Control	Address 4	QoS Control	Frame Body	FCS

Figura 2.6. Formato de trama MAC.

WPA/WPA2. Para la identificación de la estación (capa MAC) se usa el direccionamiento estándar de IEEE de 48 bits. Los dos bytes de duración indican el tiempo (en microsegundos) que el canal estará asignado para transmitir datos (MPDU). Los bits de tipo identifican a la trama como de control, de gestión o de datos. Los bits de subtipo identifican más específicamente el tipo de trama (e.g. trama de control *Clear To Send*). Finalmente, se añade un código de redundancia cíclico (CRC) de 32 bits para la detección de errores.

Función de Coordinación Distribuida

DCF es el esquema de acceso fundamental utilizado para el soporte de transferencias de datos asíncronas, basadas en *mejor esfuerzo*. Este método es obligatorio para todas las estaciones. DCF es el mecanismo exclusivo para redes *ad hoc* y puede operar en exclusiva o coexistir con PCF (*Point Coordination Function*, explicado posteriormente) en las redes de infraestructura.

Este método se basa en el protocolo CSMA/CA (*Carrier Sense Multiple Access with Collision Avoidance*), i.e. acceso múltiple con comprobación de portadora y evitación de colisiones. En este modo, una STA debe comprobar que el medio está inactivo antes de realizar la transmisión de un paquete. Existen dos mecanismos de comprobación de portadora, uno a nivel físico midiendo el interfaz radio y otro virtual en la capa MAC. La comprobación en la capa física determina la presencia de otras estaciones mediante el análisis de los paquetes detectados y la actividad del canal a través de la estimación de potencia de señal proveniente de otras estaciones. La comprobación virtual (característica opcional) se utiliza para que una STA informe al resto de STAs en el mismo BSS acerca de cuánto tiempo reservará el canal para la transmisión de sus tramas. Para ello, el emisor utilizará el campo de duración de la cabecera MAC de las tramas de datos, o de las tramas de control RTS (*Request To Send*) y CTS (*Clear To Send*). Al recibir esta duración, el resto de estaciones actualizarán sus temporizadores locales de *vector de asignación de red* (NAV, *Network Allocation Vector*).

Como se muestra en la Figura 2.7, si un paquete llega a una cola vacía y el medio ha estado inactivo durante un intervalo de tiempo superior a DIFS (*DCF InterFrame Space*), la STA origen podrá transmitir el paquete inmediatamente. Mientras tanto, el resto de STAs retardarán sus transmisiones y ajustarán sus NAVs, comenzando después un

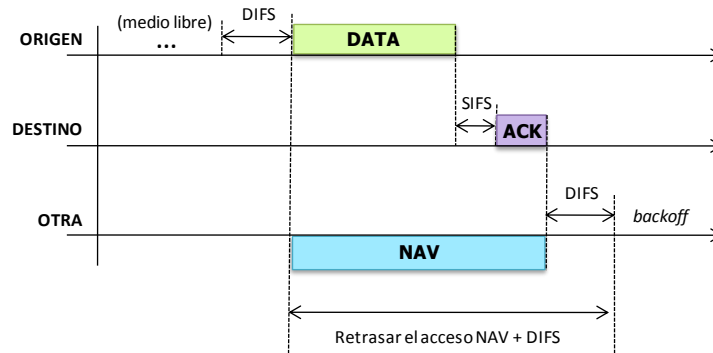


Figura 2.7. CSMA/CA en DCF.

proceso de retroceso (*backoff*). En dicho proceso, la STA calcula un intervalo de tiempo aleatorio y lo asigna al *temporizador de backoff*, valor elegido a partir de la Ventana de Contención (CW , *Contention Window*) como expresa la Ecuación 2.1.

$$\text{temporizador de backoff} = \text{aleatorio} [0, CW] \times \text{tiempo de ranura} \quad (2.1)$$

donde $CW_{\min} \leq CW \leq CW_{\max}$, y CW_{\min} , CW_{\max} y el *tiempo de ranura* dependen del tipo de nivel físico. El envío de tramas de confirmación se realiza tras un intervalo SIFS (*Short InterFrame Space*) cuya duración es inferior al resto de intervalos entre tramas, consiguiendo así que el acceso para estas tramas sea prioritario.

El temporizador de *backoff* decrece sólo cuando el medio está inactivo, parándose cuando una estación transmite. En cuanto expira, la estación tiene permiso para transmitir. En el caso de que dos estaciones transmitiesen simultáneamente ocurriría una colisión. Las estaciones no detectarían la colisión debido a la diferencia entre los niveles de potencia que transmiten y que reciben de la otra estación. Por ello, se utilizan confirmaciones positivas para indicarle al emisor que la trama ha sido recibida con éxito (véase la Figura 2.8). Si no se recibe la confirmación, el emisor supone que la trama colisionó y planifica la retransmisión, entrando de nuevo en el proceso de *backoff*. Para reducir la probabilidad de colisión, tras cada transmisión sin éxito se duplica la CW hasta su valor máximo (CW_{\max}). Si la transmisión tuvo éxito la CW se reinicia a su valor mínimo (CW_{\min}).

Otro problema solventado por DCF es la presencia de estaciones ocultas. Las estaciones ocultas son terminales que el receptor puede escuchar pero no son detectadas por los otros emisores. Por ello, los paquetes enviados por estos emisores pueden colisionar en el receptor. Para resolverlo, se utiliza el esquema de transmisión mediante tramas RTS / CTS. El emisor envía una trama corta RTS de 20 bytes solicitando permiso para enviar una trama de datos (véase la Figura 2.8), y el receptor responde con una trama CTS de 14 bytes concediéndolo. Cuando el emisor recibe esta trama, comienza la transmisión de su trama de datos. De esta forma, todas las estaciones escuchan o bien una trama RTS, o bien una trama CTS y actualizan sus NAVs de forma que no transmiten durante un tiempo, evitándose así la colisión. Como las tramas RTS y CTS son más cortas, es más difícil que se produzca una colisión (en cuyo caso utilizarían el proceso de *backoff*), mejorando el rendimiento de DCF en la mayoría de casos. Sin embargo, puede reducirse la velocidad en el caso de tramas de datos de tamaño pequeño, debido a la sobrecarga introducida por RTS / CTS. Por ello, se utiliza un umbral (*RTS_CTS_packet_size_threshold*) para decidir si se utiliza el mecanismo RTS / CTS, usando estos mensajes si la trama de datos tiene un tamaño mayor.

Función de Coordinación Puntual

PCF utiliza un esquema centralizado en el que un punto coordinador (PC, normalmente ubicado en el AP) realiza sondeos entre las estaciones para averiguar si

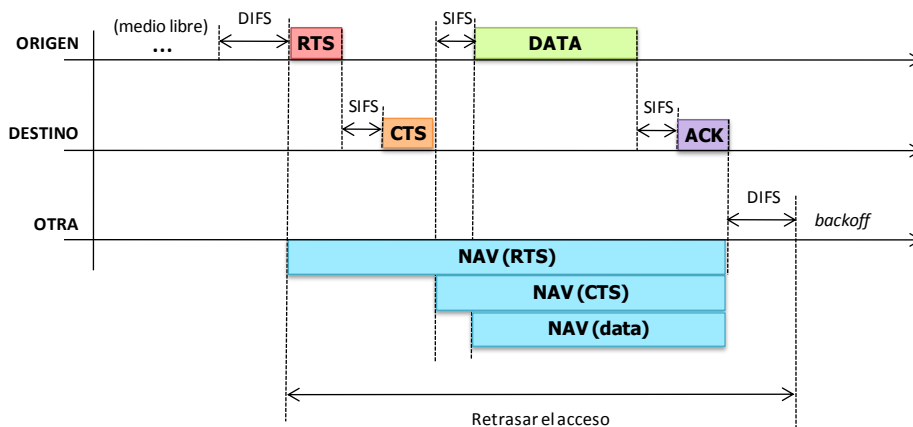


Figura 2.8. Esquema de acceso RTS / CTS.

quieren transmitir. Si un BSS tiene esta función habilitada, el tiempo de acceso al canal se divide en intervalos periódicos denominados intervalos de baliza (*beacon interval*) (véase la Figura 2.9). Este intervalo se compone de un período libre de contención (CFP) y un período de contención (CP). Durante el CFP, el PC mantiene una lista de estaciones registradas y sondea a cada estación de acuerdo a esa lista. Entonces, una STA consigue el permiso para transmitir una trama de datos cuando recibe un sondeo. Ya que cada STA sólo puede transmitir una trama con una longitud máxima dada, el CP conoce la duración máxima de este período (*duración_máxima_de_CFP*). El tiempo utilizado por el PC entre generación de tramas de baliza se denomina Tiempo Objetivo de Transmisión de Baliza (TBTT, *Target Beacon Transmission Time*). En la baliza, el PC incluye el TBTT siguiente y lo difunde a todas las estaciones en el BSS.

Para garantizar que ninguna estación en modo DCF pueda interrumpir la operación de PCF, el PC espera durante un intervalo PIFS (Espacio Entre Tramas PCF, *PCF InterFrame Space*) que es menor que DIFS para comenzar la transmisión de la trama de baliza. Cuando las estaciones reciben esta trama actualizan sus NAVs al valor de *duración_máxima_de_CFP* (o a la duración restante del CFP en el caso de una trama de baliza retrasada). Durante el CP se utiliza el esquema DCF, y es obligatorio que la trama de baliza permita al menos una trama de datos en este modo.

La Figura 2.9 muestra un acceso al medio típico durante el PCF. Cuando el PC sondea a una estación, puede enviar tramas de datos mediante *piggybacking* en la trama de sondeo *CF-Poll* (sondeo libre de contención). Entonces la STA le devuelve una trama de datos junto a la confirmación tras un intervalo SIFS. Cuando el PC sondea a la próxima STA, no sólo envía una trama de datos al destino, sino que incluye también la confirmación de la transmisión anterior. Nótese que la mayoría de las transmisiones están separadas por PIFS excepto en un caso: si la estación sondeada no responde al PC dentro del período PIFS, el PC sondeará a la siguiente estación. Las estaciones que no responden son eliminadas de la lista de sondeo hasta el siguiente CFP. En cualquier momento, el PC puede terminar este período enviando un paquete *CF-End*, cuya recepción por parte de las estaciones implica el reinicio de sus NAVs y la posibilidad de que transmitan durante el CP. Normalmente PCF utiliza un planificador *round-robin* para sondear a las estaciones de forma secuencial en el orden de la lista de sondeo, pero también se pueden utilizar otros mecanismos basados en prioridad si se requieren diferentes niveles de QoS por parte de las diferentes estaciones.

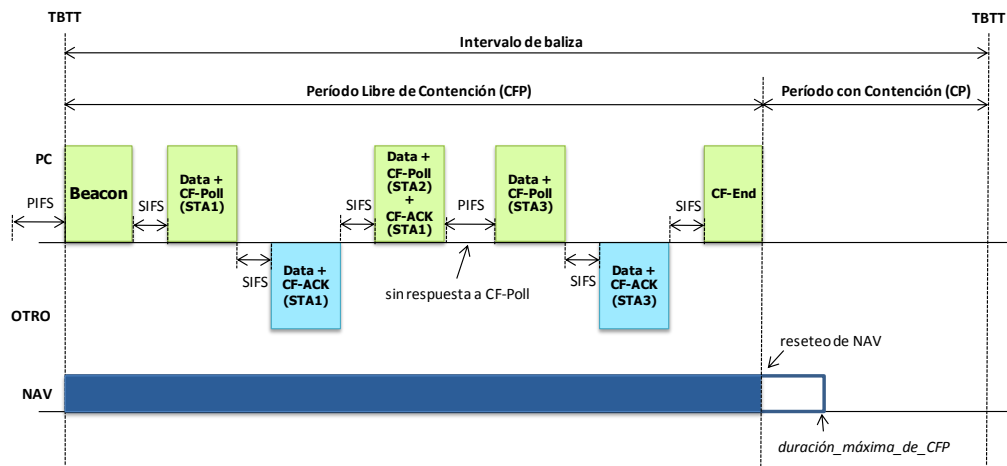


Figura 2.9. Ciclos de PCF y DCF.

Aunque PCF está diseñado para mejorar el soporte de QoS, no puede garantizar unos requisitos mínimos. Esto se debe a que el tiempo de transmisión de una estación no está acotado una vez que consigue una oportunidad para transmitir, ya que, aunque sólo recibe permiso para transmitir una trama, la posibilidad de que existan errores y sus respectivas transmisiones hace su duración no tenga un límite máximo.

Intervalos de tiempo entre tramas

Como ya se ha comentado en este apartado, la capa MAC utiliza el protocolo CSMA/CA para evitar las colisiones en el medio compartido, i.e. el interfaz inalámbrico. Para ello, las estaciones (incluidas el AP/PC) tienen que esperar un tiempo comprobando que el canal está inactivo antes de transmitir.

De forma resumida, para el cálculo de este tiempo de comprobación del canal se utiliza el intervalo SIFS para el envío de tramas de alta prioridad (confirmaciones, tramas RTS/CTS, sondeos, datos en PCF, etcétera); el intervalo DIFS se emplea en las transmisiones de datos en el modo DCF; y el intervalo PIFS se usa para el modo PCF (transmisión de tramas baliza, detección de falta de respuesta por parte de una estación, etcétera). Además, en el caso de alta probabilidad de colisión (e.g. tras la transmisión de

datos por parte de alguna estación) se emplea el mecanismo de *backoff*, basado en un tiempo aleatorio acotado por una ventana de contención (CW), cuyo valor mínimo y máximo están especificados.

Estos parámetros dependen del nivel físico utilizado. La Tabla 2.3 muestra de forma resumida sus valores para las capas físicas definidas por el estándar [26].

La relación entre los intervalos entre tramas viene dada por las Ecuaciones 2.2 y 2.3.

$$PIFS = SIFS + \text{tiempo de ranura} \quad (2.2)$$

$$DIFS = SIFS + 2 \times \text{tiempo de ranura} \quad (2.3)$$

Existe además un Espacio Entre Tramas Extendido (EIFS, *Extended InterFrame Space*) cuyo valor no está especificado pero debe ser superior al resto de de IFSs. Este intervalo se utiliza para las retransmisiones.

Estándar	Tiempo de ranura (μs)	SIFS (μs)	DIFS (μs)	PIFS (μs)	CW _{min}	CW _{max}
FHSS	50	28	128	78	15	1023
DSSS	20	10	50	30	31	1023
IR	8	10	26	18	63	1023
OFDM (20MHz)	9	16	34	25	15	1023
OFDM (10MHz)	13	32	58	45	15	1023
OFDM (5MHz)	21	64	106	85	15	1023
HR/DSSS (<i>High Rate</i>)	20	10	50	30	31	1023
ERP-OFDM (<i>Extended Rate PHY</i>)	9 ó 20 (*)	10	28 ó 50	19 ó 30	15 ó 31	1023

Tabla 2.3. Parámetros temporales de los niveles físicos de IEEE 802.11-2007.

(*) Se puede optar por transmitir usando ranuras temporales cortas de 9 μs .

2.3 Calidad de Servicio en redes WLAN: estándar IEEE 802.11e

El estándar WLAN original (IEEE 802.11-1997 [1]) no define ningún mecanismo de acceso que permita el cumplimiento de requisitos estrictos de tiempo real. Para solventar esta falta de soporte de QoS, el borrador IEEE 802.11e (posteriormente incluido en el estándar IEEE 802.11-2007 [26]) define una nueva función de coordinación llamada Función de Coordinación Híbrida (HCF, *Hybrid Coordination Function*). HCF combina los esquemas de acceso anteriores (DCF y PCF) con mecanismos de mejora de QoS para conseguir la diferenciación de los servicios. HCF proporciona mecanismos de acceso al canal de forma tanto distribuida como coordinada, de una manera similar a los modos DCF y PCF en el estándar original. El mecanismo de acceso distribuido, basado en contención, se denomina Acceso Mejorado al Canal de forma Distribuida (EDCA, *Enhanced Distributed Channel Access*), y el mecanismo de acceso coordinado, sin contención, se denomina Acceso al Canal Controlado por HCF (HCCA, *HCF Controlled Channel Access*).

La función HCF se ejecuta en el coordinador híbrido (HC, *Hybrid Coordinator*), normalmente ubicado en el mismo equipo que el QAP (*QoS Access Point*). Esta función se denomina *híbrida* porque, a diferencia de PCF, las transmisiones en HCCA pueden realizarse tanto en el período libre de contención como en el período de contención. Los mecanismos de acceso EDCA y HCCA se explican en los siguientes subapartados.

2.3.1 Acceso al canal de forma distribuida: EDCA

En el mecanismo de acceso EDCA cada estación (QSTA, *QoS Station*) tiene cuatro colas de transmisión o categorías de acceso (AC, *Access Category*), que se comportan como estaciones virtuales. Las cuatro ACs son AC_VO (para tráfico de voz), AC_VI (para tráfico de vídeo), AC_BE (para tráfico *best effort*) y AC_BK (para tráfico de *background*). De esta forma, al contrario que en DCF donde todo el tráfico compartía una cola común, en EDCA cada tipo de tráfico se encola en su AC correspondiente.

2.3. Calidad de Servicio en redes WLAN: estándar IEEE 802.11e | 31

Cada una de las estaciones virtuales (representadas en la Figura 2.10) tiene un conjunto de parámetros específicos que permiten un acceso al medio de forma diferenciada. Estos parámetros son:

- La longitud de la ventana de contención a usar para el procedimiento de *backoff*
- La cantidad de tiempo que una estación tiene que esperar antes de iniciar el proceso de *backoff* o la transmisión
- La duración máxima de la transmisión una vez que se ha accedido correctamente al medio

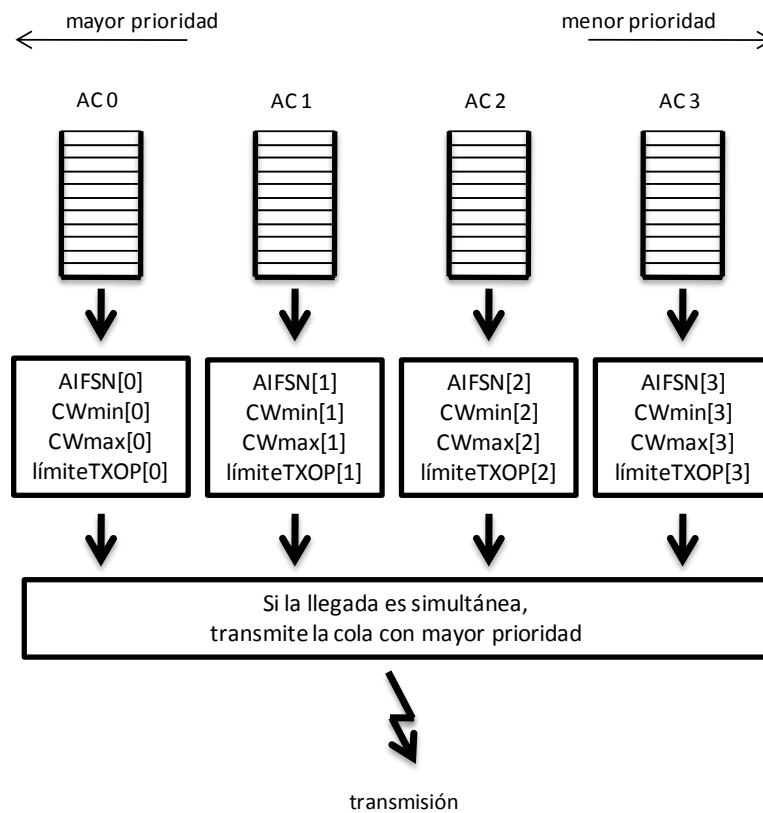


Figura 2.10. Categorías de Acceso en el mecanismo de acceso EDCA.

La Tabla 2.4 muestra cómo se consigue esta diferenciación de acceso al medio mediante la asignación de diferentes valores a estos parámetros, explicados a continuación.

Los parámetros CW_{\min} y CW_{\max} son el valor mínimo y máximo de la ventana de contención, respectivamente. La ventana de contención se utiliza para calcular el número de ranuras temporales de cesión antes de acceder al medio. Fijando valores bajos a CW_{\min} y CW_{\max} se asigna al AC una mayor prioridad, ya que estadísticamente el tiempo de espera será menor.

El *número de espacio entre tramas de arbitraje* (AIFSN, *Arbitration InterFrame Space Number*) es el número de ranuras temporales tras un período de espera SIFS que una estación debe esperar antes de comenzar un proceso de *backoff* o una transmisión. AIFSN afecta al Espacio Entre Tramas de Arbitraje (AIFS), que especifica dicha duración (en tiempo en vez de en ranuras temporales). De esta forma, asignándole un valor bajo a este parámetro para una cierta AC se su prioridad aumentará. AIFS se calcula mediante la expresión dada por la Ecuación 2.4.

$$AIFS[AC] = SIFS + AIFSN[AC] \times \text{ranura temporal} \quad (2.4)$$

El parámetro *límite de TXOP* especifica la longitud (o duración máxima) de una oportunidad de transmisión. Un valor mayor que cero significa que una AC puede transmitir varias tramas en tanto que la duración de las transmisiones no sea mayor que este parámetro. Un valor igual a cero indica que sólo se puede enviar una trama de datos o gestión (más un posible intercambio de tramas RTS/CTS). De esta forma, dándole un valor superior a cero también se aumenta la prioridad de la AC.

Cada AC compite independientemente por las TXOPs basándose en los parámetros

AC	CWmin	CWmax	AIFSN	límite de TXOP
AC_VO	7	15	2	3.264 ms
AC_VI	15	31	2	6.016 ms
AC_BE	31	1023	3	0
AC_BK	31	1023	7	0

Tabla 2.4. Parámetros por defecto para EDCA utilizando la capa física DSSS de 802.11b.

2.3. Calidad de Servicio en redes WLAN: estándar IEEE 802.11e | 33

descritos anteriormente. Una vez que la AC ha comprobado que el medio está inactivo durante al menos la duración indicada por $AIFS[AC]$, inicia su temporizador de *backoff*. Si dos o más ACs dentro de una misma estación quisieran transmitir en el mismo instante de tiempo, ocurriría una colisión interna. Esta colisión se resuelve dentro de la estación, de forma que se transmitirán las tramas de la AC con mayor prioridad (que consigue la TXOP), mientras que la AC con menor prioridad actuará como si hubiese ocurrido una colisión externa en el medio inalámbrico. Este tipo de colisiones se denomina *colisión virtual*, ya que no ha ocurrido en el medio físico real.

En ese sentido, EDCA permite la diferenciación de servicios pero no un soporte de QoS garantizado ya que, aunque una categoría de acceso de mayor prioridad tiene más probabilidades de acceder al canal, podría ocurrir eventualmente (con menor probabilidad) que una AC menos prioritaria consiguiese hacerse con el medio inalámbrico. Por ello, se dice que EDCA proporciona prioridad estadística pero no parametrizada. La Figura 2.11 presenta un ejemplo en el que una AC menos prioritaria consigue acceder al canal antes que otra con mayor prioridad. En este ejemplo, las QSTAs 3 y 4 intentan acceder al medio tras la transmisión llevada a cabo entre las estaciones 1 y 2. Aunque el flujo de datos de la QSTA 3 es más prioritario (AC 0, i.e. de voz), debido a la aleatoriedad del proceso de *backoff*, tiene que esperar más tiempo que la QSTA 4 (cuyo flujo es menos prioritario y pertenece a la AC 3, i.e. *background*).

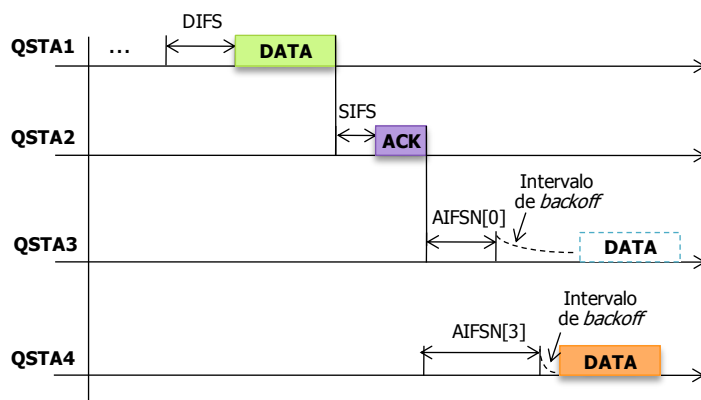


Figura 2.11. Ejemplo de prioridad estadística utilizando EDCA.

2.3.2 Acceso al canal de forma coordinada: HCCA

El mecanismo HCCA se diseñó para permitir la provisión de QoS parametrizada, utilizando un coordinador híbrido que gestiona la asignación del ancho de banda para transmitir en el medio inalámbrico. El HC tiene una prioridad mayor para el acceso que las estaciones inalámbricas, algo necesario para que pueda asignar las oportunidades de transmisión a las estaciones en los períodos con o sin contención. Una QSTA, basándose en sus requisitos, solicita TXOPs al HC, tanto para sus propias transmisiones como para aquellas desde el QAP hacia ellas mismas. El coordinador híbrido, que normalmente se ubica en el QAP, o bien acepta o bien rechaza la petición basándose en los parámetros soportados por la QSTA. Para las transmisiones en el enlace ascendente, el HC asigna turnos a la QSTA basándose en sus requisitos. Para las transmisiones en el enlace descendente, el QAP consigue directamente las TXOPs del HC y entrega las tramas encoladas hacia la QSTA.

En HCCA, cada QSTA puede establecer hasta ocho Flujos de Tráfico (TS, *Traffic Stream*), que se caracterizan por las Especificaciones de Tráfico (TSPEC, *Traffic Specification*) negociadas entre la QSTA y el QAP. Las TSPECs incluyen un conjunto de parámetros [2], la mayoría opcionales, que se utilizan para describir el tráfico y sus requisitos. Algunos de los campos principales de la TSPEC de un TS i son:

- La tasa de datos media (R_i , *Mean Data Rate*)
- El tamaño de la unidad de datos nominal del servicio (N_i , *Nominal Service Data Unit (SDU) Size*)
- El tamaño máximo de SDU (M_i , *Maximum SDU Size*)
- La tasa mínima de la capa física (I_i , *Minimum PHY Rate*)
- La cota de retardo (D_i , *Delay Bound*)
- El intervalo máximo de servicio (MSI_i , *Maximum Service Interval*).

La caracterización del flujo de tráfico incluye los cinco primeros parámetros,

mientras que MSI_i es un requisito del servicio. Más concretamente, la cota del retardo indica el tiempo de vida máximo de una SDU (antes de su descarte) y el intervalo máximo de servicio especifica el tiempo máximo que puede pasar entre dos asignaciones de TXOPs a una misma QSTA en el enlace ascendente.

El QAP inicia fases de acceso controlado (CAP, *Controlled Access Phase*) con una prioridad mayor que los mecanismos de acceso basados en contienda (i.e. DCF o EDCA) ya que sólo espera un período PIFS antes de acceder al medio. Una CAP consiste en una o varias TXOPs, durante las cuales el QAP puede transmitir tramas de datos hacia las QSTAs o recibir desde ellas (mediante el envío previo de una trama de sondeo o *polling*). Esta trama especifica el tiempo máximo que una QSTA puede utilizar el medio inalámbrico (incluyendo los espacios entre tramas y tramas de confirmaciones), de forma que al estar este tiempo acotado, se proporciona QoS para aplicaciones con requisitos estrictos de tiempo real. El QAP puede comenzar las CAPs (para TXOPs de HCCA) en los períodos con y sin contienda, mientras que las TXOPs de EDCA sólo se permiten en el período de contienda (véase la Figura 2.12).

2.3.3 Requisitos de los planificadores para HCCA

El planificador de paquetes utilizado en HCCA no está determinado por el estándar, sino que se puede utilizar cualquiera siempre que cumpla los requisitos exigidos. En concreto, el estándar especifica que se debe satisfacer la siguiente condición: la asignación de oportunidades de transmisión debe verificar los parámetros de TSPEC negociados para los flujos admitidos. De esta forma, la duración acumulada $W_i(t_1, t_2)$ de las TXOPs asignadas para el TS i , para cualquier intervalo $[t_1, t_2]$ mayor que un Intervalo

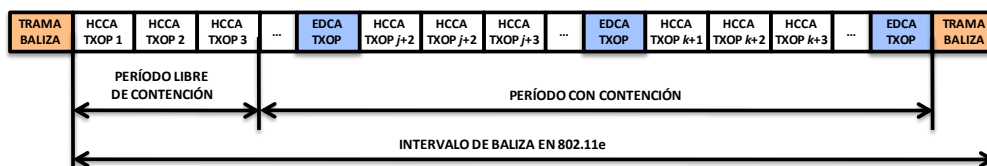


Figura 2.12. Intervalo de baliza de IEEE 802.11e.

Especificado mínimo (*Specification Interval*, anunciado por el QAP) debe satisfacer la Ecuación 2.5,

$$W_i(t_1, t_2) \geq \frac{t_2 - t_1 - \Delta_i}{\frac{N_i}{R_i}} \cdot tx(N_i, \Gamma_i) \quad (2.5)$$

donde $tx(N_i, \Gamma_i)$ es el tiempo para transmitir una SDU de tamaño nominal N_i a la tasa del medio físico Γ_i y Δ_i es igual a MSI_i si está especificada o D_i en otro caso. El tiempo de transmisión nominal para el TS i , incluyendo las tramas de control MAC y los espacios entre tramas, se puede calcular mediante la Ecuación 2.6,

$$tx(N_i, \Gamma_i) = SIFS + \left(\frac{h_{PHY} + h_{MAC} + N_i}{\Gamma_i} \right) + SIFS + \left(\frac{h_{PHY} + h_{ACK}}{\Gamma_{BASICA}} \right) \quad (2.6)$$

donde h_{PHY} , h_{MAC} y h_{ACK} son los tamaños de las cabeceras de la capa física, de la capa MAC y las tramas de confirmación respectivamente, y Γ_{BASICA} es la tasa de transmisión del medio físico para las tramas de confirmación.

2.3.4 Planificador de referencia en HCCA

El borrador IEEE 802.11e no especifica cómo planificar las TXOPs. Sin embargo, sí incluye la descripción de un planificador de ejemplo que puede ser utilizado como referencia. Este planificador primero determina un Intervalo de Servicio (SI) global, que es menor que el parámetro MSI de todos los flujos admitidos y múltiplo del intervalo de baliza (que incluye los períodos con y sin contienda, véase la). Cualquier flujo admitido conseguirá una TXOP al final de cada SI, proporcionando así al menos una TXOP dentro del límite de tiempo impuesto por el parámetro MSI de ese flujo. El planificador determina la duración necesaria de la TXOP para cada flujo considerando el número de paquetes que pueden llegar dentro de un SI, según la Ecuación 2.7,

$$TXOP_i = \max \left(\frac{nMSDU_i \times N_i}{\Gamma_i} + O, \frac{M_i}{\Gamma_i} + O \right) \quad (2.7)$$

donde $TXOP_i$ es la duración de la oportunidad de transmisión para el TS i , N_i es el

tamaño nominal de MSDU (MAC SDU), T_i es la tasa de transmisión de la capa física, M_i es el máximo tamaño permitido para la MSDU, O denota la sobrecarga (que incluye los intervalos entre tramas y el tiempo de transmisión de las tramas de confirmación y de sondeo) en unidades de tiempo y $nMSDU$ es el número de MSDUs que llegaron a la QSTA con una tasa binaria media λ_i durante un período SI, que se puede calcular de la Ecuación 2.8.

$$nMSDU_i = \left\lceil \frac{SI \times \lambda_i}{N_i} \right\rceil \quad (2.8)$$

En la norma también se sugiere un algoritmo de control de admisión para este planificador. Se admite un flujo nuevo si se satisface la expresión dada por la Ecuación 2.9.

$$\frac{TXOP_{k+1}}{SI} + \sum_{i=1}^k \frac{TXOP_i}{SI} \leq \frac{T_{baliza} - T_{período de contención}}{T_{baliza}} \quad (2.9)$$

donde k es el número de flujos existentes, T_{baliza} es la duración de un período de baliza y $T_{período de contención}$ es el tiempo utilizado para el acceso mediante EDCA. Es decir, se comprueba si el tiempo de su oportunidad de transmisión ($TXOP_{k+1}$) puede ser asignado en el tiempo total para HCCA ($T_{baliza} - T_{período de contención}$) considerando el tráfico del resto de flujos.

2.4 Trabajos relacionados

La presente tesis está basada en el mecanismo HCCA ya que, como se ha comentado, es el único modo de acceso al medio en redes IEEE 802.11 que permite garantizar el cumplimiento de requisitos estrictos de tiempo en los flujos de tráfico transportados. Además, HCCA puede implementarse en una tarjeta estándar 802.11e a través de controladores *software* [28].

En esta sección, con el objetivo de establecer el contexto de este trabajo, se proporciona una revisión de las estrategias de asignación de recursos encontradas en la

literatura sobre redes IEEE 802.11e basadas en HCCA. Éstas se han clasificado en dos clases principales: algoritmos de planificación que cumplen el requisito impuesto por el estándar (Ecuación 2.5), y otros planificadores que mejoran la eficiencia del tráfico CBR y/o VBR sin comprobar la condición anterior. En términos generales, dentro de estas categorías hay trabajos relevantes sobre disciplinas de planificación basadas en el vencimiento (*deadline*) de los paquetes [3] [4] [5] [6] [7] o en la estimación de las longitudes de las colas [8] [9] [10] [11] [12]] donde se almacenan los paquetes a transmitir.

El análisis se centrará en la primera categoría, dado que los esquemas propuestos en la tesis pretenden garantizar requisitos de retardo y de ancho de banda, cumpliendo así con las especificaciones del estándar. Tras los estudios realizados, el autor sólo identifica cuatro algoritmos que hayan demostrado proporcionar las garantías requeridas para cada flujo: el planificador de referencia (explicado en la Sección 2.3.4), el esquema SETT-EDD [3] y los planificadores RTH [4] y WTTP [13]. Además de estos algoritmos, el planificador propuesto en [5] también cumple la condición anterior ya que, como se explicará más tarde, todos los planificadores de la familia LR (*latency-rate*), por definición, proporcionan un tiempo de servicio garantizado a cada flujo (véase el Teorema 1 dentro de la Sección 4.1).

La *planificación basada en el tiempo de transmisión estimado* utiliza la política de *fecha de vencimiento más temprana* (SETT-EDD, *Scheduling based on Estimated Transmission Time with Earliest Due Date*) [3]. Este planificador estima los instantes de partida mínimo y máximo, mediante los parámetros intervalo de servicio mínimo (*mSI*) e intervalo de servicio máximo (*MSI*). Utilizando el instante de llegada de los paquetes (que es estimado en el caso del sentido ascendente), las SDUs se sirven según el algoritmo de fecha de vencimiento más temprana (EDD). De esta forma se obtiene mejor rendimiento que el planificador de referencia especialmente con tráfico VBR. Sin embargo, SETT-EDD no proporciona cotas a los parámetros de QoS tales como ancho de banda y retardo.

Cicconetti *et al* propusieron un planificador HCCA de tiempo real (RTH, *Real-Time HCCA scheduler*) [4], basado en el algoritmo de planificación *primero la fecha de vencimiento más temprana* (EDF, *Earliest Deadline First*) con la *política de recursos de pila* (SRP, *Stack Resource Policy*). Este algoritmo se puede encontrar en la literatura de

planificación de tareas periódicas de tiempo real en entornos de multiprogramación. En el esquema se identifican tres bloques: un algoritmo de control de admisión, un algoritmo en diferido de planificación de tiempos y un procedimiento de ejecución. Dado que la planificación periódica de tiempos se realiza cuando se admite un nuevo flujo, esta estrategia está orientada sólo a tráfico CBR y por tanto no puede gestionar de forma eficiente el tráfico elástico.

El *protocolo inalámbrico de paso de testigos planificado* (WTTP, *Wireless Timed Token Protocol*) [13] proporciona una capacidad fija a los flujos CBR, mientras que reserva una tasa mínima a los flujos VBR. Esta aproximación permite utilizar el ancho de banda no usado para el tráfico VBR. WTTP está basado en el Protocolo de Paso de Testigos Planificado (TTP, *Timed Token Protocol*) utilizado para redes FDDI (Interfaz de Datos Distribuidos por Fibra), que básicamente es un algoritmo de tipo *round-robin*. La lista *round-robin* para la asignación de turnos incluye únicamente TSs con paquetes a transmitir, de forma que los flujos se añaden o eliminan de la cola basándose en el tamaño de sus colas (enviado mediante *piggyback* desde las QSTAs al QAP). Este planificador garantiza que cada TS tendrá TXOPs asignadas a una velocidad suficiente. Sin embargo, no garantiza que todos los paquetes se transmitirán dentro de su límite de retardo.

Foronda *et al* [5] utilizan un control de admisión y un planificador modificado de la familia LR (*latency-rate*) con un conformador *token bucket*. Aunque tiene algunas similitudes con la estrategia de planificación adoptada en esta tesis, se han identificado algunas desventajas. Primero, la planificación se realiza de forma periódica, lo que reduce su flexibilidad y aumenta la latencia del servidor (superior a un período SI). A pesar de que su latencia puede ser aceptable en redes 802.11a debido a su alta velocidad, puede conllevar retardos extremadamente altos en redes 802.11b o en interfaces físicos más lentos. Por último, hay que especificar la tasa binaria a la que se deben servir los diferentes flujos, hecho que es adecuado para aplicaciones de tiempo real pero no para aplicaciones de tráfico elástico como las basadas en el protocolo TCP (e.g. navegación web y descarga de ficheros), ya que éstas intentarán aumentar su tasa basándose en el mecanismo de control de congestión.

2.5 Conclusiones

En este capítulo se ha descrito el funcionamiento de los principales estándares utilizados en redes inalámbricas de área local, i.e. el estándar IEEE 802.11 y sus variantes en la capa física. Además, se han descrito detalladamente los diferentes modos de acceso definidos tanto por la norma original (DCF y PCF) como por la corrección 802.11e (EDCA y HCCA), haciendo énfasis en la falta de soporte de QoS de algunos y las garantías de otros.

Las diferentes técnicas encontradas en la literatura científica para el cumplimiento de requisitos de QoS en este tipo de redes están limitadas por su ámbito de aplicación, no habiendo encontrado ninguna que satisfaga de forma eficiente los requisitos exigidos en un entorno de tráfico heterogéneo con diferentes tipos de servicios, siendo éste el principal problema a resolver en esta tesis.

Capítulo 3

Calidad de Experiencia (QoE)

La Calidad de Experiencia (QoE, *Quality of Experience*) se define como “*la aceptación global de una aplicación o servicio tal cual es percibida por el usuario final*” [29]. Por tanto, la QoE incluye los efectos del sistema completo extremo a extremo (cliente, terminal, red, infraestructura de servicios, etcétera), además de poder estar influenciada por las expectativas del usuario respecto al servicio o por el contexto en particular de aplicación.

Por otro lado, la Calidad de Servicio (QoS, *Quality of Service*) se define como “*la totalidad de las características de un servicio de telecomunicaciones que determinan su capacidad para satisfacer las necesidades explícitas e implícitas del usuario del servicio*” [30].

Ambos conceptos implican múltiples factores –a veces puede que solapados–, si bien de forma generalizada se asume que la QoE está relacionada con la percepción subjetiva del usuario final, mientras que la QoS es un concepto objetivo determinado por métricas observables del rendimiento de la red.

Tal como se comentó en el Capítulo 1, las principales contribuciones de este trabajo

42 | Calidad de Experiencia (QoE)

(véase la Sección 1.3) se basan en la hipótesis de que la incorporación de criterios de QoE en el diseño de algoritmos de Gestión de Recursos Radio, frente a la aproximación tradicional que sólo incluye parámetros de QoS, puede mejorar de forma significativa la calidad percibida por el conjunto de usuarios –los cuales generan y consumen diferentes tipos de tráfico– de una red IEEE 802.11.

Por tanto, resulta necesario tener una estimación de la calidad percibida por los usuarios finales de los diferentes servicios considerados. Esta afirmación conlleva dos consecuencias:

- Es necesario identificar una métrica común para los diferentes servicios.
- El cálculo de la métrica debe realizarse en tiempo real para reaccionar adecuadamente bajo condiciones cambiantes en la red (e.g. considerando la carga actual de tráfico).

Para estimar la QoE se propone la utilización de la Nota Media de Opinión (MOS, *Mean Opinion Score*) como medida unificadora para los diferentes servicios.

La escala MOS fue inicialmente definida para servicios de voz en [31] (Anexo A: *Pruebas de opinión sobre la conversación*) como una escala numérica (de 1 a 5) para que los usuarios, en experimentos diseñados expresamente, indicaran cuál era su nivel de satisfacción respecto a este servicio. La Tabla 3.1 indica la correspondencia entre la opinión del usuario y su valoración numérica. Esta escala ha sido posteriormente adoptada en otras recomendaciones de la ITU-T (Unión Internacional de Telecomunicación, sector de estandarización en Telecomunicación) para no sólo servicios conversacionales, sino también para aplicaciones de datos [32], video telefonía [33] e IPTV [34].

Dado que para alcanzar los objetivos de la tesis, la puntuación MOS ha de ser obtenida en tiempo real y de forma automática, es obvio que se precisa de una metodología que permita su estimación sin requerir la participación de usuarios reales (ni realizar encuestas).

<i>Opinión</i>	<i>Valor</i>
Excelente	5
Buena	4
Regular	3
Mediocre	2
Mala	1

Tabla 3.1. Escala MOS.

Para la inclusión de criterios subjetivos en los diferentes mecanismos y políticas de Gestión de Recursos Radio se definen las *funciones de utilidad*, como aquellas que establecen para un servicio dado una correspondencia entre los indicadores objetivos del rendimiento del servicio (e.g. tasa binaria, retardo, *jitter*, etc.) y la experiencia subjetiva del usuario. Disponer de esta correspondencia –o función de utilidad– permite automatizar el proceso de estimación de la QoE, posibilitando así su inclusión en cualquier procedimiento *on-line* que opere en tiempo real, como por ejemplo la asignación de recursos radio.

Para el caso particular de los algoritmos de RRM podría ser excesivamente complejo considerar individualmente cada uno de servicios prestados. Por este motivo, en lo que sigue los diferentes servicios se agrupan en *clases de tráfico*, tal y como se expone en la Sección 3.1.

El resto del capítulo se ha organizado de la siguiente manera. En la Sección 3.2 se definen las diferentes funciones de utilidad para los servicios considerados. La Sección 3.3 expone la metodología utilizada para evaluar el rendimiento de los diferentes mecanismos desarrollados. Y por último, en la Sección 3.4 se resumen las principales conclusiones extraídas.

3.1 Clases de Tráfico

Una *Clase de Tráfico* (CT) se define como un conjunto de servicios que son

44 | Calidad de Experiencia (QoE)

sensibles al mismo tipo de requisitos de QoS. Este concepto ya ha sido utilizado en otras tecnologías inalámbricas como UMTS [35] o LTE [36]. El objetivo de esta definición es simplificar el posible conjunto de servicios considerados, permitiendo a su vez realizar un tratamiento diferenciado en los algoritmos desarrollados.

En concreto, con el objetivo de cubrir los servicios más extendidos en la actualidad, se han adoptado las siguientes clases de tráfico:

- **Conversacional** – Dentro de esta clase se incluyen los tráficos de audio (VoIP) y vídeo (videoconferencia) que exijan interactividad. Estos tráficos se caracterizan por que tanto el tiempo de transferencia como la variación de la separación temporal (*jitter*) entre unidades de información (e.g. tramas de sonido, cuadros de imagen, paquetes) deben estar acotados. En esta clase el tráfico es bidireccional y entre humanos, por tanto los límites de tolerancia –retardo y *jitter*– están determinados por los mecanismos de percepción –aparatos auditivo y visual.
- **Streaming** – Esta CT incluye tráficos de vídeo o audio en tiempo real pero sin interactividad –es decir, no conversacional–, no obstante el destinatario final de la información sigue siendo humano. El transporte se realiza por tanto en un único sentido.

Este esquema, relativamente reciente, está caracterizado por que el *jitter* al igual que en la clase conversacional debe estar acotado, aunque en este caso para el retardo no hay requisitos tan estrictos. Aquí, a diferencia de la clase anterior, dado que el equipo receptor realiza una alineación temporal de los datos (por medio de un *buffer*) el *jitter* máximo sólo está condicionado por esta función de alineación. De esta forma, el *jitter* tolerable es muy superior al impuesto por los mecanismos de percepción en la clase conversacional.

- **TCP de tiempo real** – Clase de tráfico idéntica a la anterior, pero para servicios que utilizan TCP como protocolo de transporte.
- **Interactiva** – Esta clase de tráfico se define para el tráfico generado por un usuario final, en este caso una máquina o un ser humano, que estando en línea solicita información a un equipo remoto (e.g. un servidor). Ejemplos

significativos son la navegación web (interacción hombre-máquina) o el acceso a *web services* (protocolos SOAP [37] o REST [38] entre máquinas).

El tráfico interactivo corresponde al paradigma cliente/servidor, de forma que el tiempo de ida y vuelta (*round trip time*) es uno de sus características principales. Un requisito para esta CT es que el contenido de los paquetes se debe transmitir de forma fiable (sin errores y con control de flujo).

- **De fondo** – Esta clase se define para tráfico unidireccional, sin requisitos de retardo aunque sí de fiabilidad. Ejemplos significativos son la transmisión de ficheros en segundo plano (*background*), el envío de correo electrónico y el servicio de mensajes cortos (SMS).

Para cada una de las clases de tráfico propuestas, atendiendo a su amplia difusión y representatividad, se han considerado los siguientes servicios: Voz sobre IP (VoIP) para la CT *conversacional*, *streaming* de vídeo sobre RTP/UDP para la clase *streaming*, *streaming* de vídeo sobre TCP (e.g. difusión de vídeo en YouTube [39]) para la clase *TCP de tiempo real*, navegación web para la clase *interactiva* y FTP para la CT *de fondo*.

3.2 Funciones de utilidad

Esta sección describe las funciones de utilidad definidas para los servicios utilizados en la experimentación desarrollada. Toda función de utilidad mapea métricas objetivas de red (e.g. tasa binaria o retardo) en valores de calidad subjetiva. De acuerdo con [31], se ha utilizado la escala MOS para la definición de las funciones, usando un rango continuo desde 1 (usuario totalmente insatisfecho) hasta 5 (usuario totalmente satisfecho).

Un requisito indispensable para estas funciones es que permitan su estimación *on-line*, ya que los valores de MOS obtenidos serán utilizados por procedimientos ejecutados en tiempo real.

Otro requisito importante es la sencillez, ya que un único equipo (el punto de acceso) tendrá que realizar estos cálculos –en tiempo real– para múltiples flujos de tráfico.

3.2.1 Función de utilidad para tráfico VoIP

La estimación de MOS para el tráfico VoIP está basada en el Modelo-E (*E-Model*) [22]. Considerando escenarios inalámbricos de interior, y suponiendo que las fuentes VoIP utilizan el códec G.711 [40], la expresión particular del MOS se detalla en la Ecuación 3.1. Para el caso considerado los parámetros del modelo se muestran en la Tabla 3.2.

$$MOS = \begin{cases} 1.0 & \text{si } R < 0 \\ 4.5 & \text{si } R > 100 \\ 1 + 3.5 \cdot 10^{-3}R + 7 \cdot 10^{-6}R \cdot (R - 60) \cdot (100 - R) & \text{en otro caso} \end{cases} \quad (3.1)$$

donde

$$R = R_0 - I_s - I_d - I_{e,eff} + A$$

$$I_d = 0.024 \cdot D + \max\{0, 0.11 \cdot (D - 177.3)\}$$

$$I_{e,eff} = I_e + \frac{(95 - I_e) \times lp}{lp + B_{pl}}$$

siendo lp la probabilidad de pérdida de paquetes y D el retardo en milisegundos.

La Figura 3.1 muestra la evolución del valor MOS en función del retardo extremo a extremo (D) para el caso concreto de un escenario inalámbrico de interior en el que las

<i>Parámetro</i>	<i>Valor</i>	<i>Descripción</i>
A	5	Factor de ventaja (<i>advantage</i>)
R_0	93.2	Relación señal-ruido básica
I_s	0.0	Factor de deterioro por simultaneidad
I_e	0.0	Factor de deterioro por equipo
B_{pl}	0.043	Factor de robustez de pérdida de paquetes

Tabla 3.2. Parámetros del Modelo E para entornos de interior con fuentes usando el códec G.711.

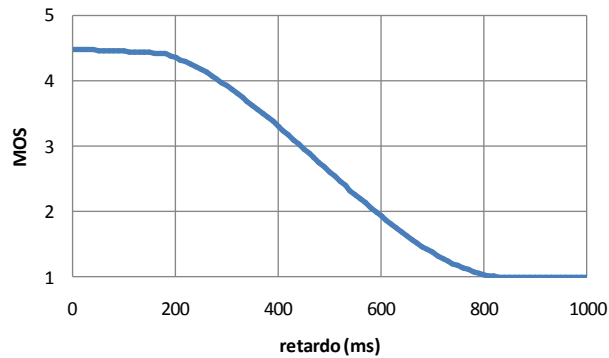


Figura 3.1. Ejemplo de función de utilidad para tráfico *VoIP* (entorno interior y códec G.711).

fuentes de tráfico utilizan el códec G.711 de acuerdo con la Ecuación 3.1. En este ejemplo se ha considerado que no hay pérdidas de paquetes.

En la literatura se han propuesto otros modelos complementarios, siendo PESQ [41] (*Perceptual Evaluation of Speech Quality*) uno de los más relevantes. PESQ también está recomendada por la ITU-T, pero al ser un algoritmo de *referencia completa* requiere de la señal original para compararla con la señal recibida y, por tanto, no puede ser calculada para las llamadas en curso en un sistema real.

Pese a la existencia de otras alternativas (se puede encontrar una clasificación de diferentes métodos en [42] [43]), el Modelo-E ha sido elegido por su sencillez y funcionamiento en tiempo real al utilizar indicadores disponibles habitualmente.

3.2.2 Función de utilidad para tráfico *streaming* de vídeo

La evaluación subjetiva de la calidad para *streaming* de vídeo está basada en *EvalVid* [44]. Esta herramienta evalúa las diferencias entre las imágenes originales y las resultantes tras su transmisión a través de la red. Sin embargo, debido a que *EvalVid*

48 | Calidad de Experiencia (QoE)

exige disponer de ambas imágenes, su adopción está limitada –sólo para escenarios que permitan realizar estimaciones de calidad *en diferido (offline)*–.

En [23] se describen dos modelos de aprendizaje para la predicción de calidad de vídeo *en directo (on-line)* utilizando *EvalVid*. Teniendo en cuenta las características temporales (e.g. movimiento) y espaciales (e.g. bordes, brillo) de los vídeos a evaluar, éstos se clasifican en tres grupos: *movimiento lento* (SM, *Slight Movement*), *paseo apacible* (GM, *Gentle Walking*) y *movimiento rápido* (RM, *Rapid Movement*). La calidad del vídeo viene dada por la Ecuación 3.2.

$$MOS = \frac{a_1 + a_2 FR + a_3 \ln(SBR)}{1 + a_4 PER + a_5 (PER)^2} \quad (3.2)$$

donde FR (*Frame Rate*, en fps) es el número de cuadros por segundo, SBR (*Sending Bit Rate*, en Kbps) es la tasa binaria de envío y PER (*Packet Error Rate*) es la tasa de error de paquetes. Los coeficientes a_i para los tres grupos considerados se muestran en la Tabla 3.3.

La Figura 3.2 muestra la evolución de la calidad experimentada (MOS) al variar tanto la tasa binaria de envío (SBR) como la tasa de error de paquetes (PER) en el caso concreto de un vídeo del grupo RM con 15 fps.

3.2.3 Función de utilidad para tráfico *streaming* de vídeo sobre TCP

Por sencillez, se ha supuesto que el tráfico de un *streaming* de vídeo sobre TCP se almacena en el receptor siguiendo una estrategia de doble umbral [45]. El primer umbral (*UTA, Umbral de Tiempo Almacenado*) define el tamaño mínimo del *buffer* (en unidades de tiempo) antes de que comience la reproducción del vídeo. El segundo umbral define el tamaño del *buffer* a partir del cual el receptor solicita al servidor que deje de enviarle datos.

Para este servicio se supondrá que la calidad subjetiva sólo está afectada por la falta de datos en el *buffer* del receptor (*buffer underruns*), lo que se traduce en una interrupción

<i>Coefficientes</i>	<i>SM</i>	<i>GW</i>	<i>RM</i>
a_1	4.5796	3.4757	3.0946
a_2	-0.0065	0.0022	-0.0065
a_3	0.0573	0.0407	0.1464
a_4	2.2073	2.4984	10.0437
a_5	7.1773	-3.7433	0.6865

Tabla 3.3. Definición de los parámetros a_i para streaming de vídeo.

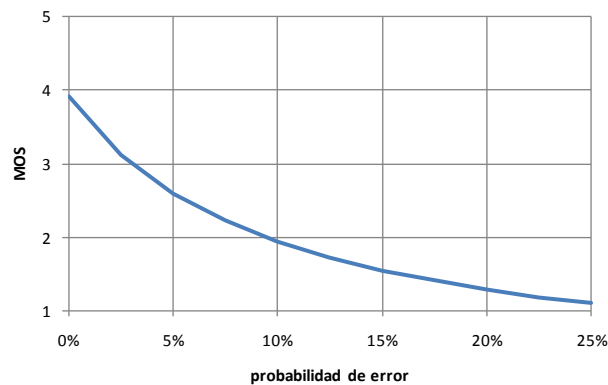


Figura 3.2. Ejemplo de función de utilidad para tráfico *streaming* de vídeo (RM, 15 fps, SBR 1Mbps).

de la reproducción. Esta suposición se debe a que, gracias a la fiabilidad en la entrega del protocolo TCP, todos los datos recibidos son correctos y por tanto la QoE no dependerá de los errores.

Dado que las ausencias de datos en el *buffer* se deben por lo general a eventos de duración variable –e.g. episodios de congestión o pérdidas en el interfaz inalámbrico [46] [47] [48]–, no tiene sentido definir una métrica de la calidad instantánea a partir de encuestas a usuarios para secuencias con duraciones del orden de decenas o centenas de segundos. Por tanto, se ha diseñado una función de utilidad artificial que estima el valor de MOS basándose únicamente en la situación instantánea del *buffer* de recepción. Los criterios para diseñar esta función han sido:

- Al comenzar la reproducción, los usuarios están acostumbrados a un retardo

50 | Calidad de Experiencia (QoE)

inicial debido a la acumulación de datos en el buffer del receptor (*playout buffer*), por lo que éste no debería considerarse como un factor negativo al estimar el valor instantáneo de MOS.

- La falta de datos en el buffer y su posterior relleno hasta el umbral de reproducción deben reducir abruptamente la nota de opinión media a su valor mínimo, ya que paralizar la reproducción tiene un impacto significativo en la opinión instantánea del usuario final.
- Al comenzar de nuevo la reproducción tras una pausa por falta de datos, el valor de MOS no debería aumentar rápidamente. Así, se reflejaría el hecho de que, tras una interrupción en el vídeo, el usuario necesita un cierto tiempo para recuperar el grado de satisfacción. Además, esta aproximación evita oscilaciones no deseadas en la función de utilidad.
- Si el nivel de ocupación del *playout buffer* está por debajo del primer umbral (UTA), entonces el valor de MOS debería disminuir, aunque no abruptamente, de forma que refleje la probabilidad de que pueda haber, en un futuro cercano, falta de datos en el *buffer*.

La Figura 3.3 (gráfica superior) muestra un ejemplo del MOS de acuerdo con el comportamiento deseado dada la evolución temporal de la ocupación instantánea del *playout buffer* (gráfica inferior).

Teniendo en cuenta las consideraciones anteriores, la función de utilidad diseñada para *streaming* de vídeo sobre TCP viene dada por las Ecuaciones 3.3 y 3.4.

$$MOS[k] = \begin{cases} 5 & \text{si } TA \geq UTA \\ 1 + (5 - 1) \times TA/UTA & \text{si } 0 < TA < UTA \\ 1 & \text{si } TA \leq 0 \end{cases} \quad (3.3)$$

$$AVGMOS[k] = \begin{cases} MOS[k] & \text{si } MOS[k] < AVGMOS[k - 1] \\ \alpha AVGMOS[k - 1] + (1 - \alpha)MOS[k] & \text{en otro caso} \end{cases} \quad (3.4)$$

donde *TA* es el *Tiempo Almacenado*, i.e. el tiempo de reproducción correspondiente a los paquetes en el *playout buffer* del receptor.

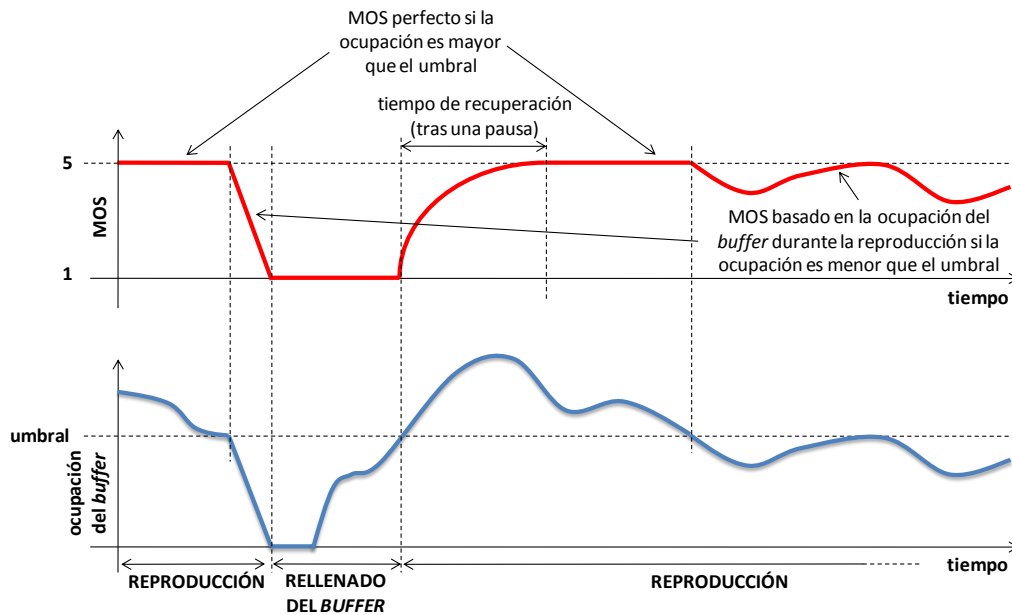


Figura 3.3. Evolución artificial de MOS para streaming de vídeo sobre TCP.

El parámetro α permite ajustar la molestia experimentada por el usuario final tras una falta de datos. Se define el *Tiempo de Recuperación* (TR) como el tiempo que transcurre desde que comienza de nuevo la reproducción hasta que el usuario vuelve a estar satisfecho ($MOS = 5 - \varepsilon$). Este tiempo depende del parámetro α siguiendo la regla expresada en la Ecuación 3.5.

$$\alpha = \left(\frac{TR}{PAM} \right)^{\frac{1}{4}} \sqrt{\frac{\varepsilon}{4}} \quad (3.5)$$

donde *PAM* es el *Período de Actualización de MOS*, i.e. intervalo temporal durante el que se estima el valor MOS expresado en segundos. A modo de ejemplo, si se pretende tener un tiempo de recuperación $TR = 5$ s, el valor de α sería 0.48 (considerando $PAM = 1$ s y $\varepsilon = 0.1$).

3.2.4 Función de utilidad para tráfico web

Esta función se obtuvo tras realizar una encuesta [25] [16] en la que 52 voluntarios puntuaron la calidad percibida tras descargar diferentes páginas web con diferentes velocidades de transmisión (hasta 2 Mbps).

Para ello, se utilizó el navegador *Mozilla Firefox* [49] para descargar un total de 14 páginas web, incrementando en cada una el tiempo de descarga (variando entre 600 milisegundos hasta 60 segundos aproximadamente). El tiempo de descarga (i.e. el período transcurrido entre el instante en que se realiza la petición hasta el momento en que la página ha sido completamente descargada) se controla de forma artificial por medio de un servidor HTTP hecho a medida para este propósito. Este servidor implementa los métodos GET y HEAD de HTTP/1.1, y controla el tiempo de descarga mediante la parada momentánea de su ejecución (usando la función *usleep* de Linux) cuando tiene que enviar un paquete. El servidor se ejecutó en la misma máquina que el cliente para evitar otras fuentes de retardo.

Para evitar la influencia del contenido de la página en la percepción del usuario, se utilizaron páginas web con el mismo formato y un contenido similar. En concreto, se crearon 14 documentos con información (texto, enlaces y fotos) sobre monumentos. Los sujetos eran 52 estudiantes de Ingeniería de Telecomunicación con una edad media entre 21 y 22 años, muy familiarizados con la navegación web a través de la red WLAN del campus. Además, se ha supuesto que la duración total de la sesión de navegación, con sólo 14 páginas, no afecta a la tolerancia de los usuarios respecto al tiempo de descarga.

La Figura 3.4 muestra los resultados obtenidos –tanto el valor medio como su desviación estándar– en función del tiempo de descarga de las páginas web.

Sin embargo, si se requiere una estimación instantánea de la calidad percibida por el usuario no se puede utilizar el tiempo de descarga, ya que éste puede ser muy elevado. Por ello, la Ecuación 3.6 muestra la relación entre la tasa binaria –que sí se puede calcular *online*– y el valor de MOS, realizando un ajuste mediante una función lorentziana y suponiendo un tamaño medio de páginas web de 130 KBytes [50]. Este tamaño medio de página permite establecer una correspondencia entre el tiempo de descarga y la tasa

binaria requerida.

$$MOS = 5 - \frac{578}{1 + \left(\frac{THR + 541.1}{45.98}\right)^2} \quad (3.6)$$

donde *THR* es la tasa binaria (*throughput*) en kbps. La Figura 3.5 representa gráficamente esta relación.

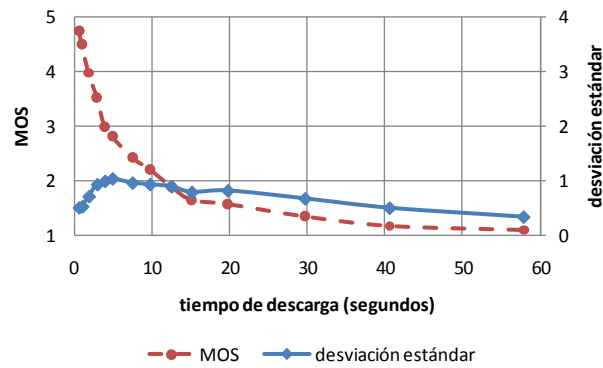


Figura 3.4. Función de utilidad para tráfico web en función del tiempo de descarga.

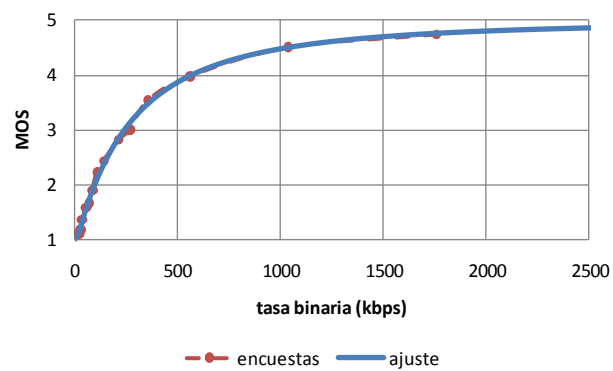


Figura 3.5. Función de utilidad para tráfico *web* en función de la tasa binaria.

3.2.5 Función de utilidad para tráfico FTP

Y. Sato *et al* [24] estudiaron la estimación de la calidad subjetiva de las comunicaciones en la transmisión de grandes volúmenes de datos. Para ello encuestaron a 91 usuarios, con tamaños típicos de ficheros –con un valor medio de 40 MBytes, calculado a partir de los diez ficheros más descargados de un sitio web muy popular en Japón [51]– y considerando anchos de banda similares a los utilizados en la tecnología FTTH (*Fiber To The Home*) –valores medios (de 6 a 18 Mbps) de los anchos de banda publicados en [52], que es un sitio web de comparación de velocidades entre diferentes ISPs– con un máximo de 25 Mbps.

Debido a la diferente tecnología de acceso utilizada, se ha realizado un escalado de la tasa binaria máxima para considerar el ancho de banda más escaso que utilizan las redes IEEE 802.11e. Considerando la capa física de IEEE 802.11b, se ha tomado 1 Mbps como la velocidad objetivo para el servicio FTP, i.e. el valor por encima del cual el usuario final estaría plenamente satisfecho. Por consiguiente, se ha utilizado un factor de escalado igual a 25. Con estas suposiciones, la función de utilidad para tráfico FTP se expresa según la Ecuación 3.7.

$$MOS = 6.5 \times THR - 0,54 \quad (3.7)$$

donde *THR* es la tasa binaria (*throughput*) expresada en Mbps. La Figura 3.6 muestra esta función de utilidad.

3.3 Metodología para la evaluación de rendimiento

La metodología utilizada para evaluar de forma unificada el rendimiento de los diferentes esquemas desarrollados se basa en el uso de una métrica común para los diferentes servicios, la Nota Media de Opinión, de forma que se pueda medir y comparar su eficiencia. Esto no sería posible con medidas como la tasa binaria, el retardo, la probabilidad de pérdidas, etcétera, ya que cada servicio se ve influenciado por estos

3.3. Metodología para la evaluación de rendimiento | 55

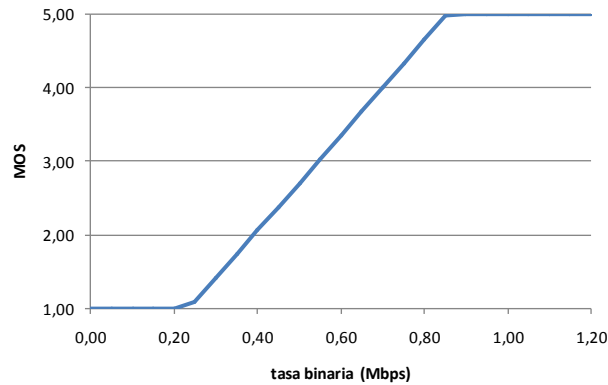


Figura 3.6. Función de utilidad para tráfico FTP.

indicadores de una manera diferente. Por ejemplo, la calidad percibida en las llamadas de voz se ve afectada de forma notable por el retardo mientras que éste tiene poca influencia en las descargas de ficheros. Otros servicios, como la descarga progresiva de vídeo sobre TCP (e.g. YouTube), tienen una relación compleja entre la calidad subjetiva y los indicadores objetivos de rendimiento habituales (velocidad de descarga, retardo, *jitter*, pérdidas, etcétera). Estos razonamientos originan la necesidad de una métrica común para evaluar la calidad de los diferentes servicios. Esta necesidad, unida a la importancia de considerar la experiencia del usuario final, ha motivado la adopción de la Nota Media de Opinión como métrica de referencia.

Además de evaluar la calidad de cada servicio, resulta interesante comprobar si unos servicios sufren una degradación excesiva mientras que otros perciben buena calidad, i.e. si el sistema es justo o no. Existen índices de justicia habituales como la varianza (Ecuación 3.8), el coeficiente de varianza (Ecuación 3.9) y el cociente mín – máx (Ecuación 3.10). En estas expresiones x_i es la métrica de asignación de recursos a analizar, n es el número de medidas y μ es el valor medio de x_i .

$$\mu = \frac{1}{n} \sum_{i=1}^n x_i$$

$$\text{varianza} = \frac{1}{n-1} \sum_{i=1}^n (x_i - \mu)^2 \quad (3.8)$$

$$\text{coeficiente de varianza} = \frac{\text{varianza}}{\text{media}} \quad (3.9)$$

$$\text{cociente mín - máx} = \frac{\text{mín}_j \{x_j\}}{\text{máx}_i \{x_i\}} \quad (3.10)$$

En general, cualquier medida o índice de justicia debe verificar las siguientes propiedades:

- *Independencia respecto al tamaño de la población.* El índice elegido debería ser válido con independencia del tamaño de la población o número de medidas. Este requisito lo cumplen los índices anteriores.
- *Independencia de la escala y la métrica.* El índice debería ser independiente de la escala, i.e. la unidad de medida no debería importar. Esta característica descarta el uso de la varianza como índice de justicia, si bien el coeficiente de varianza y el cociente mín – máx sí verifican esta propiedad.
- *Acotación.* El índice debería estar acotado entre 0 y 1, i.e. un sistema totalmente justo debería tener una justicia de 1 y uno totalmente injusto debería tener un índice de 0. De esta forma, la justicia se puede expresar como un porcentaje y así permite definir un índice de discriminación como $(1 - \text{índice de justicia})$. El coeficiente de varianza tiene un rango entre 0 e infinito, no siendo sencillo de interpretar qué nivel de justicia implica un valor concreto. El cociente mín – máx sí respeta esta condición.
- *Continuidad.* El índice debería ser continuo. Cualquier pequeño cambio en la asignación de recursos debería mostrarse en el índice de justicia. El cociente mín – máx no cumple este punto, ya que los cambios que no afecten a los valores extremos (mínimo y máximo) no modificarán dicho índice.

Como se puede comprobar ninguno de los índices definidos anteriormente (Ecuaciones 3.8, 3.9 y 3.10) cumplen con todas las propiedades identificadas.

Por este motivo, Jain *et al* [53] definieron un nuevo índice de justicia que verifica

todas las propiedades. Este índice se suele utilizar para medir la justicia en la asignación de recursos en términos de tasa binaria, si bien Jain lo definió de forma genérica para cualquier métrica x . La expresión del índice de justicia de Jain viene dada por la Ecuación 3.11.

$$f(x) = \frac{(\sum_{i=1}^n x_i)^2}{n \sum_{i=1}^n x_i^2} \quad (3.11)$$

Esta función devuelve un valor entre 0 y 1 (realmente entre $1/n$ y 1), indicando que el sistema bajo evaluación es más justo cuanto mayor es su cuantía. Su fórmula se puede desarrollar [53] obteniéndose la Ecuación 3.12.

$$f(x) = \frac{(\frac{1}{n} \sum_{i=1}^n x_i)^2}{\frac{1}{n} \sum_{i=1}^n x_i^2} = \frac{E^2[x]}{E[x^2]} = \frac{E^2[x]}{E^2[x] + E[x^2] - E^2[x]} = \frac{\mu^2}{\mu^2 + \sigma^2} \Rightarrow$$

$$f(x) = \frac{1}{1 + \sigma^2 / \mu^2} \quad (3.12)$$

donde μ es la media de las muestras y σ^2 es su varianza. Esta expresión indica que la función será más sensible frente a variaciones (recogidas en σ^2) cuanto menor sea su relación con el valor medio de las muestras (μ).

Teniendo en cuenta la Ecuación 3.12, la métrica elegida ha sido $x = \overline{MOS}_i - 1$, donde \overline{MOS}_i es el valor medio de MOS para la clase de tráfico i . El término -1 desplaza la escala MOS al rango [0, 4] reduciéndose así el valor medio de las muestras x_i sin modificar su varianza. De esta forma se aumenta la sensibilidad del *índice de justicia* respecto a la diferencia entre muestras.

3.4 Conclusiones

En este capítulo se ha presentado el concepto de *Calidad de Experiencia* que se utilizará en los diversos mecanismos diseñados en la tesis. Los conceptos y funciones

58 | Calidad de Experiencia (QoE)

expuestos servirán de base para los capítulos posteriores, considerando como hipótesis de partida que la incorporación de criterios de QoE en estos algoritmos mejorará de forma significativa la calidad percibida por los diferentes usuarios de la red en escenarios en los que el tráfico proviene de diferentes servicios.

Para simplificar la gestión de los servicios considerados, se ha expuesto una clasificación de los mismos introduciendo el concepto de *Clase de Tráfico (CT)*, agrupando servicios con el mismo tipo de requisitos de QoS. Esta clasificación considera las siguientes CTs: *conversacional*, *streaming*, *TCP de tiempo real*, *interactiva* y *de fondo*.

Debido a la necesidad de estimar de forma automática –y en tiempo real– la calidad percibida por los usuarios, se ha presentado el concepto de *función de utilidad*, definida como aquella que permite mapear indicadores objetivos del rendimiento (QoS) en la calidad subjetiva que experimentaría un usuario (QoE). Se han definido funciones de utilidad para el conjunto de servicios elegidos, algunas basadas en la bibliografía existente (VoIP, *streaming* de vídeo, FTP) y otras diseñadas expresamente para las CTs *interactiva* y *TCP de tiempo real* (tráfico web y *streaming* de vídeo sobre TCP).

Por último, se ha descrito una metodología unificada para evaluar el rendimiento de las diferentes mejoras que se describirán en los capítulos posteriores. Esta evaluación se basa en la Nota de Opinión Media, métrica común para los diferentes servicios permitiendo así su comparación. También se ha definido un índice que permite estimar la *justicia* de un sistema, en términos de calidad experimentada, para así evaluar las prestaciones de los algoritmos propuestos en escenarios que contemplen mezcla de distintas clases de tráfico.

Capítulo 4

Una variante del Planificador– Conformador Monolítico para HCCA

El Planificador–Conformador Monolítico (MSS [20][54], *Monolithic Shaper–Scheduler*) fue propuesto para gestionar tareas de tiempo real en entornos multitarea. Inspirado en MSS, en este trabajo se propone una variante de esta disciplina de planificación para ser utilizada por el coordinador híbrido en HCCA.

El funcionamiento de este planificador se expone en la Sección 4.1 mientras que su adaptación al mecanismo HCCA del estándar IEEE 802.11e se detalla en la Sección 4.2. La Sección 4.3 describe la evaluación de su rendimiento y finalmente, las conclusiones extraídas se resumen en la Sección 4.4.

4.1 Funcionamiento del planificador MSS

El Planificador–Conformador Monolítico de referencia está descrito en [54]. Este planificador ha sido elegido por la idoneidad de sus características, ya que garantiza un ancho de banda mínimo y un retardo acotado. Además, introduce la menor latencia de entre todos los planificadores de la familia GPS (*General Processor Sharing*), tal y como se demuestra en [54]. En el contexto de redes de paquetes, estos planificadores son conocidos como *Packet Fair Queueing* [55] [56] [57], siendo una aproximación a la familia GPS en la que la unidad mínima de asignación de recursos es un paquete.

MSS puede operar tanto con flujos con tráfico conformado como no conformado. Si el flujo es conformado, un regulador *dual-leaky-bucket* [58] es el encargado de conformar la generación de tráfico. Este tipo de conformadores está formado por dos componentes: un *leaky bucket* de tasa de bit sostenible (SBR, *Sustainable-Bit-Rate*) y un segundo *leaky bucket* de tasa de bit de pico (PBR, *Peak-Bit-Rate*). El *bucket* SBR construye el comportamiento del flujo a largo plazo (i.e. tasa binaria media) mientras que el *bucket* PBR impone una limitación en la tasa de transmisión durante las ráfagas (i.e. tasa binaria de pico y duración de la ráfaga). Ambos reguladores, para un flujo i , se definen por sus tamaños ($B_{i,s}$ y $B_{i,p}$, respectivamente) y sus tasas de generación de turnos (*tokens*) ($\rho_{i,s}$ y $\rho_{i,p}$, respectivamente). En la Figura 4.1 se representa gráficamente este conformador.

Cuando MSS recibe un paquete, el flujo correspondiente recibe una etiqueta que incluye un tiempo de comienzo (*tiempo elegible*) y un tiempo de fin (*tiempo de vencimiento*). Dicho paquete se guarda hasta que llega su tiempo elegible. El planificador ordena la transmisión de los paquetes elegibles de acuerdo a sus tiempos de vencimiento. Hecho esto, se aplica la política de selección de paquetes de Primero el paquete Elegible con Menor tiempo de Vencimiento (SEFF, *Smallest Eligible Finishing time First*). Dado un paquete de un flujo i con una tasa garantizada R_i , sus tiempos (elegible y de vencimiento) se calculan como los tiempos de comienzo y fin para transmitir dicho paquete en una red ideal en la que el flujo fuera servido por un enlace dedicado de capacidad R_i . Más concretamente, el tiempo elegible de un paquete se calcula como el máximo entre el tiempo de llegada del paquete y el tiempo de vencimiento del paquete anterior del mismo flujo. El tiempo de vencimiento se calcula como la suma entre el tiempo de llegada del paquete y el tiempo que implicaría su transmisión en el sistema

4.1. Funcionamiento del planificador MSS | 61

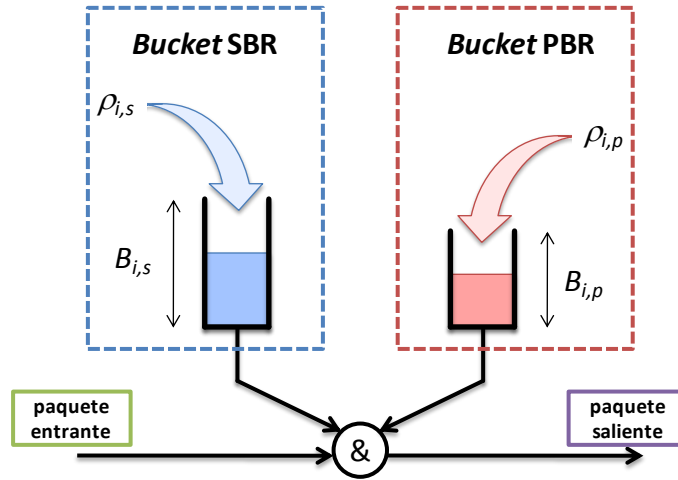


Figura 4.1. Conformador dual-leaky-bucket.

ideal, i.e. N_i/Γ_i , donde N_i representa la longitud del paquete k -ésimo del flujo i y Γ_i es la tasa mínima de transmisión del nivel físico. Para comprobar el instante en que el flujo se vuelve elegible, éste recibe un nuevo sello de tiempo F_i^k cuando recibe el paquete k -ésimo de longitud N_i en el instante $t_{i,a}^k$, de acuerdo con la Ecuación 4.1.

$$F_i^k = \max(t_{i,a}^k, F_i^{k-1}) + \frac{N_i}{R_i} \quad (4.1)$$

La operación de MSS consta de cuatro actividades básicas:

- *Test de conformidad virtual*
- *Encolado de flujo*
- *Desencolado de flujo*
- *Selección de flujo*

El test de conformidad virtual evalúa la posibilidad de que la transmisión inmediata de un nuevo paquete viole uno de los perfiles de los *buckets* SBR o PBR. La actividad de

62 | Una variante del Planificador–Conformador Monolítico para HCCA

encolado de flujo actualiza el sello de tiempo del flujo (F_i^k). Estas dos tareas se ejecutan en el momento en que un nuevo paquete llega a la cabeza de la cola, bien por la llegada de un paquete a una cola vacía, bien por la transmisión de un paquete en una cola con más paquetes. La actividad de desencolado de flujo comprueba cuando un flujo se vuelve o no elegible (i.e. el siguiente paquete a transmitir de dicho flujo es o no elegible). Si un flujo no es elegible, se realiza una recalibración del sello de tiempo para evitar un crecimiento incontrolado de la distancia entre dicho sello de tiempo y la referencia real de tiempo. Finalmente, el procedimiento de selección de flujo identifica el flujo del siguiente paquete a transmitir escogiendo aquel que sea elegible y que tenga el sello de tiempo mínimo. En caso de que ninguno sea elegible, se seleccionará entre los flujos que sean conformes virtualmente. Tanto la actividad de desencolado de flujo como el procedimiento de selección se realizan tras la transmisión de un paquete.

Las Figuras 4.2 y 4.3 muestran los diagramas de flujo de estas operaciones. Detalles adicionales pueden obtenerse en [54].

El planificador MSS pertenece a la familia de planificadores de tasa-latencia (LR, *Latency-Rate*) [59], ya que es una modificación de una disciplina de planificación de Reloj Virtual (*Virtual Clock* [60]) con las mismas garantías de ancho de banda y latencia. Un servidor LR se define como un planificador cuyo servicio ofrecido a un flujo es al menos igual a su tasa reservada (durante cualquier intervalo de tiempo posterior a su latencia Θ y siempre que la suma de las tasas reservadas no superen la capacidad del servidor). Dado que el planificador MSS es un servidor LR con un regulador *dual-leaky-bucket*, el análisis realizado en [59] es aplicable.

De esta manera, en un planificador MSS el retardo máximo D_i que puede experimentar un paquete perteneciente a un flujo i –conformado por un regulador *dual-leaky-bucket*– verificará que [59]:

$$D_i \leq \left(\frac{\rho_{i,p} - R_i}{\rho_{i,p} - \rho_{i,s}} \right) \left(\frac{B_{i,s}}{R_i} \right) + \Theta_i^{MSS} - \frac{N_i}{\rho_{i,s}} \quad (4.2)$$

donde Θ_i^{MSS} es la latencia del planificador para el flujo i . En [54] se demuestra que Θ_i^{MSS} es la menor posible de entre todos los planificadores GPS (*Generalized Processor Sharing*), y además verifica que:

4.1. Funcionamiento del planificador MSS | 63

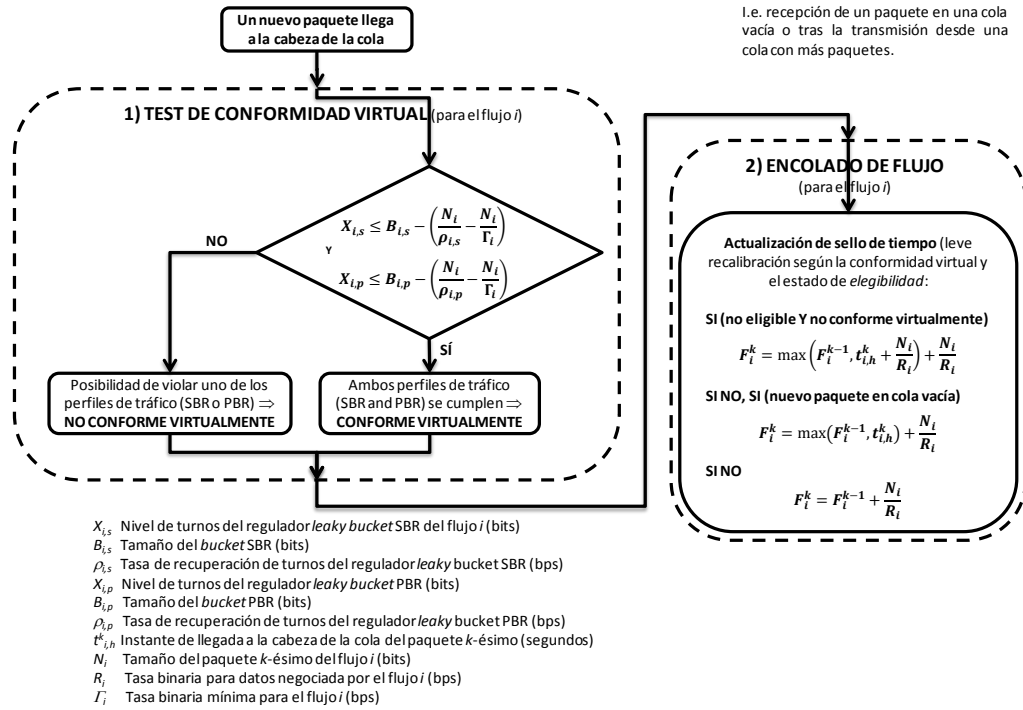


Figura 4.2. Actividades de *test de conformidad virtual* y *encolado de flujo* del planificador MSS.

$$\theta_i^{MSS} \leq \frac{N_i}{R_i} + \frac{N_{max}}{\Gamma_i} \quad (4.3)$$

donde N_{max} es el tamaño del paquete –SDU (*Service Data Unit*)– máximo de todos los flujos.

Como se deduce de la Ecuación 4.2, el retardo máximo está influenciado por el tamaño del *bucket* SBR. Francini *et al* [54] proporcionaron los tamaños mínimos tanto del *bucket* SBR ($B_{i,s}$) como PBR ($B_{i,p}$) para los que se garantiza que el planificador MSS no provoca desbordamientos.

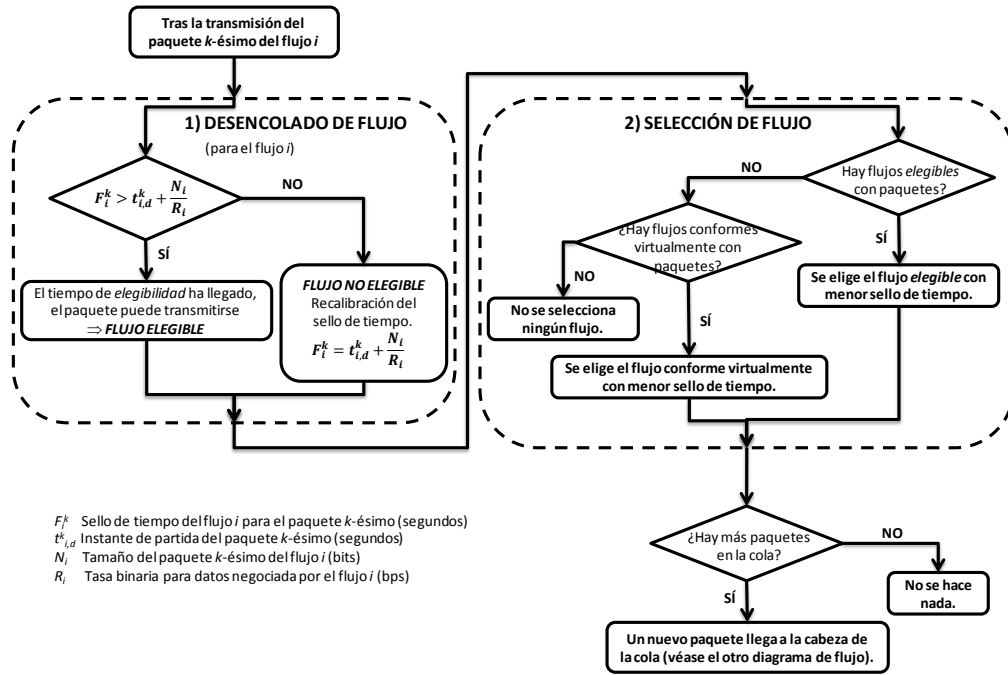


Figura 4.3. Actividades de desencolado de flujo y selección de flujo en el planificador MSS.

$$B_{i,s}^{min} = \left(2N_i \left(\frac{1}{R_i} - \frac{1}{\Gamma_i} \right) + \frac{N_{max}}{\Gamma_i} \right) \cdot \rho_{i,s} \quad (4.4)$$

$$B_{i,p}^{min} = \left(2N_i \left(\frac{1}{\rho_{i,p}} - \frac{1}{\Gamma_i} \right) + \frac{R_i}{\rho_{i,p}} \cdot \frac{\Gamma_i - \rho_{i,p}}{\Gamma_i - R_i} \cdot \frac{N_{max}}{\Gamma_i} \right) \cdot \rho_{i,p} \quad (4.5)$$

Por tanto, como en un planificador MSS la tasa de servicio garantizada R_i es igual a la tasa $\rho_{i,s}$ del bucket SBR, y como la latencia θ_i^{MSS} está acotada –Ecuación 4.3–, suponiendo que el tamaño del bucket SBR es el mínimo dado por la Ecuación 4.5 se puede simplificar la expresión del retardo máximo:

$$D_i \leq 2N_i \left(\frac{1}{R_i} - \frac{1}{\Gamma_i} \right) + \frac{2 \times N_{max}}{\Gamma_i} \quad (4.6)$$

4.1. Funcionamiento del planificador MSS | 65

En la expresión anterior, no se ha considerado el retardo extra introducido por la capa MAC de IEEE 802.11e debido a cabeceras y procedimientos de contención, ya que su valor (típicamente menor que 1 ms, véase el Apéndice B) es generalmente despreciable frente a esta cota. Sin embargo, siempre es posible su inclusión para así tener una mejor estimación.

Teorema 1 – Cumplimiento del requisito IEEE 802.11e para planificadores HCCA

El planificador MSS satisface el requisito impuesto por el estándar IEEE 802.11e para los planificadores a utilizar con el mecanismo de acceso HCCA (Ecuación 2.5), asignando a Δ_i un valor superior a la latencia de MSS (θ_i^{MSS}). Nótese que Δ_i será el *intervalo de servicio máximo* (MSI_i) o la *cota de retardo máximo* (D_i), dependiendo de lo especificado en el TSPEC (véase la Sección 2.3.2).

Demostración

De la definición de un servidor LR [59], el servicio total W_i^{MSS} (expresado en bits transmitidos) proporcionado por un planificador MSS al flujo i durante un período ocupado (comenzado en τ) tras un tiempo t se puede expresar como:

$$W_i^{MSS}(\tau, t) \geq \max\left(0, R_i(t - \tau - \theta_i^{MSS})\right) \quad (4.7)$$

El servicio ofrecido –en unidades temporales– se puede calcular incluyendo la tasa de transmisión del nivel físico Γ_i en la Ecuación 4.7. Suponiendo, sin pérdida de generalidad, que $t > \tau + \theta_i^{MSS}$ y que la tasa de transmisión Γ_i incluye las tramas de control MAC y los espacios entre tramas, la Ecuación 4.7 se puede expresar como:

$$W_i^{MSS}(\tau, t) \geq \frac{R_i}{\Gamma_i}(t - \tau - \theta_i^{MSS}) = \frac{N_i/\Gamma_i}{N_i/R_i}(t - \tau - \theta_i^{MSS}) = \frac{t - \tau - \theta_i^{MSS}}{\frac{N_i}{R_i}} \cdot tx(N_i, \Gamma_i) \quad (4.8)$$

Esta expresión coincide con la Ecuación 2.6, identificando t como t_1 , τ como t_2 y especificando $\Delta_i \geq \theta_i^{MSS}$, lo que demuestra que MSS verifica el requisito exigido a los planificadores HCCA.

4.2 Adaptación para HCCA

En el Planificador–Conformador Monolítico, el servidor mantiene toda la información necesaria para elegir el flujo adecuado tras la transmisión de cada paquete. Esto es posible porque el servidor encola los paquetes de todos los flujos de tráfico antes de elegir qué flujo debe transmitir.

Sin embargo, en las redes IEEE 802.11e, el planificador (ubicado en el QAP) sólo encola paquetes enviados *hacia* las QSTAs (dirección descendente), pero no *desde* las QSTAs (dirección ascendente). Los paquetes de flujos originados por las QSTAs se encolan en las propias QSTAs, de forma que el planificador no tiene información sobre esos flujos *a priori*.

La información necesaria para cada flujo consiste en:

- La existencia o no de paquetes encolados
- Su sello de tiempo
- Su condición de elegibilidad
- Su estado de conformidad virtual

Esta información se puede calcular y mantener en los flujos descendentes de forma sencilla, pero las QSTAs deben proporcionarla para los flujos ascendentes de forma que el planificador tenga toda la información necesaria para la selección de flujo.

Para saber si hay paquetes esperando a ser transmitidos, se puede utilizar el mecanismo opcional de *piggybacking* del tamaño de las colas (subcampo *Queue Size* presente en las tramas de datos QoS [2]). El resto de la información se debe añadir como señalización extra en la cabecera de las tramas de datos o de confirmación. Esta información requiere pocos bytes (según la precisión requerida para los sellos de tiempo), por lo que la sobrecarga introducida puede ser despreciable.

Una posible solución que evitaría incluir este tipo de información en las tramas entre las estaciones y el punto de acceso consistiría en que éste último estimara estos datos. La existencia o no de paquetes en la cola de una estación se puede evaluar a partir del campo *Queue Size* anteriormente comentado, siempre que la estación envía una trama de datos. En el caso de que la cola quedara vacía, el QAP podría utilizar el Máximo Intervalo de Servicio (parámetro MSI de TSPEC) para suponer la llegada de un nuevo paquete.

En ambos casos, el QAP puede estimar los instantes de desencolado y la llegada de un nuevo paquete a la cabeza de la cola (ya sea suponiendo la recepción de un nuevo paquete en una cola vacía, ya sea por la transmisión de un paquete en una cola con más paquetes), lo que permite al QAP calcular las variables de estado asociadas al flujo (sello de tiempo, condición de *elegibilidad*, estado de *conformidad virtual*). De esta manera, no haría falta ningún tipo de señalización asociada al estado de los flujos ascendentes.

Sin embargo, la estimación de la llegada de un nuevo paquete a una cola vacía podría reducir el rendimiento del planificador debido a los retardos introducidos (no se conoce con certeza el momento de llegada de un nuevo paquete en una QSTA) y a falsos positivos (creencia de llegada de un nuevo paquete cuando realmente no ha ocurrido). No obstante, la llegada de paquetes a colas vacías suele implicar una baja carga de tráfico en el sistema (dado que la QSTA ha podido transmitir todos sus paquetes anteriormente), situación en la que el rendimiento del planificador no es crítico. Si la carga fuese más elevada, las colas tenderían a tener varios paquetes y su longitud se indicaría de forma exacta mediante *piggybacking*, por lo que el rendimiento debería acercarse al obtenido usando señalización para los flujos ascendentes.

Una vez que el QAP dispone de esta información para los flujos de ambas direcciones, es capaz de realizar las operaciones necesarias. Si el flujo seleccionado es de dirección descendente, el QAP enviará sus paquetes directamente. Sin embargo, si se trata de un flujo de dirección ascendente, primero enviará un mensaje de sondeo a la QSTA antes de la transmisión del paquete. Es importante señalar que, dado que todos los flujos (ascendentes y descendentes) se tratan de forma similar en el proceso de selección, la injusticia intrínseca de las redes 802.11 entre ambos sentidos [61] se resuelve de forma automática.

4.3 Evaluación de rendimiento

En este apartado se describen, en primer lugar, las condiciones experimentales bajo las que se llevó a cabo la campaña de simulación para evaluar el comportamiento de la variante del planificador MSS propuesta para redes inalámbricas con soporte de HCCA. Posteriormente se analizan los resultados obtenidos, comparándolos con los correspondientes a otros planificadores seleccionados como referencia. En concreto, en el estudio se incluyen los siguientes planificadores:

- El especificado en el estándar IEEE 802.11e (véase la Sección 2.3.4), debido a su amplio uso como referencia respecto al rendimiento de otros planificadores.
- El esquema WTTP, ya que cumple con los requisitos impuestos por el estándar (comentados en la Sección 2.3.3) y permite gestionar tanto tráfico CBR como VBR.
- La disciplina FHCF, que si bien sus autores [8] no demuestran que cumpla este requisito, resulta una alternativa interesante ya que utiliza todo el ancho de banda disponible en el canal.

El análisis realizado se basa –como se comentó en la Sección 3.3– en la estimación de la calidad experimentada por los usuarios de los diferentes servicios así como la justicia entre servicios de los esquemas estudiados. Para ello, se ha definido como calidad mínima aceptable aquella que consigue un valor de MOS igual a 3.5 (posteriormente, en el Capítulo 5, se definirá el rango de calidad deseable como al intervalo de valores de MOS entre 3.5 y 4). La máxima carga soportada que verifica este requisito de aceptabilidad QoE ($MOS \geq 3.5$) y el índice de justicia permiten la evaluación del rendimiento de la solución propuesta, así como su comparación con otros planificadores de referencia.

4.3.1 Condiciones experimentales

Los experimentos realizados se han llevado a cabo utilizando la herramienta *Network Simulator 2* (NS-2 [62]), conjuntamente con algunos módulos añadidos. Para la realización de las campañas de simulación se ha utilizado el entorno *ANSWER* [63], que permite la automatización de los experimentos realizados. Éste se configuró de forma que el nivel de confianza para todos los experimentos fuese del 95 por ciento (con un máximo de 10 réplicas por punto de carga simulado), criterio utilizado para garantizar la relevancia estadística de los resultados obtenidos –los intervalos de confianza se incluyen en las gráficas mediante barras de error–. La duración de las simulaciones ha sido de 300 segundos con un intervalo inicial de calentamiento (*warm-up*) de 30 segundos sin recogida de estadísticas.

La simulación del mecanismo HCCA bajo NS-2 ha sido realizada mediante el entorno de trabajo *ns2hcca* [64]. Además, el planificador de paquetes utilizado ha sido la variante propuesta del Planificador–Conformador Monolítico (MSS) adaptada para el mecanismo HCCA tal como se ha descrito en este capítulo, siendo ésta implementada sobre el módulo *ns2hcca*.

El escenario simulado (representado en la Figura 4.4) ha sido un entorno heterogéneo, es decir, una red cableada y una red inalámbrica conectadas a través de un punto de acceso. Este AP se conecta a varias estaciones y a su vez a la parte cableada, de forma que las comunicaciones se llevan a cabo entre las estaciones inalámbricas y servidores o equipos en la red cableada.

Respecto al interfaz inalámbrico, se utiliza la capa física HR-DSSS (*High Rate – Direct Sequence Spread Spectrum*), también conocida como 802.11b, cuyos parámetros se resumen en la Tabla 4.1.

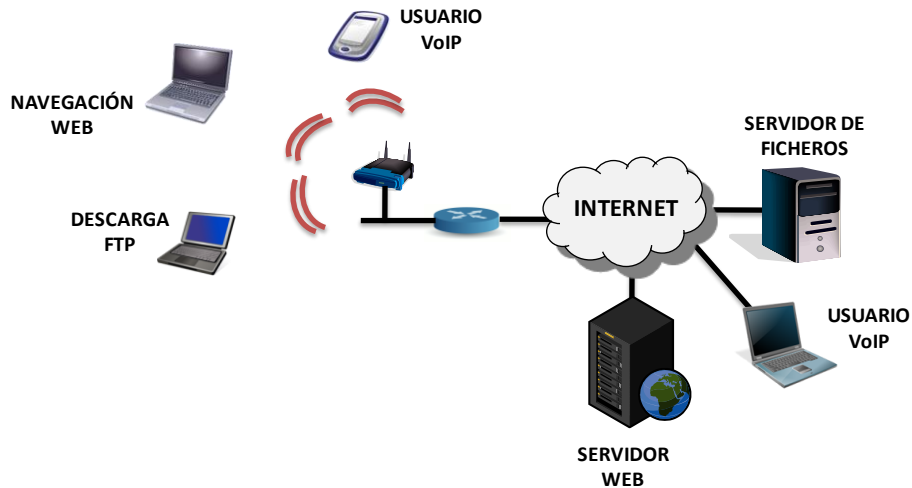


Figura 4.4. Escenario heterogéneo utilizado en la experimentación.

Se ha supuesto un canal inalámbrico libre de errores y, por tanto, la fragmentación a nivel MAC y el soporte de múltiples velocidades se han deshabilitado. Esta suposición permite enfocarse específicamente en el rendimiento del sistema en condiciones ideales. Los errores del canal se deberían considerar para tener un modelo de sistema más realista, pero el análisis de su influencia en los algoritmos de planificación se sale del ámbito de

<i>Parámetro</i>	<i>Valor</i>
<i>SIFS</i>	10 μ s
<i>PIFS</i>	30 μ s
<i>DIFS</i>	50 μ s
Ranura temporal (<i>slot time</i>)	20 μ s
Tasa binaria básica (<i>basic rate</i>)	1 Mbps
Cabecera de la capa física (h_{PHY})	192 bits
Cabecera de la capa MAC (h_{MAC})	288 bits
Longitud de la trama ACK (h_{ACK})	112 bits
Longitud de la trama CF-POLL (h_{POLL})	288 bits
Tasa binaria de datos	11 Mbps

Tabla 4.1. Parámetros de la capa física inalámbrica.

este estudio. En cualquier caso, la ausencia de errores no debería afectar a la generalidad de las conclusiones extraídas.

Además, se supone que todos los nodos se pueden comunicar directamente entre ellos y no existe el problema del nodo oculto, por lo que el mecanismo de protección RTS/CTS tampoco está habilitado.

Los mecanismos QACK y de *piggyback* del estándar 802.11e se encuentran activos dado que mejoran el rendimiento del sistema. Gracias al mecanismo de *piggyback*, el QAP puede enviar información de confirmación o de sondeo en una trama de datos. Al estar activado el bit QACK, el QAP puede combinar mensajes de control y datos dirigidos a diferentes estaciones.

Respecto a los enlaces en la parte cableada, éstos se han configurado para transmisión a alta velocidad (100 Mbps) y baja latencia (2 ms), de forma que pueda suponerse un impacto despreciable en los resultados debido a estos enlaces y así los resultados estén motivados por el rendimiento de la red inalámbrica.

En las simulaciones se empleó tráfico de voz (VoIP), de difusión de vídeo (*streaming* de vídeo), de vídeo sobre TCP, de navegación web y FTP, aplicaciones representativas de las clases de tráfico consideradas (véase la Sección 3.1). Las características de la generación de estos tipos de tráfico en NS-2 se describen a continuación.

Tráfico VoIP

Se ha elegido Voz sobre IP (VoIP) como aplicación representativa de la clase de tráfico conversacional. El tráfico VoIP ha sido generado tal como se especifica en la recomendación P.59 [65], en la que describen las características temporales –intervalos de actividad y silencio– de las conversaciones de voz humana. Tanto la duración como la tasa de las ráfagas de voz están distribuidas exponencialmente (con una media de 1.004 segundos y una tasa del 38.33 por ciento) y las duraciones de las pausas siguen una distribución *constante-más-exponencial* (con una media de 1.587 segundos). Para seguir estas especificaciones, el módulo *ns2voip* [66] fue actualizado para considerar conversaciones bidireccionales, i.e. incluyendo los estados de habla simple, habla doble y silencio mutuo. La transición entre estos estados se describe en la Figura 4.5, donde p_1 , p_2

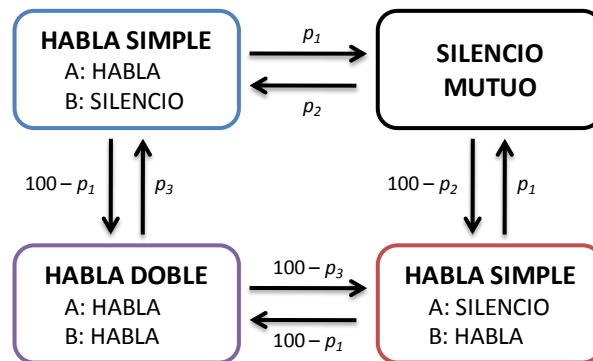


Figura 4.5. Modelo de transición de estados para una conversación.

y p_3 son las probabilidades de dichas transiciones y toman los valores 0.4, 0.5 y 0.5 respectivamente.

El códec utilizado ha sido G.711 [40], debido a su uso extensivo y su disponibilidad en prácticamente todos los sistemas VoIP. Además, dado que requiere un gran ancho de banda (64 kbps para la carga útil) se puede considerar como peor caso, comparado con otros *codecs* con menores requisitos de velocidad. Este *códec* genera una trama de voz (80 bytes) cada 10 milisegundos. En este caso, las fuentes VoIP se han configurado para concatenar tres tramas dentro del campo de datos de un único paquete UDP / RTP (240 bytes + 40 bytes de cabecera) enviado cada 30 milisegundos. Este valor, 30 ms, es el intervalo de servicio medio para las fuentes VoIP.

Tráfico streaming de vídeo

Para este tipo de tráfico se ha utilizado la herramienta *EvalVid* [44], que soporta la generación a partir de ficheros de traza para vídeos con formato MPEG-4, H.263 y H.264. Esta herramienta se puede utilizar en el simulador NS-2 para generar tráfico a partir de trazas reales [67].

Se han utilizado secuencias de vídeo con resolución QCIF (176x144) codificadas en formato MPEG-4 siguiendo la descripción en [23]. La traza utilizada ha sido *football*, que

se engloba en la categoría de *Movimiento Rápido* [23], la más exigente en cuanto a necesidad de recursos. El tráfico generado es de tasa de bit variable (VBR), con una tasa de cuadro de 15 fps y un tamaño medio de cuadro de 9445 bytes.

En la parte receptora se ha implementado un *buffer* de reproducción para mitigar los efectos del *jitter*. En las simulaciones, los paquetes se descartan si llegan después de 5 segundos (valor por defecto del tamaño del *buffer* en Microsoft® Windows Media™ Player [68]), medidos respecto a su momento de reproducción original.

Tráfico streaming de vídeo sobre TCP

Este tráfico se ha generado suponiendo las características típicas de los vídeos disponibles en el servidor YouTube [69], i.e. un tráfico con tasa binaria constante (CBR) generado a 330 kbps. El *Umbral de Tiempo Almacenado* (UTA) fue fijado a 5 segundos. Los vídeos considerados son lo suficientemente largos como para que la transmisión no se interrumpa durante la duración de las simulaciones realizadas.

Tráfico web

El generador de tráfico web modela los tamaños de las páginas con una variable aleatoria de distribución exponencial, cuya media es 130 KBytes [50]. El tiempo de lectura también sigue una distribución exponencial con una media de 61.73 segundos [70].

Tráfico FTP

Se ha supuesto ficheros con tamaño infinito en los servidores, de forma que las fuentes de tráfico FTP siempre tienen paquetes que transmitir.

Para las especificaciones de tráfico (TSPEC) negociadas se han considerado las tasas binarias asociadas a los servicios empleados cuando están definidas de forma implícita o explícita por el servicio. Así, para el tráfico de voz la velocidad especificada es

$$64 \text{ kbps} \times \frac{280 \text{ bytes incluyendo cabeceras}}{240 \text{ bytes de datos}} = 74.7 \text{ kbps}$$

74 | Una variante del Planificador–Conformador Monolítico para HCCA

En el caso de *streaming* de vídeo se ha utilizado su valor medio,

$$8 \frac{\text{bits}}{\text{byte}} \times 9445 \frac{\text{bytes}}{\text{trama}} \times 15 \frac{\text{tramas}}{\text{segundo}} = 1.13 \text{ Mbps}$$

ya que este tráfico es VBR. El tráfico de *streaming* de vídeo sobre TCP requiere una tasa binaria de 330 kbps debido al modelo de generación de tráfico usado.

Sin embargo, tanto el tráfico FTP como de navegación web no tienen asociadas unas velocidades de transmisión motivadas por su generación de tráfico, ya que el control de congestión del protocolo TCP hará que intente transmitir tan rápido como la red permita. Por tanto, hay que elegir un criterio para especificar las tasas de transmisión en sus TSPECs respectivos. Siguiendo la filosofía de este trabajo, se han usado las funciones de utilidad descritas en la Sección 3.2 para calcular las tasas de transmisión requeridas para conseguir un valor de MOS concreto en estos servicios. El valor elegido ha sido 3.75 (punto medio entre 3.5 y 4.0, rango elegido en el Capítulo 5 como criterio de calidad experimentada deseable), lo que implica unas velocidades de 660 y 450 kbps para FTP y navegación web, respectivamente.

Por otro lado, el tráfico de los servicios *streaming* de vídeo sobre TCP, web y FTP se transmiten sobre el protocolo TCP, de forma que los resultados incluyen los efectos debidos al mecanismo de control de congestión. Debido a las diferencias significativas entre la implementación de este protocolo en NS-2 y las implementaciones reales [71], se ha añadido el módulo *TCPLinux* [72] para reflejar el comportamiento real de TCP. El mecanismo de control de congestión utilizado en las simulaciones ha sido *New Reno* [73] debido a su amplia difusión [74] y a su uso en multitud de artículos como protocolo de transporte de referencia (e.g. [75] [76] [77] [78]).

En cuanto a la evaluación de rendimiento se utilizó el módulo *ns2measure* [79], módulo que permite la generación de estadísticas complejas en tiempo de simulación. Además, el uso de este módulo es necesario para la automatización de las simulaciones con el entorno *ANSWER*. De esta forma, se simplifica el proceso de obtención y análisis de resultados en NS-2 ya que no es necesario un postprocesado de los ficheros de trazas de las simulaciones y además se permite la elaboración de estadísticas que no podrían ser calculadas a partir de estos ficheros (e.g. el estado de una variable del sistema no se puede

obtener a partir de los eventos sucedidos a los paquetes en la red). La implementación de *ns2measure* fue modificada radicalmente para la incorporación de las funciones de utilidad descritas en la Sección 3.2.

4.3.2 Resultados

A continuación se explican los resultados obtenidos para evaluar el rendimiento de la variante de MSS para HCCA, así como su comparación con otros planificadores tomados como referencia. El Apéndice C incluye una comprobación de su implementación, en NS-2, en algunos escenarios relevantes.

Rendimiento de la variante de MSS para HCCA

Esta campaña de simulaciones tuvo como objetivo caracterizar el rendimiento de la variante del planificador MSS propuesta para su adopción en HCCA. Se han considerado dos cargas distintas de tráfico. El primer caso, denominado *carga media*, incluye una sesión de *streaming* de vídeo, una descarga progresiva de vídeo (*streaming* de vídeo sobre TCP), una sesión de navegación web y una descarga FTP. En el segundo caso, denominado *carga alta*, se utilizó una sesión de *streaming* de vídeo, dos descargas de vídeo sobre TCP, tres sesiones de navegación web y dos descargas FTP. En ambos casos el número de conversaciones (VoIP) varía entre 0 y 30.

La Figura 4.6 muestra la calidad de experiencia para los diferentes servicios en el experimento de *carga media*, incluyendo los intervalos de confianza (mediante barras de error). Para el mismo caso, la Figura 4.7 muestra la justicia entre servicios estimada mediante el índice de justicia definido en la Sección 3.3. Como se comentó en esa sección, en la evaluación se adopta la Nota Media de Opinión (es decir, un criterio de calidad de experiencia) para así tener una métrica común y unificadora entre los diferentes servicios y planificadores considerados.

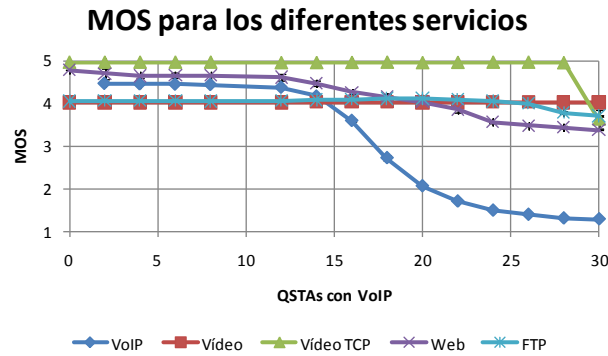


Figura 4.6. Calidad experimentada para los diferentes servicios usando el planificador MSS con carga de tráfico media.

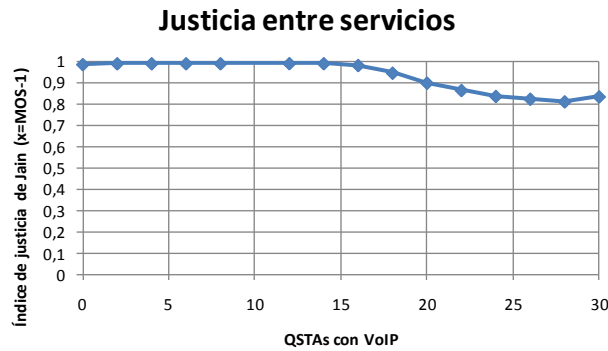


Figura 4.7. Justicia entre servicios del planificador MSS con carga de tráfico media.

Con esta carga, todos los usuarios tienen una calidad subjetiva superior a $MOS = 3.5$ hasta con 16 conversaciones. A partir de esa carga de tráfico VoIP, la calidad de este servicio decrece abruptamente mientras que el resto de servicios siguen manteniendo una calidad aceptable (MOS superior o similar a 3.5) hasta el último punto de carga simulado (30 conversaciones).

Esta reducción de calidad coincide con una bajada del índice de justicia a partir de 16 fuentes de voz, llegando a un valor mínimo de 0.8. Este valor refleja la diferencia entre

la calidad percibida por los usuarios de VoIP y el resto de usuarios, que sí reciben un servicio satisfactorio (MOS superior o similar a 3.5).

Siguiendo un análisis similar, las Figuras 4.8 y 4.9 muestran la calidad subjetiva y la justicia entre servicios, respectivamente, para el experimento denominado de *carga alta*.

En este caso, al aumentar el número de usuarios de los diferentes servicios, decrecen los recursos disponibles para cada usuario individual y por tanto se podrán admitir menos conversaciones con una calidad aceptable. Por este motivo, con *carga alta* el número de

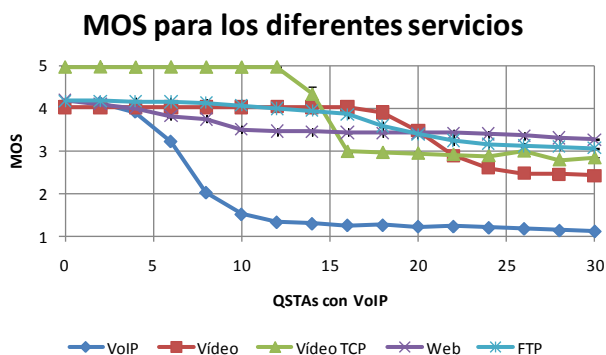


Figura 4.8. Calidad experimentada para los diferentes servicios usando el planificador MSS con carga de tráfico alta.

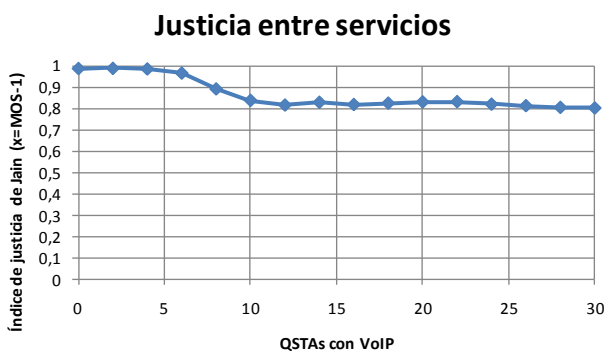


Figura 4.9. Justicia entre servicios del planificador MSS con carga de tráfico alta.

78 | Una variante del Planificador–Conformador Monolítico para HCCA

usuarios de VoIP simultáneos que experimentan un MOS superior a 3.5 se reduce a 5. En el caso de *streaming* de vídeo, la QoE se mantiene en el rango deseado hasta con 20 conversaciones. Sin embargo, a partir de 15 llamadas la calidad de *streaming* de vídeo sobre TCP disminuye por debajo de 3.5. En el caso de las sesiones web, su calidad se mantiene aproximadamente en 3.5 en todo el rango de tráfico VoIP simulado. Las descargas de tráfico consiguen un MOS entre 3.5 y 3 para 20 – 30 llamadas. En cualquier caso, si se utilizara un control de admisión basado en MOS sólo se permitiría el acceso de 5 usuarios del servicio de voz sobre IP, ya que la calidad decrecería a partir de esa carga por debajo de los límites establecidos. Respecto a la justicia del planificador, en este caso también se consigue un índice de justicia de 1 hasta 5 ~ 6 llamadas de voz, bajando hasta 0.8 con más llamadas. Al igual que en el experimento de *carga media*, VoIP ve su calidad muy mermada en comparación con el resto de servicios, que obtienen una QoE más homogénea.

Además, se incluye un tercer caso, *carga muy alta*, que será utilizado para su comparación con los esquemas desarrollados en los posteriores capítulos. Esta carga incluye dos sesiones de *streaming* de vídeo, tres descargas de vídeo sobre TCP, cinco sesiones de navegación web y tres descargas FTP. Como se aprecia en la Figura 4.10, el excesivo nivel de tráfico hace que la experiencia de los usuarios sea insatisfactoria cuando la red utiliza el esquema MSS (e.g. VoIP y *streaming* de vídeo presentan siempre una QoE *mediocre o mala*), por lo que su análisis en este capítulo no resulta relevante. No obstante, el bajo rendimiento en este escenario y las diferencias de QoE entre los distintos servicios motiva la necesidad de superponer un mecanismo que mejore la calidad de experiencia así como el índice de justicia de determinadas clases de tráfico según distintos criterios personalizables por el administrador de la red. Para ello en el Capítulo 5 se propondrá y evaluará una arquitectura que permitirá priorizar los servicios con requisitos de tiempo real, y en el Capítulo 6 se definirán y evaluarán otros criterios o políticas para la asignación de los recursos basándose en la arquitectura comentada.

A modo de conclusión, se pueden extraer las siguientes observaciones de los resultados obtenidos para las diferentes cargas (Figuras 4.6, 4.8 y 4.10). Así, debido a la dependencia de la QoE con las métricas objetivas de rendimiento (funciones de utilidad) para cada servicio:

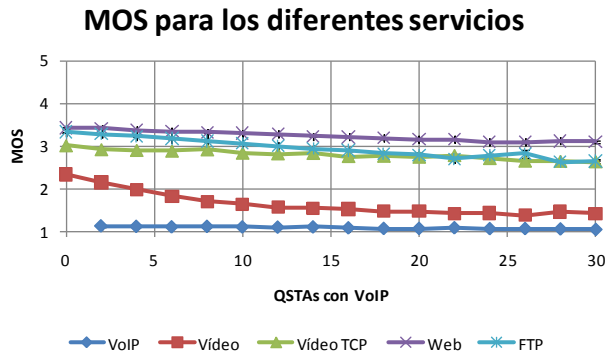


Figura 4.10. Calidad experimentada para los diferentes servicios usando el planificador MSS con carga de tráfico muy alta.

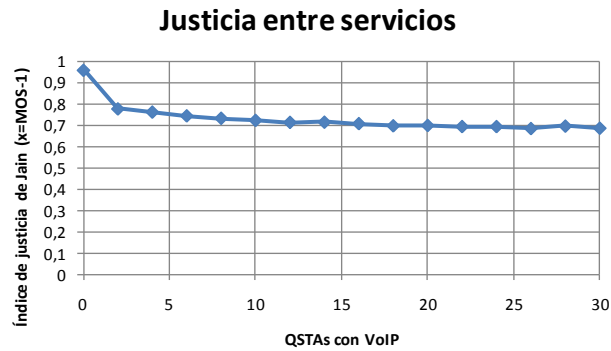


Figura 4.11. Justicia entre servicios del planificador MSS con carga de tráfico muy alta.

- El servicio VoIP es el más afectado ante la falta de recursos, llegando a tener mala calidad incluso cuando el resto de servicios obtiene un valor de MOS superior a 3.5 (e.g. véanse los puntos de mayor carga en la Figura 4.6). Esto se debe a que sus requisitos son más estrictos que los del resto de servicios, siendo más sensible ante condiciones de alta carga.
- Los usuarios del servicio de *streaming* de vídeo experimentan una degradación relativamente abrupta ante cambios en la carga –buena con

carga media, cambio abrupto entre *buena* y *mediocre* con *carga alta*, y *mediocre* o *mala* para *carga muy alta*–.

Así, si se cumplen los requisitos de velocidad de transmisión y retardo, el vídeo no experimentará degradación en la imagen y el usuario estará satisfecho. Si empiezan a retrasarse paquetes y no llegan antes de su momento de reproducción (incluyendo los efectos del *buffer* del receptor), se perderán tramas y la calidad de imagen dejará de ser aceptable. Esto motiva que la variación de la QoE con la carga sea tan brusca.

- *Streaming* de vídeo sobre TCP también presenta una transición abrupta en su QoE, siendo *excelente* para *carga media*, variando entre *excelente* y *regular* para *carga alta* y manteniéndose en *regular* para *carga muy alta*.

En este caso, este cambio se debe al estado el *buffer* de reproducción. Si tiene suficientes datos (por encima de un umbral) la calidad es *excelente*, mientras que si se queda sin datos la calidad es *mala*. Sólo en el caso de tener datos por debajo del umbral la calidad tomaría valores intermedios. Sin embargo, normalmente esta última situación suele ocurrir como transición de corta duración entre las dos situaciones anteriores provocando que, en media, las curvas de QoE no sean más graduales.

- La calidad experimentada en la navegación web varía de forma muy gradual con el aumento de carga, siendo ésta *buena* para *carga media*, entre *buena* y *regular* para *carga alta* y estando por encima de *regular* para *carga muy alta*.

Esto se debe a la suavidad de la curva de su función de utilidad (Figura 3.5) y que ésta sólo depende de la velocidad de transmisión.

- Algo parecido ocurre con el tráfico FTP, debido a que tienen unos requisitos similares (en términos de tasa media negociada en sus TSPECs respectivos para conseguir la QoE requerida). De esta forma, su calidad también es *buena* para *carga media*, entre *buena* y *regular* para *carga alta* y *regular* para *carga muy alta*.

De forma similar a la navegación web, el motivo radica en la variación gradual de su función de utilidad (Figura 3.6) y a la dependencia exclusiva

con la tasa de transmisión.

Comparación de MSS con otros planificadores usados con HCCA

A continuación se compara el rendimiento del planificador MSS con otros planificadores descritos en la literatura para ser usados con el mecanismo HCCA. Como se comentó previamente, esta comparación se realiza con los siguientes planificadores:

- *El planificador especificado en el estándar IEEE 802.11e*. A pesar de su sencillez, cumple los requisitos impuestos para HCCA y se utiliza ampliamente en la literatura relacionada como referencia. Esto permite, con las necesarias precauciones, comparar diferentes soluciones aun no teniendo sus implementaciones (no disponibles en la mayoría de los casos). La implementación de este esquema se encuentra disponible a través del módulo *ns2hcca* [64].
- *El planificador WTTP (Wireless Timed Token Protocol [13])*. Este planificador está diseñado para soportar tanto tráfico CBR como VBR, asignando un ancho de banda fijo a cada flujo y repartiendo el ancho de banda sobrante al final de la ronda de turnos. Su implementación ha sido cedida por el grupo *Computer Network Group* de la Universidad de Pisa.
- *El planificador FHCF (Fair scheduling scheme for HCF [8])*. Sus autores no demuestran si verifica las restricciones necesarias para HCCA, pero se trata de un planificador interesante para la comparativa ya que emplea todo el ancho de banda disponible. Además, su implementación se encuentra disponible [80].

Comparación con el planificador de referencia

Las Figuras 4.12 y 4.13 muestran el comportamiento del planificador de referencia, en términos de calidad experimentada y de justicia, para el experimento de *carga media*. Para esta campaña de simulaciones se ha deshabilitado el control de admisión ya que éste no permitía introducir tantas fuentes en el sistema, incluso sin tráfico VoIP.

En este escenario, los usuarios que descargan ficheros aprecian una calidad por

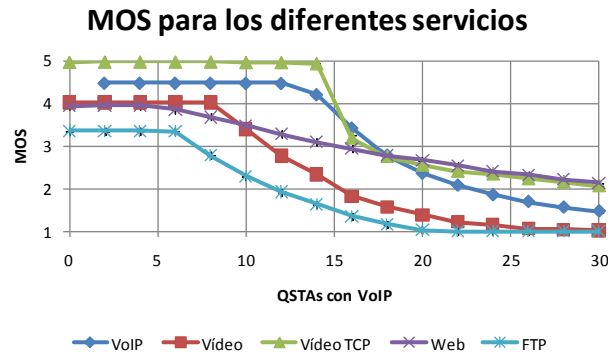


Figura 4.12. Calidad experimentada para los diferentes servicios usando el planificador de referencia con carga de tráfico media.

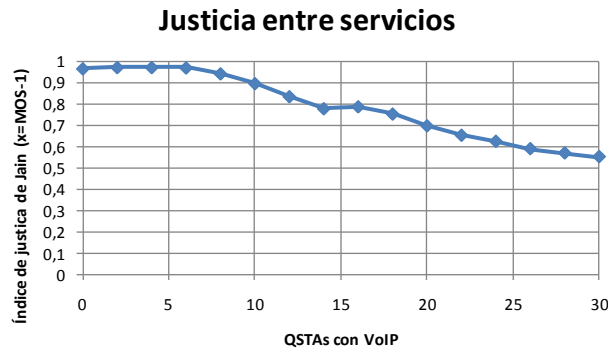


Figura 4.13. Justicia entre servicios del planificador de referencia con carga de tráfico media.

debajo del límite establecido incluso sin carga de VoIP, decreciendo notablemente a partir de 6 conversaciones. Sin embargo, los usuarios de *streaming* de vídeo y de navegación web consiguen una calidad aceptable hasta con 10 usuarios de VoIP, mientras que los usuarios de vídeo sobre TCP y VoIP no sufren una degradación notable hasta que la carga de voz es de 15 ~ 16 fuentes.

El índice de justicia también refleja estas diferencias en las curvas de MOS entre los diferentes servicios, decreciendo a partir de 6 usuarios y teniendo un mínimo aproximadamente de 0.55 en el punto de máxima carga.

4.3. Evaluación de rendimiento | 83

Por tanto, en el escenario simulado, tanto la calidad experimentada como la justicia entre servicios son inferiores a los valores obtenidos con la variante propuesta del planificador MSS (calidad aceptable para todos los usuarios hasta con 16 llamadas de voz e índice de calidad mínimo de 0.8).

Las conclusiones son similares para el caso de *carga alta*, como se aprecia en las Figuras 4.14 y 4.15.

Nuevamente, los usuarios de tráfico FTP tienen una calidad inferior a la requerida

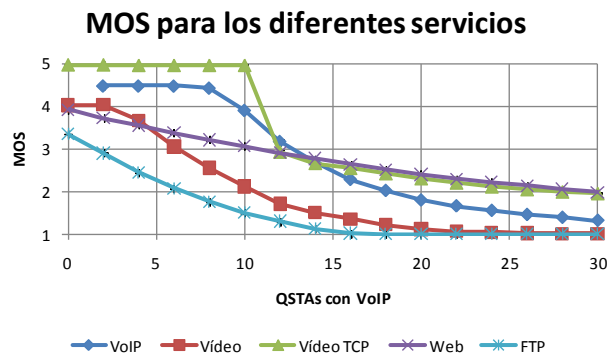


Figura 4.14. Calidad experimentada para los diferentes servicios usando el planificador de referencia con carga de tráfico alta.

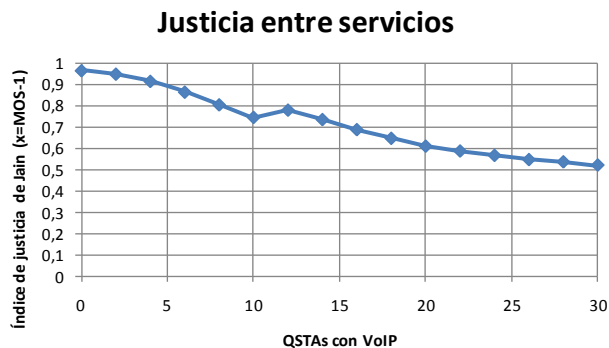


Figura 4.15. Justicia entre servicios del planificador de referencia con carga de tráfico alta.

84 | Una variante del Planificador–Conformador Monolítico para HCCA

con cualquier carga de tráfico de voz, decreciendo significativamente desde la primera llamada. El tráfico de *streaming* de vídeo y de navegación web tiene una calidad aceptable hasta con 5 conversaciones, mientras que el de VoIP y descarga progresiva de vídeo se mantiene en el rango deseable hasta aproximadamente con 11 llamadas. No obstante, debido al tráfico FTP, no hay ningún punto de carga en el que todos los usuarios estén satisfechos.

De forma análoga a lo que ocurría con *carga media*, el índice de justicia refleja estas diferencias entre los distintos servicios, reduciéndose desde la primera llamada y llegando a un mínimo inferior a 0.55.

El análisis expuesto indica que en este caso los resultados también son peores comparados con los conseguidos para el planificador MSS (todos los usuarios de los diferentes servicios satisfechos hasta con 5 llamadas VoIP y un índice de justicia mínimo de 0.8).

Comparación con el planificador WTTP

Los resultados para *carga media* utilizando el esquema WTTP se muestran en las Figuras 4.16 y 4.17.

Con esta carga, el planificador WTTP es capaz de satisfacer a todos los usuarios (MOS > 3.5) hasta con 10 llamadas de voz. Sin embargo, en la Figura 4.16 se aprecia el tratamiento diferente, en términos de QoE, para cada servicio. El servicio con mejor calidad es VoIP, ya que ésta se mantiene aceptable hasta con 20 llamadas de voz. En el caso de navegación web este rango se reduce hasta 15 llamadas aproximadamente. El resto de servicios ven su calidad degradada notablemente a partir de 10 ~ 11 conversaciones.

El comportamiento observado se debe al propio funcionamiento del planificador, que asigna un ancho de banda fijo a cada flujo (basado en su tasa binaria media) y reparte el resto de la capacidad sobrante para los flujos con tráfico VBR. Por ello, VoIP, que tiene un patrón CBR, siempre consigue su asignación fija por lo que se ve menos afectado que el resto de servicios.

4.3. Evaluación de rendimiento | 85

Tal como ocurría con el planificador de referencia, esta disparidad en la calidad experimentada por los diferentes servicios implica un alto grado de injusticia entre los mismos, reflejándose en el índice de justicia representado en la Figura 4.17. En este caso, el índice decrece a partir de 10 llamadas, coincidiendo con el rango de calidad aceptable para todos los usuarios, llegando hasta 0.55 cuando la carga aumenta.

Estos resultados indican, en este escenario, un peor rendimiento del planificador WTTP, en términos de calidad subjetiva y de justicia, comparado con MSS. WTTP

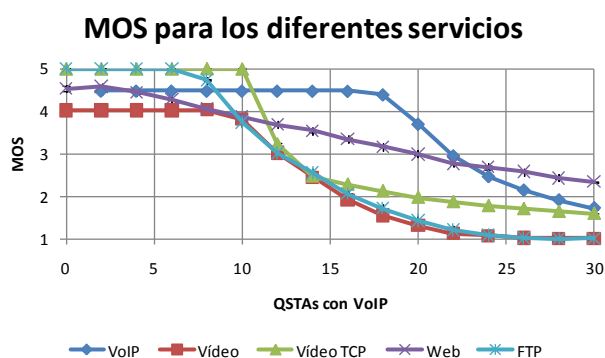


Figura 4.16. Calidad experimentada para los diferentes servicios usando el planificador WTTP con carga de tráfico media.

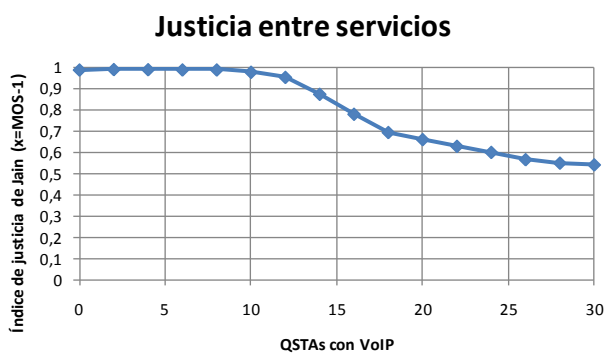


Figura 4.17. Justicia entre servicios del planificador WTTP con carga de tráfico media.

86 | Una variante del Planificador–Conformador Monolítico para HCCA

permite la admisión de 10 llamadas de voz manteniendo la satisfacción de todos los usuarios, mientras que MSS permitía 16 llamadas. El menor valor del índice de justicia, 0.55 frente a 0.8, también indica que MSS es más justo en términos de calidad percibida cuando los recursos no son suficientes para mantener la calidad requerida.

Las Figuras 4.18 y 4.19 muestran las mismas curvas para el experimento de *carga alta*.

Tal como ocurría con los planificadores MSS y de referencia, las conclusiones son

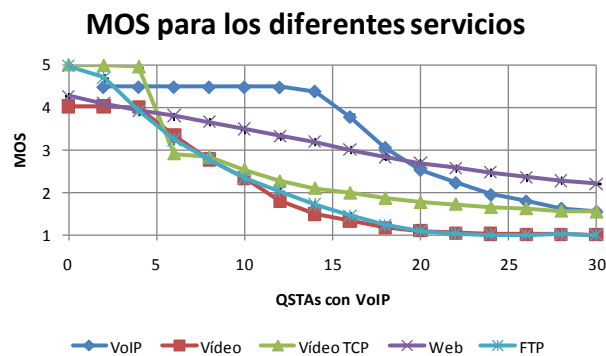


Figura 4.18. Calidad experimentada para los diferentes servicios usando el planificador WTTP con carga de tráfico alta.

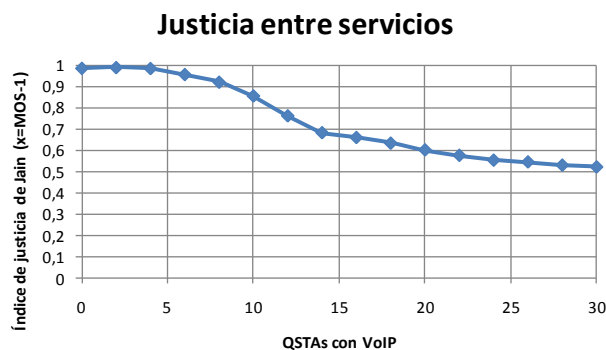


Figura 4.19. Justicia entre servicios del planificador WTTP con carga de tráfico alta.

similares a las deducidas para *carga media*, salvo el lógico desplazamiento en cuanto al número de estaciones con tráfico VoIP soportadas debido a la mayor carga de tráfico en el sistema.

Para esta carga, WTTP sólo cumple la calidad requerida hasta con 5 conversaciones, como se observa para las curvas de *streaming* de vídeo, de descarga de vídeo sobre TCP y de ficheros. La calidad para la navegación web se mantiene aceptable hasta con 10 usuarios VoIP, y para el tráfico de voz hasta con 15 ~ 16 llamadas.

Nuevamente, proporcionar una respuesta diferente para cada servicio hace que se reduzca la justicia de esta disciplina, tal y como se observa en la Figura 4.19. El valor mínimo del índice de justicia es de 0.5.

El número de usuarios que puede admitir el sistema, con esta configuración, es similar al soportado por MSS, i.e. hasta 5 llamadas de voz. Sin embargo, la justicia entre servicios es muy inferior usando WTTP una vez que faltan recursos, ya que su índice es de 0.5 (en el caso peor) comparado con el valor 0.8 conseguido con MSS.

Comparación con el planificador FHCF

Para esta comparativa se han eliminado los conformadores de tráfico *dual leaky buckets* utilizados para limitar el tráfico TCP cuando el planificador utilizado es MSS, para que se pueda realizar una comparación justa en términos de la tasa binaria obtenida por usuario. Si no se deshabilitasen estos reguladores, se estaría limitando la generación de tráfico al usar MSS (que no podría superar los 660 kbps negociados en la TSPEC) y no al utilizar FHCF.

Debido a que la implementación disponible para FHCF [80] está basada en una versión antigua de *Network Simulator 2* (v2.1b7), la mezcla de tráfico se ha reducido a VoIP (utilizando fuentes CBR) y FTP. No se ha deshabilitado el control de admisión utilizado por FHCF, limitándose la comparación al rango de carga permitido. Por este motivo sólo se ha incluido una fuente de tráfico FTP.

88 | Una variante del Planificador–Conformador Monolítico para HCCA

La Figura 4.21 muestra la calidad experimentada por los usuarios del servicio VoIP. La curva para FHCF está limitada a 11 llamadas debido al control de admisión, consiguiéndose un valor de MOS superior a 4.4 en todos los puntos de carga. Estos valores son similares a los obtenidos con MSS, si bien en este caso, un control de admisión basado en MOS (e.g. $MOS > 3.5$) habría permitido hasta 25 conversaciones.

La calidad subjetiva para el servicio FTP está representada en la Figura 4.20. De nuevo, los valores obtenidos para el planificador FHCF son *excelentes* para cualquier número de llamadas de voz permitidas por el control de admisión (11 llamadas). En ese rango la opinión de los usuarios de FTP cuando el planificador es MSS también es *excelente*. Un control de admisión basado en QoE habría permitido hasta 26 ~ 27 conversaciones con una calidad aceptable para FTP.

Sin embargo, en este caso resulta interesante la comparación en términos de velocidad de transmisión, ya que la calidad subjetiva es prácticamente perfecta en ambos casos (suponiendo el control de admisión de FHCF). Como se observa en la Figura 4.22, en el escenario simulado, la tasa de transmisión para la descarga de ficheros es ligeramente superior en el caso del planificador MSS.

De esta forma se comprueba que, en este escenario concreto, el planificador MSS

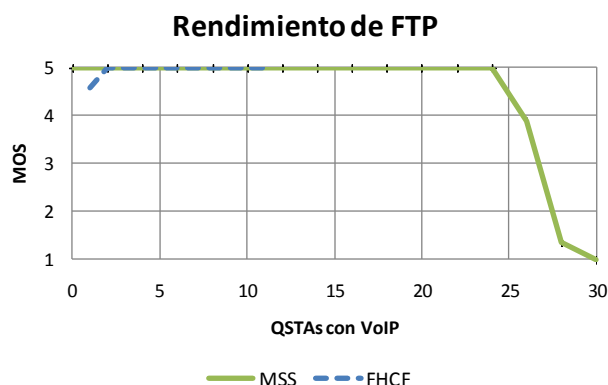


Figura 4.20. Comparación de la calidad experimentada en FTP para los planificadores MSS y FHCF.

consigue mejor rendimiento que el planificador FHCF y además cumple los requisitos impuestos para los planificadores HCCA.

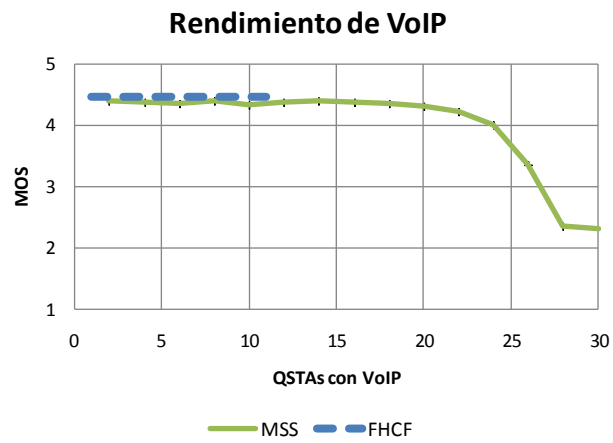


Figura 4.21. Comparación de la calidad experimentada en VoIP para los planificadores MSS y FHCF.

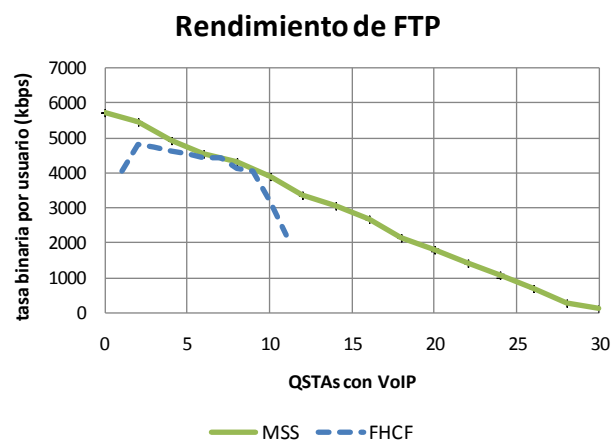


Figura 4.22. Comparación del rendimiento (tasa binaria) en FTP para los planificadores MSS y FHCF.

4.4 Conclusiones

En este capítulo se ha propuesto la adopción de un planificador de paquetes desarrollado como una variante del Planificador–Conformador Monolítico (MSS), para su uso en el mecanismo de acceso HCCA. Este algoritmo de planificación fue elegido por la idoneidad de sus características, ya que es capaz de garantizar tanto la velocidad de transmisión negociada, como un retardo acotado en la transmisión de los paquetes.

Se ha descrito el funcionamiento de esta disciplina así como un análisis teórico de sus principales características (tasa binaria garantizada, retardo máximo incurrido y tráfico servido). Posteriormente se ha explicado cómo adaptar este planificador para poder ser utilizado en un entorno inalámbrico con soporte de HCCA. En el Apéndice B se ha comprobado, en los escenarios simulados, que la variante propuesta de MSS cumple los requisitos impuestos para los planificadores según el estándar IEEE 802.11e (Ecuación 2.5).

Finalmente, se ha analizado el rendimiento de este esquema en términos de calidad experimentada por los usuarios y de justicia entre los diferentes servicios, incluyendo una comparativa con los planificadores de referencia, WTTP y FHCF. Los resultados obtenidos permiten concluir que, en los escenarios estudiados, MSS es más eficiente (en términos de QoE y de justicia) que los planificadores de referencia considerados.

Capítulo 5

Gestión de recursos para servicios heterogéneos: Método de los Territorios

Los planificadores disponibles en la literatura para el mecanismo de acceso HCCA, incluyendo la variante del planificador MSS, no son adecuados para soportar simultáneamente tráfico elástico y tráfico de tiempo real.

Un ejemplo típico de tráfico elástico es la carga generada por las aplicaciones transportadas sobre TCP, ya que su velocidad varía debido al algoritmo de control de congestión. El estándar IEEE 802.11e requiere que todos los flujos especifiquen su tasa de datos media (*mean data rate*), algo que, en general, no es posible para este tipo de servicios. Cualquier valor elegido conduciría a un comportamiento no óptimo del planificador. Por ejemplo, si se elige un valor bajo para la tasa media, los flujos TCP tendrían una velocidad reducida incluso en condiciones de carga baja (en el caso del planificador MSS, debido a la política SEFF de selección de paquetes, los flujos TCP

92 | Gestión de recursos para servicios heterogéneos: Método de los Territorios

además se verían muy afectados por otros tipos de flujo). Por el contrario, si se estipulara una tasa alta, el resto de flujos se verían degradados (incluso aquellos con restricciones estrictas de tiempo real). En consecuencia, el resultado no es óptimo en ninguno de los casos, por lo que no son adecuados para su adopción en HCCA cuando simultáneamente se tiene tráfico heterogéneo.

Por otro lado, no todos los escenarios requieren el cumplimiento estricto de requisitos de QoS. Aunque este tipo de garantías es imprescindible en situaciones críticas, como las de procesos industriales o en instrumentación aeronáutica, en la mayoría de escenarios se podría tolerar una pequeña degradación si la experiencia subjetiva del usuario final es satisfactoria. Esta pequeña degradación podría permitir liberar recursos que se podrían utilizar para flujos con peor calidad experimentada.

Teniendo en cuenta estas observaciones, se ha diseñado un mecanismo dinámico de asignación de recursos con dos objetivos claramente diferenciados: incrementar la capacidad del sistema manteniendo –aproximadamente– la satisfacción del usuario final, y adaptar los planificadores HCCA –este estudio se centrará en el planificador MSS– para soportar de forma eficiente y simultánea flujos heterogéneos, es decir, tanto de tráfico elástico como de tiempo real. El algoritmo propuesto –denominado como *Método de los Territorios*– divide la capacidad del medio inalámbrico (en términos de oportunidades de transmisión) en *territorios* para diferentes tipos de tráfico, usando como criterio para asignar recursos una estimación de la Calidad de Experiencia percibida por los usuarios finales.

Hasta donde llega el conocimiento del autor, no existe ningún mecanismo –basado en HCCA– en la literatura actual sobre redes IEEE 802.11e que considere la compartición de los recursos entre diferentes clases de tráfico o servicios. Sin embargo, otras tecnologías inalámbricas como UMTS [35] o LTE [36] utilizan este concepto (*QoS classes*) para controlar el tratamiento en el reenvío de paquetes en los niveles inferiores (e.g. pesos en las disciplinas de planificación de paquetes, umbrales de admisión, gestión de colas, etc.). Algunos trabajos [81] [82] [83] [84] defienden el uso de los valores de MOS para asignar los recursos dentro de una red. De hecho, proponen resolver un problema de optimización *cross-layer* cuyo objetivo es optimizar la suma del MOS para un conjunto de usuarios. Suponen que éste se puede calcular como una función no lineal de la tasa de transmisión. Sin embargo, estas propuestas requieren que las aplicaciones adapten su tasa de forma

continua, lo que no permite su utilización en muchas aplicaciones existentes que no son adaptables..

Por otra parte, ningún algoritmo de gestión de recursos radio (planificador, control de admisión, etc.) para HCCA ha basado sus decisiones *online* en métricas de calidad subjetiva como MOS. Sin embargo, esta aproximación no es nueva en otras tecnologías radio como OFDMA, con algunas propuestas [25] que persiguen la maximización del agregado de la calidad subjetiva experimentada por los usuarios a partir de funciones de utilidad.

El resto del capítulo se organiza de la siguiente manera. El funcionamiento del Método de los Territorios se describe en la Sección 5.1. La Sección 5.2 presenta la evaluación de su rendimiento, y las conclusiones extraídas se exponen en la Sección 5.3.

5.1 Funcionamiento del Método de los Territorios

Este algoritmo de gestión dinámica de recursos, el *Método de los Territorios*, tiene como objetivo mejorar el uso del ancho de banda disponible manteniendo simultáneamente la satisfacción de los usuarios finales en un escenario con tráfico heterogéneo, i.e. proveniente de múltiples aplicaciones con diferentes requisitos. Además, permite priorizar los servicios de tiempo real y mejorar el rendimiento de las aplicaciones TCP.

Los principios y objetivos fundamentales que han guiado el diseño de este mecanismo son los siguientes:

- La asignación de recursos se basará en la Calidad de Experiencia de los distintos flujos de tráfico.
- Los servicios de tiempo real no deben sufrir degradación debido al tráfico sin requisitos temporales estrictos.
- Las aplicaciones basadas en TCP podrán variar su velocidad para adaptarse a la

94 | Gestión de recursos para servicios heterogéneos: Método de los Territorios

capacidad del canal –gracias al funcionamiento de su mecanismo de control de congestión–. Esto es, la solución diseñada no debe restringir la velocidad de transmisión de estos flujos más allá de la limitación impuesta por la falta de recursos.

- Se debe evitar la inanición de cualquiera de los servicios soportados.
- La propuesta debe mejorar la utilización del canal inalámbrico.

Para cumplir estos objetivos, este método reparte el conjunto común de oportunidades de transmisión (TXOPs) en diferentes territorios (véase la Figura 5.1), uno por cada clase de tráfico definida (véase la Sección 3.1). De esta forma, todos los servicios con el mismo tipo de requisitos de Calidad de Servicio compartirán el mismo territorio, y los servicios con distintos tipos de requisitos harán uso de territorios también diferentes.

Cada territorio se define por una *capacidad mínima* (CM) –configurable para evitar la inanición de sus servicios– y una *capacidad reservada* (CR) –cuyo tamaño puede variar solicitando o devolviendo TXOPs del conjunto de recursos libres–.

La CR fluctuará de acuerdo a la calidad experimentada (MOS) por el usuario final, tal como se muestra en la Figura 5.2. Como se puede observar en esta figura, este algoritmo se ha diseñado para que el MOS promedio de una cierta clase de tráfico (media de los valores de MOS de los distintos flujos pertenecientes a dicha CT) permanezca entre un *umbral superior de MOS* (USM) y un *umbral inferior de MOS* (UIM). Si el MOS estimado es inferior a UIM, la percepción del usuario final es peor que la requerida y su CR se incrementará para mantener el valor de MOS dentro del rango requerido –respetando siempre las CM de los otros tipos de tráfico–. Por el contrario, si está por

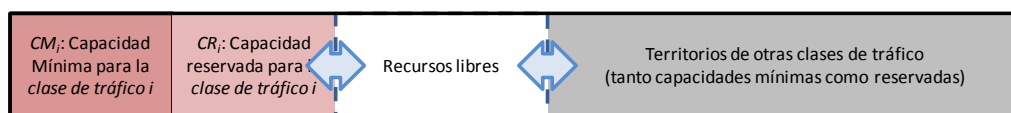


Figura 5.1. Ejemplo de definición de territorios para una clase de tráfico genérica.

5.1. Funcionamiento del Método de los Territorios | 95

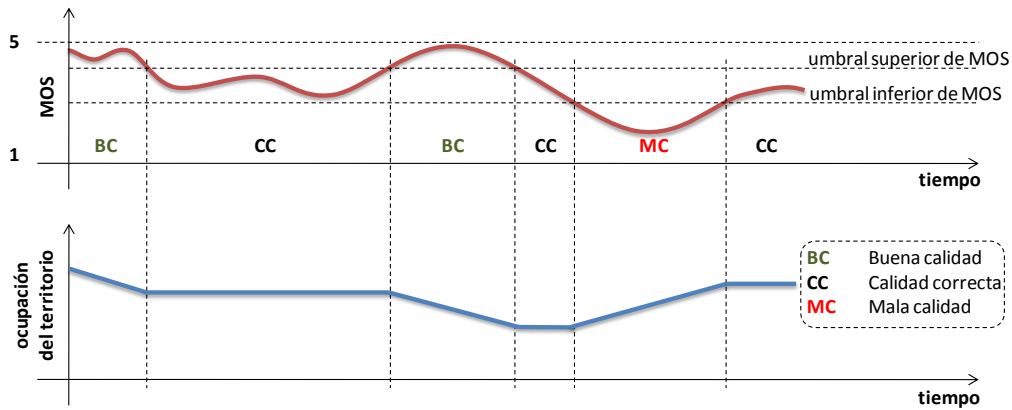


Figura 5.2. Ejemplo de la relación entre la calidad subjetiva y la evolución del territorio para una clase de tráfico genérica.

encima de USM, significa que la experiencia del usuario es mejor que la requerida y, por tanto, se puede devolver parte de los recursos de la CR.

De esta manera, el Método de los Territorios requiere del cálculo *on-line* de los valores de QoE de los distintos flujos de tráfico, que serán promediados para estimar el valor de MOS medio para una clase de tráfico dada. Estos cálculos se realizan utilizando los valores observables en la red y las funciones de utilidad descritas en la Sección 3.2.

Comportamiento dinámico del Método de los Territorios

A continuación se describen los parámetros que afectan al comportamiento dinámico de la estrategia propuesta. Estos parámetros están relacionados con:

- La estimación de la calidad percibida
- La asignación de recursos a los territorios

Respecto a la estimación de la calidad percibida por los usuarios, los valores o muestras de MOS de cada flujo se calcularán considerando un período fijo y configurable, denominado *Período de Filtrado de MOS* (PFM). Estas muestras serán calculadas según las funciones de utilidad definidas en la Sección 3.2.

96 | Gestión de recursos para servicios heterogéneos: Método de los Territorios

Además, el procedimiento incluye un filtrado exponencial para evitar oscilaciones en los territorios debido a variaciones abruptas e instantáneas del valor de MOS de una clase de tráfico dada (media del MOS estimado para los flujos pertenecientes a dicha clase). El proceso de promediado se muestra en la Ecuación 5.1.

$$AVGMOS_i[k] = \alpha \cdot AVGMOS_i[k-1] + (1 - \alpha) \cdot MOS_i[k] \quad (5.1)$$

donde $AVGMOS_i[k-1]$ es el valor promediado del MOS de la clase de tráfico i para el $(k-1)$ -ésimo intervalo PFM, α es un factor adimensional o peso –definido entre 0 y 1– para ponderar los valores promedio pasados con el valor actual $MOS_i[k]$ –medida de MOS del período k -ésimo–.

Respecto a la asignación de recursos a los diferentes territorios, la solución propuesta necesita adaptarse de forma dinámica al tráfico cambiante existente en la red. Para ello, tanto las medidas de MOS como las actualizaciones de los territorios deben considerar las diferencias entre el comportamiento dinámico de las distintas clases de tráfico (e.g. el tráfico web es rafagueante e intermitente, requiriendo así variaciones rápidas de su territorio, mientras que el tráfico FTP es por lo general continuo, sugiriendo así variaciones más suaves).

Por ello, cada CT puede seleccionar su propio período de promediado de la QoE (PFM_{*i*}) dependiendo de las características de su tráfico. De esta manera, los servicios con una generación de tráfico estable (e.g. FTP) pueden utilizar un período superior que los servicios rafagueantes (e.g. navegación web). En este último caso, el patrón de tráfico varía más rápidamente –y por tanto su QoE–, y el uso de períodos demasiado largos podría hacer que los territorios no reaccionaran de forma adecuada a estos cambios.

Siguiendo la misma filosofía, también se puede configurar el valor de los períodos para ajustar los territorios. Así, se consideran diferentes *Períodos de Actualización de Territorio* (PAT_{*i*}) –uno por cada CT– con el objetivo de tener en cuenta sus diferentes propiedades. Independientemente de la clase de tráfico considerada, parece razonable suponer que se deberían ponderar varias muestras de MOS para el cambio de tamaño de los territorios: pocas muestras podrían conducir a fluctuaciones en los territorios debidas a picos de MOS, conduciendo así a un rendimiento inestable de los servicios; demasiadas medidas podrían inducir un retardo excesivo en la respuesta a los cambios, de nuevo

causando un rendimiento inadecuado.

El valor del parámetro α afecta de manera similar, ya que un valor bajo ($\alpha \approx 0$) hace que sólo se considere una muestra de MOS –la última– pudiendo causar oscilaciones en las decisiones tomadas, mientras que un valor alto ($\alpha \approx 1$) ralentizaría la actualización de los territorios –al suavizar excesivamente las estimaciones realizadas– frente a cambios en la calidad percibida.

Otra decisión importante respecto a la dinámica de este esquema será el criterio utilizado para modificar el tamaño de los territorios cuando se detectan violaciones de los umbrales USM o UIM. Nuevamente se debe llegar a un compromiso entre estabilidad y velocidad de respuesta. Eligiendo un paso pequeño para incrementar un territorio, se produciría una falta de respuesta ante un crecimiento rápido de tráfico. Sin embargo, un valor bajo para reducir un territorio –e.g. ante un descenso de la carga– podría ser adecuado para evitar que la QoE se degradase excesivamente. En consecuencia, parece lógico pensar que se podrían obtener mejores resultados manteniendo un *Paso de Decremento* (PD) y un *Paso de Aumento* (PA) diferentes y adaptables. Esta idea de utilizar incrementos y decrementos diferentes no es nueva sino que aparece en otros entornos, como es el caso de la política AIMD (*Additive Increase Multiplicative Decrease*) para el control de congestión en TCP.

Siguiendo esta hipótesis, se propone utilizar un valor de PA proporcional a la distancia de la calidad subjetiva respecto al rango de *buena calidad* (MOS entre UIM y USM), i.e. el paso será mayor si la calidad es muy inferior de la requerida y menor si está próxima. La Ecuación 5.2 muestra cómo se calcula el valor de PA tras un período de actualización de territorio.

$$PA = \left[\text{paso_máximo} \times \frac{\text{diferencia_MOS_actual}}{\text{diferencia_MOS_máxima}} \right] \quad (5.2)$$

donde

$$\text{diferencia_MOS_máxima} = \left(\frac{USM + UIM}{2} - 1 \right)$$

$$\text{diferencia_MOS_actual} = \left(\frac{USM + UIM}{2} - AVGMOS \right)$$

98 | Gestión de recursos para servicios heterogéneos: Método de los Territorios

y *paso_máximo* es el valor máximo (configurable) que puede tomar PA. De esta forma, el territorio se actualiza rápidamente cuando se reduce la QoE (e.g. por aumentos en la carga).

Respecto al parámetro PD se ha optado por asignarle un valor fijo y pequeño –si bien es configurable–, ya que una reducción excesiva del territorio provocaría una inmediata disminución de la QoE percibida por los usuarios y probablemente un incremento posterior de dicho territorio. De esta manera se evitarían estas oscilaciones (tanto en QoE como en recursos asignados).

Política de asignación de recursos del Método de los Territorios

Este método utiliza un esquema con *derecho preferente*, es decir, las clases de tráfico se ordenan por prioridad, y una clase puede *apropiarse* de recursos del territorio de otra clase con menor prioridad (siempre que no sea de su capacidad mínima). De esta forma, los servicios de tiempo real (*conversacional* y *streaming*) son favorecidos a costa de los servicios sin este tipo de restricciones.

En el punto de acceso, el planificador de paquetes (MSS en el análisis realizado) arbitra las CTs de acuerdo a estos territorios. Primero, el Método de los Territorios comprueba mediante una lista *round-robin* qué CT tiene reservada la TXOP actual. Después, el planificador selecciona un flujo perteneciente a dicha CT para la transmisión del paquete. Por tanto, puede decirse que el Método de los Territorios actúa como un recubrimiento previo sobre la acción del planificador utilizado.

Para aumentar el uso del ancho de banda disponible, en el caso de no haber ningún paquete a transmitir de la CT correspondiente a la TXOP actual se seleccionará la clase con peor calidad subjetiva (mayor distancia porcentual respecto al punto medio del rango requerido, i.e. $\frac{USM+UIM}{2}$). Si la clase seleccionada no tiene flujos con paquetes a transmitir, se seleccionará la siguiente clase con peor QoE y así sucesivamente hasta encontrar un paquete a transmitir. De esta manera, ninguna TXOP se desaprovecha y, en consecuencia, se aumenta la utilización del canal.

La Figura 5.3 detalla el diagrama de flujo del Método de los Territorios usando la política descrita.

5.1. Funcionamiento del Método de los Territorios | 99

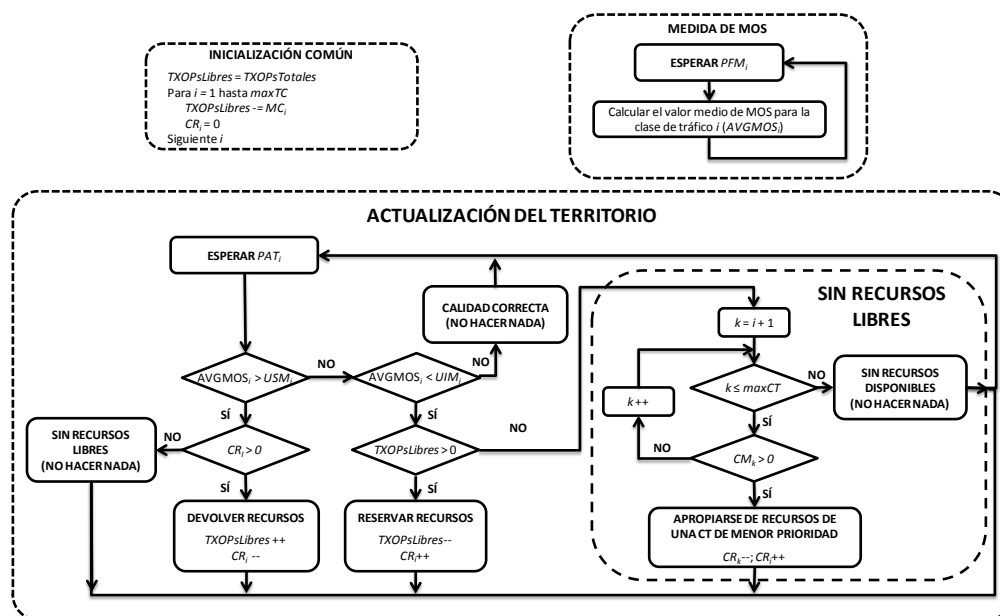


Figura 5.3. Diagrama de flujo del Método de los Territorios.

Tal como se ha comentado, el Método de los Territorios es independiente del algoritmo de planificación seleccionado, pudiendo así ser aplicado a cualquier planificador HCCA. Además, siempre que el planificador implemente una disciplina justa, la disciplina de planificación elegida debería tener un impacto reducido.

En el caso concreto del planificador utilizado (MSS) resulta necesario deshabilitar los conformadores de tráfico (*dual-leaky-buckets*) para las clases de tráfico elástico (*TCP de tiempo real, interactiva y de fondo*), dado que sus velocidades de transmisión no deben estar acotadas *a priori*. Esto no representa una degradación de las CTs de tiempo real (*conversacional y streaming*) debido a la definición de los territorios por clases y al esquema con prioridad adoptado.

5.2 Evaluación de rendimiento

5.2.1 Condiciones experimentales

En general, las condiciones experimentales han sido las expuestas anteriormente en la Sección 4.3.1, utilizando Network Simulator 2 como entorno de simulación y la herramienta ANSWER para su automatización. Al igual que en el capítulo anterior, el nivel de confianza para todos los experimentos ha sido del 95 por ciento (con un máximo de diez réplicas por punto simulado) para garantizar la validez estadística de los resultados.

El Método de los Territorios se ha implementado en NS-2 como recubrimiento a la función desempeñada por el planificador de paquetes. En concreto, este estudio se ha centrado en la variante de MSS para HCCA explicada en el Capítulo 4, si bien se podría utilizar cualquiera de los planificadores disponibles para este mecanismo de acceso.

La capacidad mínima de cada territorio se ha configurado al 10%, de forma que la mitad de los recursos queda disponible para ser reservada en función de las necesidades de las cinco clases de tráfico. Estos valores han sido escogidos buscando un compromiso entre asignación estática de los recursos –para garantizar un rendimiento mínimo de los diferentes servicios– y asignación dinámica –para permitir la adaptación de esta arquitectura a las condiciones cambiantes, basándose en la QoE de los usuarios–.

Para calcular el valor promediado de MOS de cada CT, se ha utilizado un Período de Filtrado de MOS –dependiente del servicio– igual a 1 segundo salvo para el servicio de navegación web, debido a la intermitencia de su tráfico [70]. Considerando este valor, el número medio de paquetes durante un período sería:

- $1\text{s} / 0.030\text{ s/paquete} \approx 33$ paquetes en el caso de VoIP
- $1\text{s} \times 10\text{ paquetes/cuadro} \times 15\text{ cuadros/s} = 150$ paquetes para *streaming* de vídeo
- $330\text{ Kbps} / 1500\text{ bytes/paquete} \approx 28$ paquetes para vídeo sobre TCP

- 660 Kbps / 1460 bytes/paquete \approx 56 paquetes para FTP (suponiendo que su calidad está en el valor medio del rango deseado, i.e. MOS = 3.75)

De esta forma, cada cálculo de MOS utilizará un número suficiente de paquetes tal que la estimación sea fiable.

Este razonamiento sería también aplicable al tráfico HTTP. Sin embargo, debido a los períodos de descarga y lectura, un período de un segundo sería excesivo. Considerando el modelo empleado, una página web tiene un tamaño medio de 130 KBytes [50]. Suponiendo una velocidad de transmisión de 450 Kbps (valor que permite un MOS = 3.75), tardaría en descargarse $130 \text{ KBytes} / 450 \text{ Kbps} = 2.31\text{s}$. Suponiendo que la QoE fuese inferior –siendo deseable un aumento de su territorio si no superase UIM–, este tiempo sería superior pero del mismo orden (e.g. tardaría 4.33 s para un MOS igual a 3, a una velocidad de 240 Kbps). De forma equivalente, si su MOS fuese superior al considerado –devolviendo recursos si superase USM– el tiempo de descarga sería inferior. Teniendo en cuenta que se deberían considerar varias muestras de MOS para actualizar su territorio –por motivos de estabilidad–, resulta evidente que un segundo no es un período adecuado para estimar la QoE de este servicio. Por este motivo –permitir la actualización del territorio mientras dure la descarga de la página–, se ha utilizado un valor de 0.2 como PFM para el servicio web (con aproximadamente 7.5 paquetes enviados durante ese período).

Para el Período de Actualización de Territorio –dependiente de la clase de tráfico–, se propone considerar al menos 5 muestras de MOS –para tener una estimación adecuada de las condiciones de la red que permita tomar decisiones (y evitar fluctuaciones)–. Así, el PAT valdrá 5 segundos para todas las clases de tráfico salvo la interactiva, cuyo territorio se actualizará cada segundo. Además, se ha asignado 0.5 como valor para el peso de ponderación de muestras pasadas α para todos los servicios.

El Apéndice D incluye un estudio preliminar que justifica los valores utilizados para los parámetros PFM_i , PAT_i y α_i .

Respecto a la asignación de recursos a los territorios, se han utilizado rondas de 40 TXOPs, lo que implica que el paso mínimo de actualización del territorio es $1/40 = 2.5\%$ de la capacidad del canal (en términos de oportunidades de transmisión).

102 | Gestión de recursos para servicios heterogéneos: Método de los Territorios

El Paso de Decremento (PD) ha sido establecido en una TXOP (2.5% de la capacidad) que es el valor mínimo posible, de forma que se eviten oscilaciones que podrían afectar de forma negativa a la calidad experimentada por los usuarios del territorio que reducen su tamaño.

Por el contrario, el Paso de Aumento (PA) se ha configurado para que un territorio pueda crecer rápidamente en el caso de necesidad (e.g. por un aumento repentino de la carga en esa CT). Por ello, se ha asignado un valor de 8 TXOPs (20% del total) al *paso_máximo*. Como se comentó, PA es adaptativo y su valor depende de la diferencia entre el valor medio de MOS actual y el valor deseado. Así, por ejemplo, si la QoE de los servicios conversacionales fuese *mala* (MOS = 1) por un incremento sustancial del número de usuarios de estos servicios, se utilizaría el *paso_máximo* (20%) para aumentar su territorio, tardando así tres Períodos de Actualización de Territorio en atribuirse –en caso necesario– toda la capacidad reservable ($\lceil 0.5/0.2 \rceil = 3$), i.e. 15 segundos considerando el intervalo PAT elegido. Un valor inferior (e.g. 5%) haría que el territorio conversacional tardase excesivamente en conseguir dicha capacidad ($\lceil 0.5/0.05\% \rceil = 10$ PATsen este ejemplo, i.e.50 segundos).

5.2.2 Resultados

Para la evaluación del rendimiento del Método de los Territorios se han considerado tanto condiciones estáticas –la carga no varía con el tiempo– como dinámicas –el número de fuentes varía durante la simulación–.

Las primeras campañas, similares a las realizadas en la Sección 4.3, han incluido simulaciones en las que la carga es estática. Al igual que en el capítulo anterior, se han considerado experimentos con *carga media*, *carga alta* y *carga muy alta* para que los resultados obtenidos sean directamente comparables.

El segundo conjunto de simulaciones tiene como objetivo analizar las características dinámicas del sistema propuesto. Por ello, el número de fuentes VoIP varía durante el

tiempo simulado para comprobar que este mecanismo se adapta a las circunstancias cambiantes –la carga en este caso–.

Rendimiento del Método del Territorio

Para cada una de las cargas (*media*, *alta* y *muy alta*) se muestra la calidad experimentada por los usuarios de los diferentes servicios, la justicia de esta arquitectura entre los diferentes servicios y el porcentaje de recursos destinados a cada clase de tráfico.

Los resultados del experimento de *carga media* se detallan en las Figuras 5.4, 5.5 y 5.6.

La primera figura muestra cómo evoluciona la calidad que experimentan los usuarios de los servicios considerados. Se puede apreciar que la QoE para el servicio de descarga progresiva de vídeo es perfecta, así como la calidad para la navegación web se mantiene entre *buena* y *regular* ($MOS \approx 3.5$) para todo el rango de carga simulado. Estos resultados se deben a la capacidad mínima de los territorios de las clases *TCP de tiempo real* e *interactiva*, respectivamente, que es suficiente para los usuarios de ambos tipos de tráfico. Nótese que un usuario de vídeo sobre TCP genera 330 Kbps –según el modelo utilizado– y una sesión web produce poca carga debido a los altos tiempos de lectura.

Esto no ocurre para el resto de servicios, que ven su calidad degradada conforme aumenta el número de usuarios de VoIP, consiguiendo una calidad *regular* ($MOS \approx 2$) para 30 llamadas de voz. El MOS experimentado en las descargas de ficheros comienza con un valor *excelente* (velocidad de transmisión superior a 1 Mbps), pero decae hasta una calidad *regular* (correspondientes a aproximadamente 400 Kbps). Algo similar ocurre con el tráfico de *streaming* de vídeo, que comienza con una calidad *buena* y decrece también hasta *regular*.

Esta reducción se puede explicar por el mecanismo de gestión de los territorios. Cuando se utilizaba sólo el planificador MSS, ambos servicios conseguían una calidad *buena* hasta con 30 llamadas de voz. Sin embargo, la política adoptada para el Método de los Territorios hace que la clase conversacional (servicio VoIP) tenga la máxima prioridad y pueda *robar* recursos del resto de clases (excepto de sus capacidades

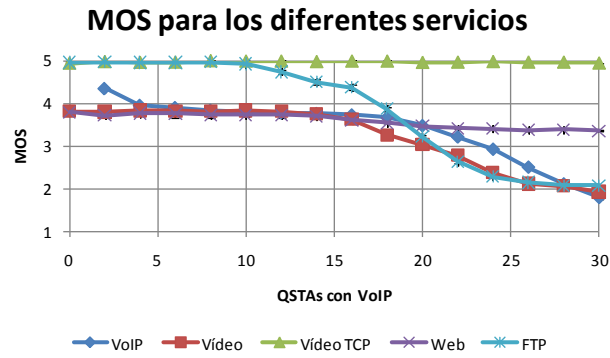


Figura 5.4. Calidad experimentada para los diferentes servicios utilizando el Método de los Territorios con carga de tráfico media.

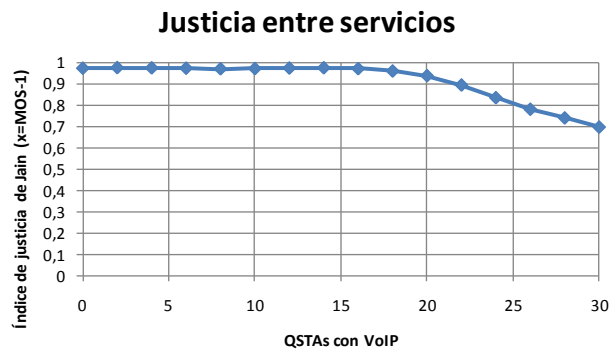


Figura 5.5. Justicia entre servicios utilizando el Método de los Territorios con carga de tráfico media.

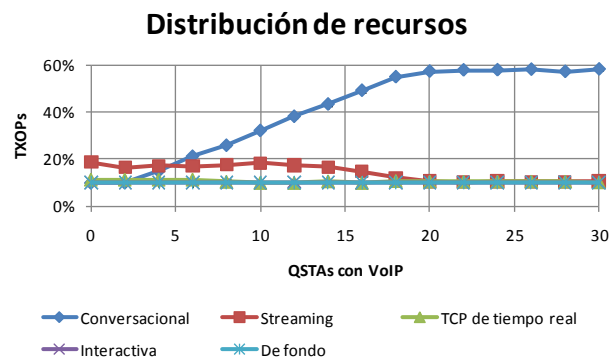


Figura 5.6. Distribución de recursos utilizando el Método de los Territorios con carga de tráfico media.

mínimas). Por ello, la QoE del resto de servicios decrece a partir, aproximadamente, de 15 llamadas (véase la Figura 5.4), ya que el territorio conversacional ha reservado prácticamente todos los recursos disponibles (véase la Figura 5.6).

Gracias a esta priorización, la capacidad para el servicio VoIP (en términos de llamadas admitidas con una calidad aceptable) aumenta hasta 20 usuarios, número superior a las 16 admitidas sin el territorio.

Lógicamente, este aumento de capacidad para las llamadas de voz tiene un precio, y es la reducción de la calidad percibida por el resto de usuarios. Este aumento en la injusticia entre servicios se ve reflejado en la Figura 5.5, ya que el índice de justicia llega hasta 0.7 mientras que con el planificador MSS no bajaba de 0.8 (más justo en términos de QoE entre servicios).

Para las simulaciones con *carga alta* (Figuras 5.7, 5.8 y 5.9) se pueden realizar el mismo tipo de observaciones, lo que demuestra la estabilidad del planificador utilizado y del Método de los Territorios.

Con esta carga, la capacidad mínima no es suficiente para hacer que el tráfico de vídeo sobre TCP mantenga una calidad *buen*a una vez que el territorio *conversacional* consume los recursos reservables, llegando a ser *mediocre* cuando el tráfico de voz aumenta.

Sin embargo, sí es suficiente para las sesiones de navegación web, debido al reducido tráfico generado (alrededor de 2 ~ 3 segundos transmitiendo y cerca de un minuto leyendo la página, i.e. sin transmitir paquetes).

Al igual que antes, la difusión de vídeo mantiene una QoE dentro del rango estipulado ($3.5 \leq \text{MOS} \leq 4$) hasta 16 ~ 18 conversaciones de voz. Esto se debe a que el número de usuarios de este tipo es el mismo (un *streaming* de vídeo) y, debido a la priorización, su calidad sólo depende del tráfico generado por clases de tráfico más prioritarias (*conversacional*). Por tanto, un aumento de carga en las clases menos prioritarias (*TCP de tiempo real, interactiva y de fondo*) no afecta a su calidad.

Esto también sucede para el tráfico VoIP, cumpliéndose así uno de los objetivos

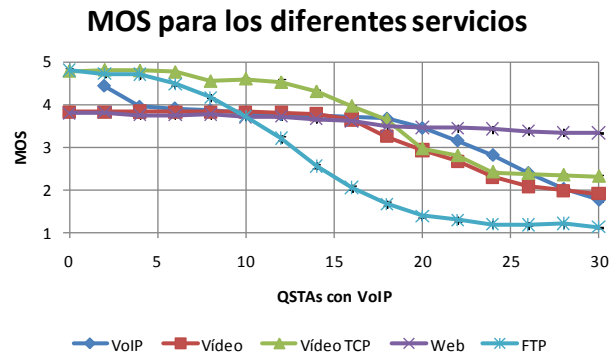


Figura 5.7. Calidad experimentada para los diferentes servicios utilizando el Método de los Territorios con carga de tráfico alta.

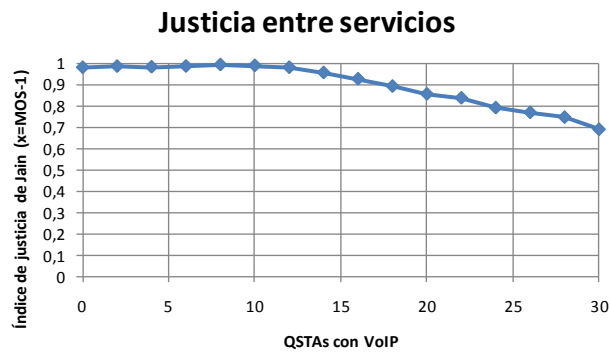


Figura 5.8. Justicia entre servicios utilizando el Método de los Territorios con carga de tráfico alta.

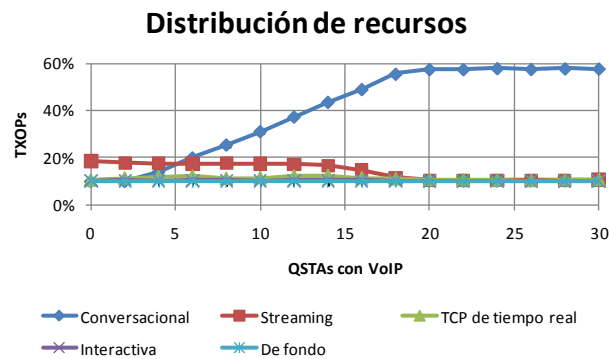


Figura 5.9. Distribución de recursos utilizando el Método de los Territorios con carga de tráfico alta.

principales del Método. El número de usuarios satisfechos de la clase más prioritaria que puede admitir el sistema (alrededor de 20, véase la Figura 5.7) es independiente del resto del tráfico. Esto se corrobora observando el porcentaje de oportunidades de transmisión para el territorio conversacional (véase la Figura 5.9), que coincide con la distribución en el experimento de *carga media*, i.e. es independiente del resto de servicios.

Nuevamente, el servicio de voz mejora utilizando la gestión basada en territorios, ya que con MSS sólo se conseguían 5 usuarios VoIP con una calidad aceptable, incrementándose con esta solución en un 300% (20 usuarios). No obstante, el servicio menos prioritario –descargas de ficheros– se ve muy degradado en comparación, pasando de una calidad *regular* (véase la Figura 4.8) a prácticamente *mala* (Figura 5.7).

Estos resultados se traducen en un mayor grado de injusticia entre servicios al aplicar esta solución, reduciéndose el índice de justicia de 0.8 (mínimo en la Figura 4.8) a 0.7 (Figura 5.8) debido al esquema por prioridades adoptado. Este aumento en la injusticia entre servicios puede sugerir que, en otros entornos, se podrían utilizar criterios alternativos que aumentarían la flexibilidad del sistema y que permitiesen al administrador de red ajustar el sistema de acuerdo con sus requisitos.

La campaña con *carga muy alta* justifica en gran medida la propuesta de esta arquitectura, ya que el tráfico prioritario consigue una calidad experimentada dentro de los límites establecidos incluso en cuando el resto del tráfico es extremadamente elevado.

Gracias a la política de reparto con derecho preferente adoptada, la capacidad para el servicio conversacional sigue siendo independiente del resto de servicios, consiguiéndose hasta aproximadamente un máximo de 20 usuarios satisfechos simultáneamente en el sistema.

La calidad de los usuarios de *streaming* de vídeo se reduce en este caso, teniendo la calidad requerida hasta con 12 llamadas de voz. Este decremento respecto a los experimentos de *carga media* y *carga alta* (hasta 16 ~ 18 conversaciones) se debe exclusivamente al aumento de usuarios de difusión de vídeo (de 1 a 2), no al aumento del tráfico perteneciente a clases menos prioritarias.

Los servicios menos prioritarios, i.e. vídeo sobre TCP y especialmente FTP, ven muy

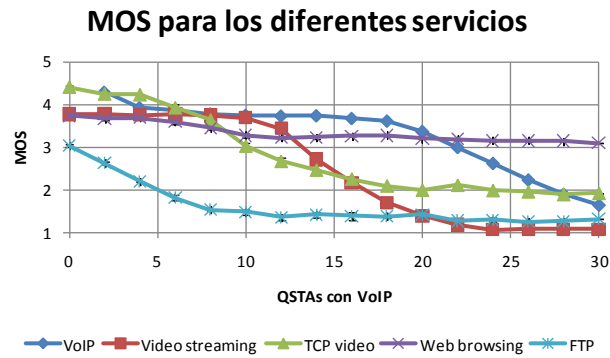


Figura 5.10. Calidad experimentada para los diferentes servicios utilizando el Método de los Territorios con carga de tráfico muy alta.

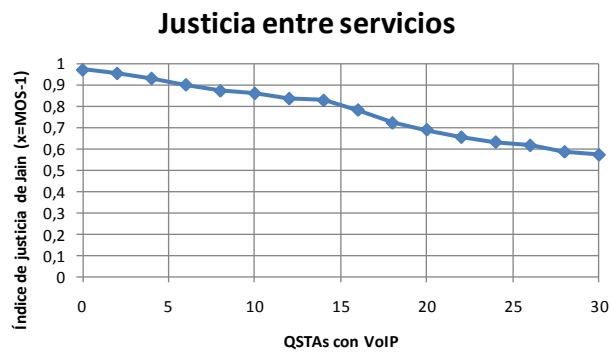


Figura 5.11. Justicia entre servicios utilizando el Método de los Territorios con carga de tráfico muy alta.

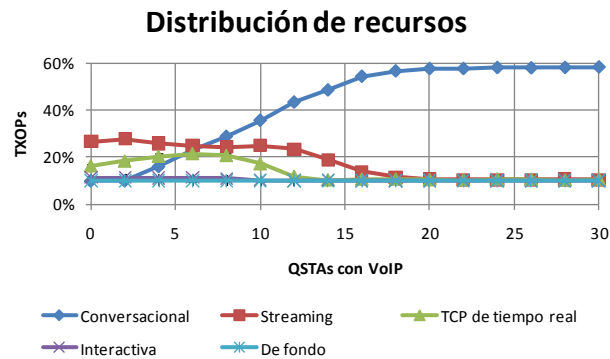


Figura 5.12. Distribución de recursos utilizando el Método de los Territorios con carga de tráfico muy alta.

reducida su calidad –*mala*–. No obstante el tráfico web consigue una experiencia similar (compárese las Figuras 4.10 y 5.10) gracias a su capacidad mínima.

Dado que el tráfico prioritario consigue una QoE independiente del resto, y que los servicios no prioritarios ven reducidos sus recursos –a favor de los prioritarios–, este experimento presenta la máxima injusticia entre tipos de tráfico, llegando a ser inferior a 0.6 (véase la Figura 5.11) mientras que con el planificador MSS no bajaba de 0.7 (Figura 4.11).

Como resumen de este análisis, se pueden extraer las siguientes conclusiones:

- La QoE de la clase de tráfico más prioritaria (*conversacional*) es independiente de las condiciones de carga –por otros servicios– del sistema.
- De forma similar, la calidad de una clase sólo depende de la carga de dicha clase y del tráfico de las clases más prioritarias.
- Las clases de tráfico menos prioritarias ven degradado su rendimiento al utilizar este mecanismo, debido a que los recursos disponibles son reservados por las clases más prioritarias. Así, la clase *de fondo* es la más afectada al ser la menos prioritaria.
- Gracias a la capacidad mínima, incluso los servicios menos prioritarios consiguen un rendimiento mínimo. Este es el caso del servicio de navegación web, que consigue una calidad aceptable gracias a su capacidad mínima y a que genera poco tráfico (incluso con varias sesiones). También FTP consigue aprox. 250 Kbps para cada usuario FTP en el caso de *carga muy alta* gracias a esta capacidad mínima –aunque su QoE sea muy baja–.

Características dinámicas del Método del Territorio

Para comprobar el comportamiento dinámico del Método del Territorio, se ha realizado una simulación basada en el experimento con *carga alta* (un usuario de *streaming* de vídeo, dos descargas progresivas de vídeo, tres sesiones web y dos descargas de ficheros), modificando el número de fuentes VoIP durante la simulación entre 1 y 30 (siguiendo el perfil de la Figura 5.13).

Las siguientes figuras muestran los resultados en el escenario considerado, para cada



Figura 5.13. Perfil de tráfico VoIP durante el experimento con carga dinámica.

uno de los servicios utilizados. Cada figura incluye cómo evoluciona con el tiempo el MOS instantáneo y promediado, así como la variación del territorio de su clase de tráfico.

La Figura 5.14 muestra el MOS instantáneo, el MOS promedio y el territorio utilizado para el tráfico VoIP (CT *conversacional*). Al ser éste el servicio más prioritario, el territorio reacciona rápidamente conforme aumenta la carga (Figura 5.13), gracias al Paso de Aumento utilizado. Sin embargo, cuando la carga decrece el territorio varía lentamente, evitando que la QoE sufra oscilaciones por falta de recursos. El MOS promediado sigue correctamente las muestras de MOS instantáneo, debiéndose sus picos más abruptos a los aumentos de carga y no al proceso de promediado. El MOS –tanto instantáneo como promediado– sólo está por debajo del rango requerido cuando hay 25 ó 30 fuentes ($t \in [180, 270]$), de forma similar a los resultados obtenidos con condiciones estáticas (Figura 5.7).

En consecuencia, se comprueba que para la dinámica considerada la asignación de recursos varía de forma adecuada para que se mantenga la QoE, estando por debajo del rango requerido sólo cuando no hay más recursos disponibles en el sistema.

Las siguientes figuras (Figuras 5.15, 5.16, 5.17 y 5.18) muestran las mismas estadísticas para los servicios de *streaming* de vídeo, vídeo sobre TCP, navegación web y descarga de ficheros.

Debido a que la mayoría de los recursos han sido asignados a VoIP, sólo se aprecian ligeros cambios en el territorio de la CT *streaming* cuando hay pocas conversaciones de voz, intentando mantener la calidad experimentada de los usuarios de difusión de vídeo. Sin embargo, esto no es posible cuando el servicio conversacional utiliza los recursos

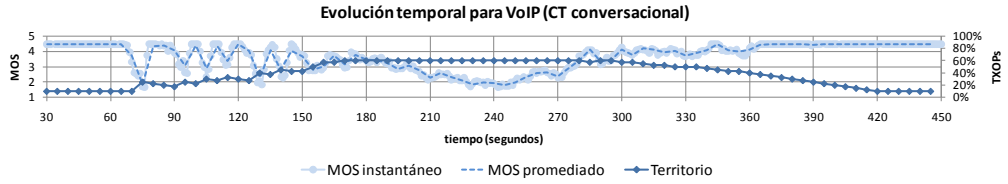


Figura 5.14. Evolución temporal de los resultados utilizando el Método de los Territorios para VoIP.

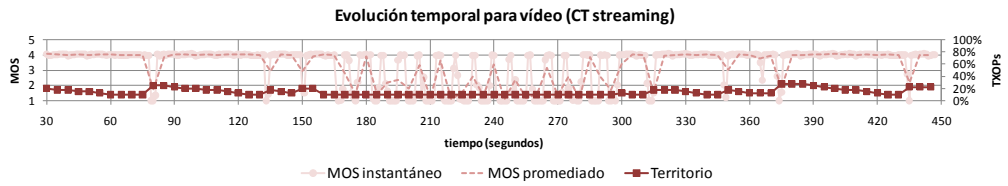


Figura 5.15. Evolución temporal de los resultados utilizando el Método de los Territorios para streaming de vídeo.

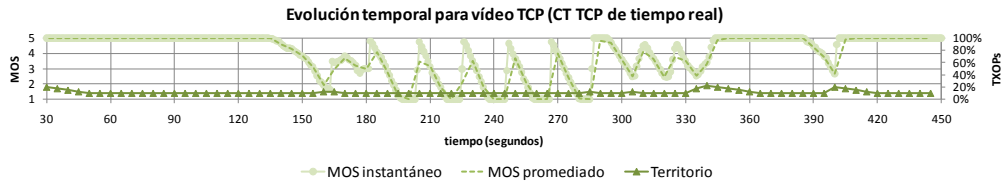


Figura 5.16. Evolución temporal de los resultados utilizando el Método de los Territorios para vídeo sobre TCP.

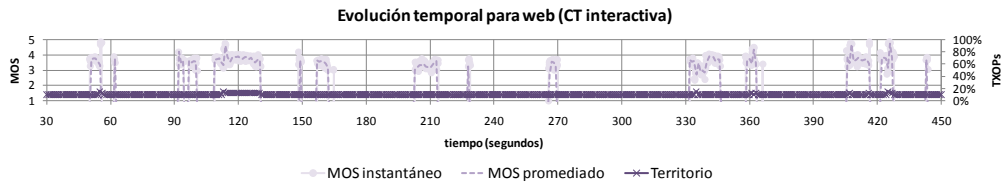


Figura 5.17. Evolución temporal de los resultados utilizando el Método de los Territorios para tráfico web.

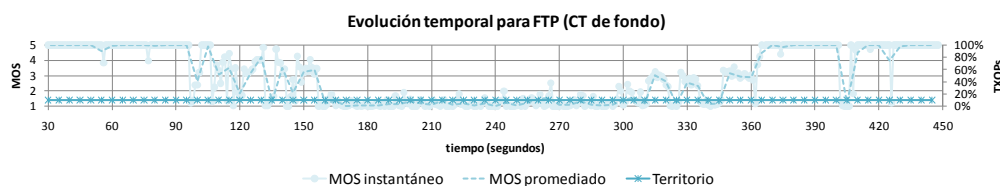


Figura 5.18. Evolución temporal de los resultados utilizando el Método de los Territorios para FTP.

disponibles, reduciéndose drásticamente el MOS promedio e instantáneo ($t \in [150, 300]$, i.e. 20 o más conversaciones).

El vídeo sobre TCP sólo puede utilizar su capacidad mínima (salvo algunos pequeños períodos con poco tráfico VoIP) y las TXOPs no utilizadas, como se observa en la Figura 5.16. Al decrecer éstas con la carga de voz, la QoE de este servicio también decrece (nuevamente para el rango con 20 o más usuarios VoIP).

La navegación web se limita a utilizar su capacidad mínima, si bien se observa un leve crecimiento del territorio cuando hay recursos libres y el tiempo de descarga permite actuar al Método de los Territorios, como se observa alrededor de $t = 120$ s. En cualquier caso, queda claro que su rendimiento se debe al establecimiento de una reserva mínima (10% de las TXOPs).

Finalmente, el territorio para la CT *de fondo* no varía durante toda la simulación. Las sesiones FTP utilizan su capacidad mínima y las TXOPs sin paquetes a transmitir por cualquier otra clase de tráfico, ya que es la menos prioritaria. Estas TXOPs no utilizadas son suficientes para conseguir una calidad *excelente* cuando hay pocas conversaciones ($t < 90$ s y $t > 360$), pero decrece al aumentar el tráfico VoIP debido a no quedan turnos de transmisión sin utilizar.

Los resultados analizados permiten comprobar, para este escenario, el correcto comportamiento dinámico del Método de los Territorios ante situaciones cambiantes –variación de la carga más prioritaria en este ejemplo–.

A modo de resumen, se pueden extraer las siguientes conclusiones:

- Los territorios evolucionan rápidamente ante decrementos de la calidad experimentada por los usuarios, e.g. cuando aumenta rápidamente el número de conversaciones de voz.
- Sin embargo, se reducen lentamente cuando no necesitan tantos recursos con el objetivo de evitar oscilaciones en la QoE de los servicios afectados –tal como se puede observar cuando se reduce rápidamente el tráfico VoIP–.
- Al igual que en los experimentos *estáticos*, los servicios más prioritarios consiguen los recursos disponibles cuando lo necesitan. De esta forma, su rendimiento es independiente del tráfico de las clases menos prioritarias.
- Los servicios menos prioritarios utilizan los recursos sobrantes, estando su rendimiento muy influenciado por el tráfico más prioritario.
- Gracias a la capacidad mínima definida en cada territorio, todos los servicios consiguen un rendimiento mínimo incluso cuando hay muchos usuarios de servicios más prioritarios.

5.3 Conclusiones

En este capítulo se ha presentado una arquitectura para la gestión de recursos radio denominada el Método de los Territorios.

Este mecanismo actúa de recubrimiento sobre la asignación de turnos realizada por el planificador de paquetes utilizado en HCCA, con los objetivos de garantizar la QoE de los servicios más prioritarios y de aumentar la utilización del ancho de banda disponible. Para ello, basa sus decisiones en la calidad experimentada por los usuarios de estos servicios, aumentando o disminuyendo los recursos asignados para los mismos en función del MOS percibido. Debido al esquema adoptado de *asignación con derecho preferente*, las CTs más prioritarias pueden *atribuirse* recursos de otras clases con menos prioridad. Sin embargo, gracias a la definición de una capacidad mínima para cada clase de tráfico se garantiza un rendimiento mínimo para cada tipo de servicio.

El rendimiento de esta solución ha sido evaluado y analizado tanto para condiciones estáticas –sin variaciones en el número de usuarios durante la simulación– como

dinámicas.

En el primer caso se ha mostrado su correcto funcionamiento para diferentes niveles de carga –*media, alta y muy alta*–, comprobándose la priorización de servicios (e.g. el rendimiento de los flujos VoIP fue independiente de la carga del resto de servicios) y el rendimiento mínimo de los servicios (e.g. las sesiones web conseguían una *buena* calidad percibida gracias a la capacidad mínima de su territorio).

En el segundo caso se ha examinado el comportamiento dinámico de la solución propuesta en situaciones cambiantes, modificándose –en el escenario simulado– el tamaño de los territorios de forma adecuada según la carga por servicio. Este análisis ha permitido evidenciar la rápida respuesta ante crecimientos de tráfico y la reacción más gradual en el caso de decrementos (evitándose posibles inestabilidades de QoE y territorios).

Por último, se ha realizado –para los experimentos con carga estática– una comparativa con el rendimiento obtenido con el planificador MSS sin utilizar territorios. Esta comparación ha mostrado el mejor comportamiento del Método de los Territorios para los servicios de tiempo real, siendo su rendimiento independiente de la carga de servicios menos prioritarios. Así, incluso con *carga muy alta* el sistema ha mostrado una capacidad muy superior para el servicio VoIP –hasta 20 usuarios satisfechos–, siendo éste uno de los objetivos requeridos para esta arquitectura y que justificaba su propuesta (e.g. con MSS y esta carga, ninguna conversación conseguía la calidad requerida).

Estas conclusiones permiten afirmar que el esquema propuesto cumple los requisitos iniciales (priorización del tráfico de tiempo real y utilización del ancho de banda en función de la calidad percibida). Sin embargo, se podrían haber adoptado otros criterios –para el reparto de la capacidad del medio inalámbrico– con objetivos diferentes pero igualmente válidos. Ésta es la aproximación adoptada en el siguiente capítulo.

Capítulo 6

Políticas de gestión de recursos basadas en Calidad de Experiencia

El *Método de los Territorios* expuesto en el Capítulo 5 fue diseñado inicialmente para permitir una priorización de las clases de tráfico definidas. De esta forma, el tráfico con restricciones estrictas de tiempo (*conversacional* y *streaming*) no verá afectado su rendimiento por otras aplicaciones sin este tipo de requisitos (clases de tráfico *TCP de tiempo real, interactiva y de fondo*).

Si bien esta política es interesante en términos de garantías de calidad de servicio, en entornos diferentes pueden existir otros criterios alternativos igualmente válidos. De esta manera, al permitir la creación de nuevas políticas de gestión de recursos en el Método de los Territorios se consigue una mayor flexibilidad ya que el operador de la red podrá ajustar el sistema de acuerdo con sus requisitos. Por ello, en este capítulo se proponen y estudian diferentes políticas de gestión de recursos, incluyendo la inicialmente considerada en el capítulo anterior. En particular, se consideran las siguientes políticas:

- *Con derecho preferente*: su objetivo es priorizar el tráfico con requisitos de

- tiempo real
- *Sin derecho preferente*: definida para adaptarse a la evolución temporal del tráfico
 - *Justa en términos de calidad de experiencia*: definida para igualar la Calidad de Experiencia de las diferentes clases de tráfico consideradas
 - *Aumento de beneficio*: su objetivo es incrementar los ingresos obtenidos con la red

6.1 Reparto con derecho preferente

Ésta es la política inicialmente utilizada por el Método de los Territorios descrita en el Capítulo 5. Su objetivo es asegurar que el rendimiento de los servicios de tiempo real no se vea afectado por el tráfico elástico a la vez que soportar de forma eficiente la mezcla de diferentes tipos de tráfico. Para ello, las clases de tráfico más prioritarias pueden apropiarse recursos de la capacidad reservada para CTs con menor prioridad. De esta forma, los servicios conversacionales pueden utilizar siempre todas las oportunidades de transmisión excepto aquellas que pertenecen a la capacidad mínima de otras clases.

Para describir formalmente esta política, la Figura 6.1 muestra el correspondiente diagrama de flujo de esta aproximación. La caja “*sin recursos libres*” establece las diferencias entre las distintas políticas, siendo el resto del esquema –el Método de los Territorios– común a todas las políticas.

6.2 Reparto sin derecho preferente

Esta aproximación supone que ninguna clase de tráfico tiene prioridad sobre el resto. En consecuencia, si un recurso ha sido reservado por una CT, éste no puede ser *robado* por otra clase. Al igual que en la política anterior, se sigue garantizando un cierto nivel de servicio para cada clase de tráfico a través de la definición de su capacidad mínima, que es ajustable.

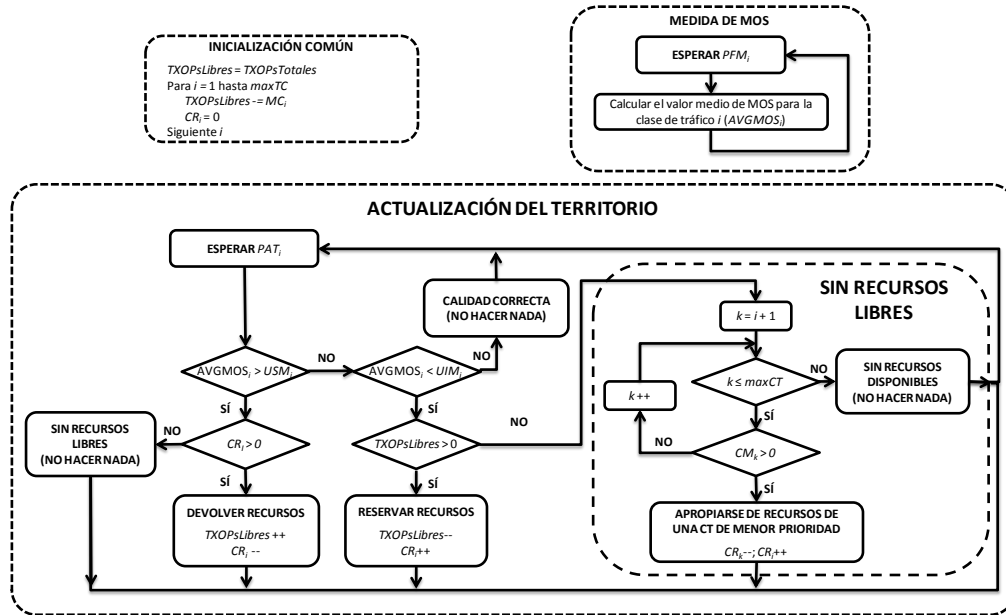


Figura 6.1. Diagrama de flujo del Método de los Territorios utilizando la política de reparto con derecho preferente.

Por tanto, esta política asigna los recursos teniendo en cuenta la calidad experimentada por las distintas CTs. Si el valor de MOS global de una clase (MOS promediado para todos los flujos pertenecientes a dicha clase) está por debajo de su *UIM* (véase la Sección 5.1), su territorio crecerá si quedan recursos libres. De forma similar, si el MOS promedio está por encima del parámetro *USM*, se liberará parte de los recursos reservados. Al tratarse de una política *el primero que llega es el primero servido*, no existen prioridades entre las distintas clases de tráfico.

Su diagrama de flujo es semejante al de la Figura 6.1, sustituyendo el recuadro “sin recursos libres” de dicho diagrama por el mostrado en la Figura 6.2.

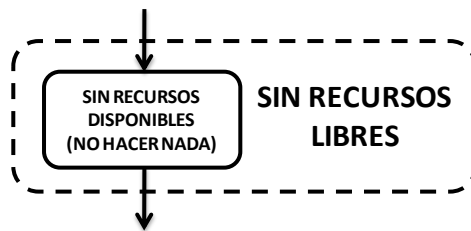


Figura 6.2. Diagrama de flujo de la política de *reparto sin derecho preferente*.

6.3 Reparto justo en términos de Calidad de Experiencia

Ninguno de las dos políticas anteriores puede asegurar que exista justicia entre las diferentes clases de tráfico respecto a la calidad percibida por sus usuarios. La política de *reparto justo en términos de Calidad de Experiencia* se comporta de forma similar a la política de *reparto sin derecho preferente* siempre que el valor de MOS de todas las CTs sea adecuado. Sin embargo, intentará igualar los valores de MOS de las diferentes CTs cuando no haya suficientes recursos. En esta situación, las CTs con menor MOS podrán atribuirse recursos de las CTs que experimenten mayor calidad.

Tal como ocurría con la política anterior, en este caso el diagrama de flujo consiste en reemplazar la caja “*sin recursos libres*” por la mostrada en la Figura 6.3.

6.4 Reparto para aumento de beneficio

La política de *reparto para aumento de beneficio*, de forma similar al *reparto justo en términos de Calidad de Experiencia*, sólo actúa cuando la calidad de una CT está por debajo de su umbral de QoE. En ese caso, las clases de tráfico se ordenan según su *Eficiencia de Beneficio (EB_i)*, i.e. los ingresos por unidad de recurso (Ecuación 6.1).

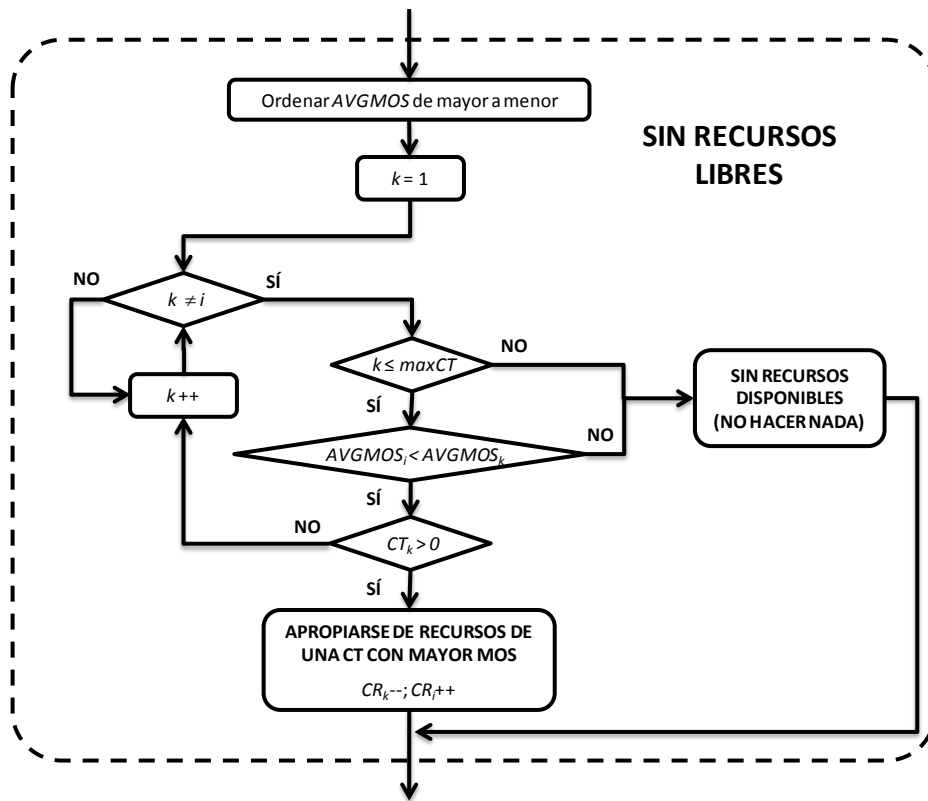


Figura 6.3. Diagrama de flujo de la política de reparto justo en términos de Calidad de Experiencia.

El indicador EB_i se calcula dividiendo el MOS global de la CT i –se ha supuesto que un usuario está más dispuesto a pagar si está satisfecho con el servicio– entre la cantidad de recursos utilizados (CR_i) –un servicio es más eficiente cuanto menos recursos utiliza para un beneficio dado–. El resultado se pondera mediante un peso W_i que es específico de la clase y puede ser ajustado por el administrador de red para reflejar sus políticas económicas. Una CT que proporcione un mayor beneficio podrá apropiarse de recursos pertenecientes a una CT con menor beneficio (véase la Figura 6.4).

$$EB_i = W_i \cdot \left(\frac{AVGMOS_{i-1}}{CR_i} \right) \quad (6.1)$$

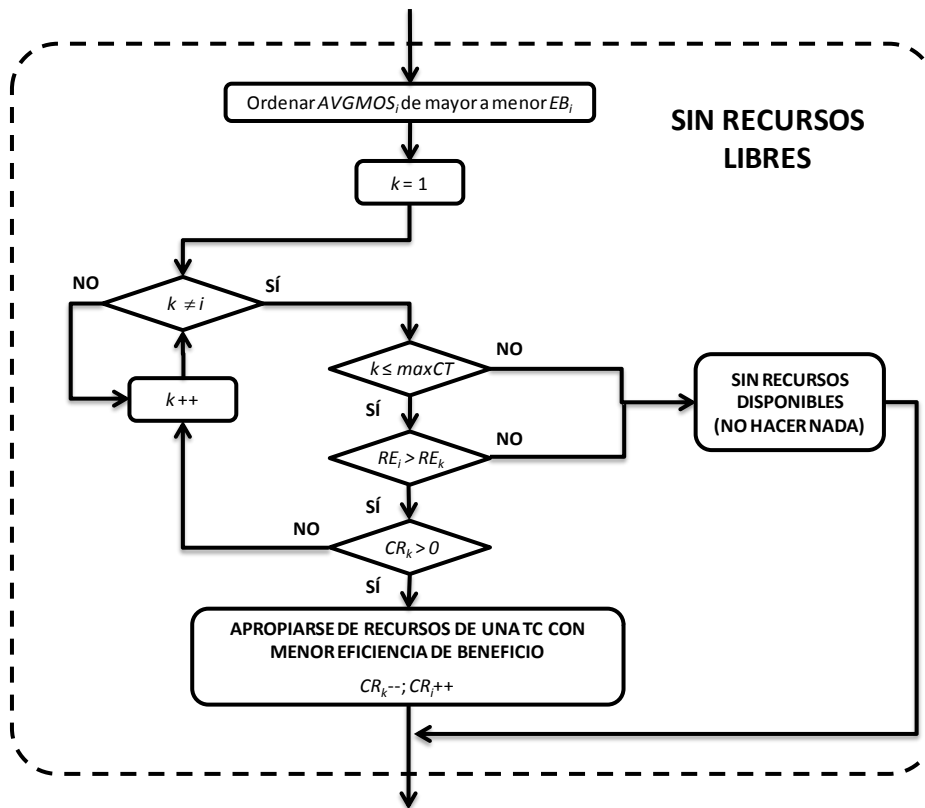


Figura 6.4. Diagrama de flujo de la política de reparto para aumento de beneficio.

6.5 Evaluación de rendimiento

El rendimiento de las diferentes políticas propuestas ha sido evaluado a través de simulaciones. La Sección 6.5.1 detalla la configuración del entorno de simulación, mientras que en la Sección 6.5.2 se analizan los resultados obtenidos.

6.5.1 Condiciones experimentales

El entorno de simulación está basado en el descrito en la Sección 5.2.1 para la evaluación del rendimiento del Método de los Territorios. Además, las políticas descritas han sido implementadas en la herramienta NS-2 como parte del Método de los Territorios, siendo su selección configurable.

Las simulaciones realizadas han sido *estáticas*, i.e. el número de usuarios de cada servicio permanece constante durante toda la simulación, utilizándose las *cargas media*, *alta* y *muy alta* definidas en la Sección 4.3.1.

6.5.2 Resultados

Este apartado analiza los resultados obtenidos con las diferentes políticas definidas en este capítulo.

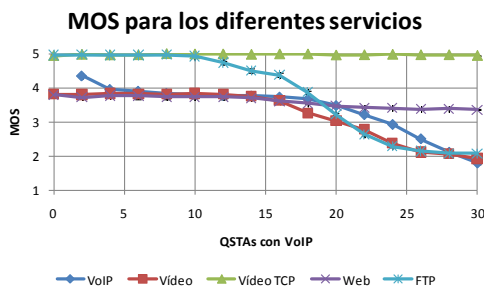
Resultados de la política de reparto con derecho preferente

El análisis comienza mostrando los resultados del esquema *con derecho preferente* –resumen de la Sección 5.2.2, dado que es la misma política– en función del aumento de tráfico (número de estaciones con tráfico VoIP). De esta forma, para las diferentes cargas simuladas, se proporciona la evolución del MOS al aumentar el número de conversaciones, así como la compartición de recursos (en términos de porcentaje de TXOPs utilizadas) para las diferentes clases de tráfico consideradas. También se incluye la métrica de beneficio (B_i , definida por la Ecuación 6.2) individualmente por tipo de servicio y de forma agregada.

$$B_i = W_i \cdot (AVGMOS_i - 1) \quad (6.2)$$

La Figura 6.5 muestra los resultados para el experimento de *carga media*, mientras que la Figura 6.6 los presenta para *carga alta*. Finalmente la Figura 6.7 los muestra para el experimento de *carga muy alta*. Las conclusiones que se pueden extraer a partir de la observación de estas curvas son similares para las diferentes cargas, lo que pone de manifiesto la estabilidad de esta política.

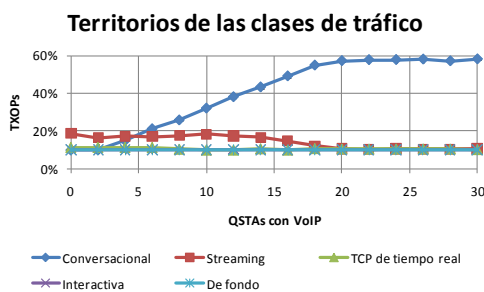
La distribución de los territorios (subfiguras (c)) muestra que la clase conversacional (VoIP) utiliza tantos recursos como sean necesarios, aumentando conforme se incrementa su carga. Si los servicios conversacionales no reservan todos los recursos disponibles (debido a la buena experiencia del usuario final), los servicios de *streaming* (*streaming* de vídeo) se los adjudican. Cualquier otra CT (*TCP de tiempo real, interactiva o de fondo*)



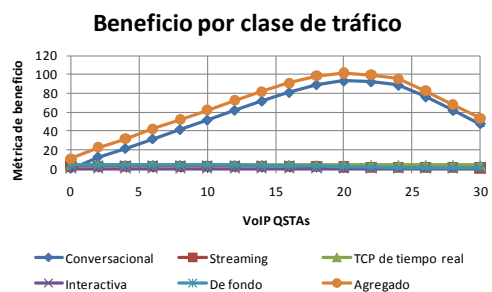
(a)



(b)



(c)



(d)

Figura 6.5. Resultados usando la política de *reparto con derecho preferente* para carga media.

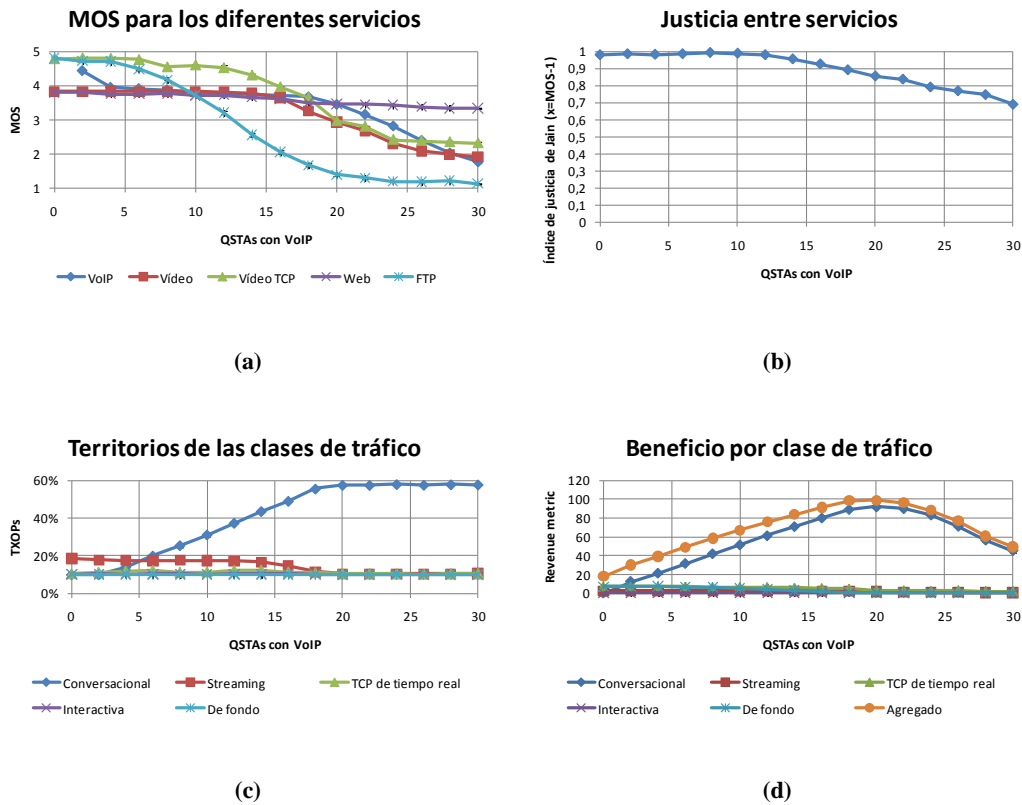


Figura 6.6. Resultados usando la política de reparto con derecho preferente para carga alta.

sólo puede utilizar su capacidad mínima (10% de las TXOPs) en la mayoría de los casos, dado que los servicios más prioritarios se han atribuido los recursos existentes.

Estos resultados son esperables dada una política de *reparto con derecho preferente*, donde el rendimiento de la CT con mayor prioridad (*conversacional*) es independiente del resto (como se mostró en el Capítulo 5).

Los valores de MOS (subfiguras (a)) dependen directamente del servicio y de los recursos asignados. Para estas condiciones experimentales concretas, el tráfico VoIP mantuvo un buen rendimiento (MOS superior a $UIM_I=3.5$) hasta aprox. 20 llamadas independientemente de la carga de tráfico, a partir de las cuales se saturó el territorio *conversacional* y su calidad se redujo.

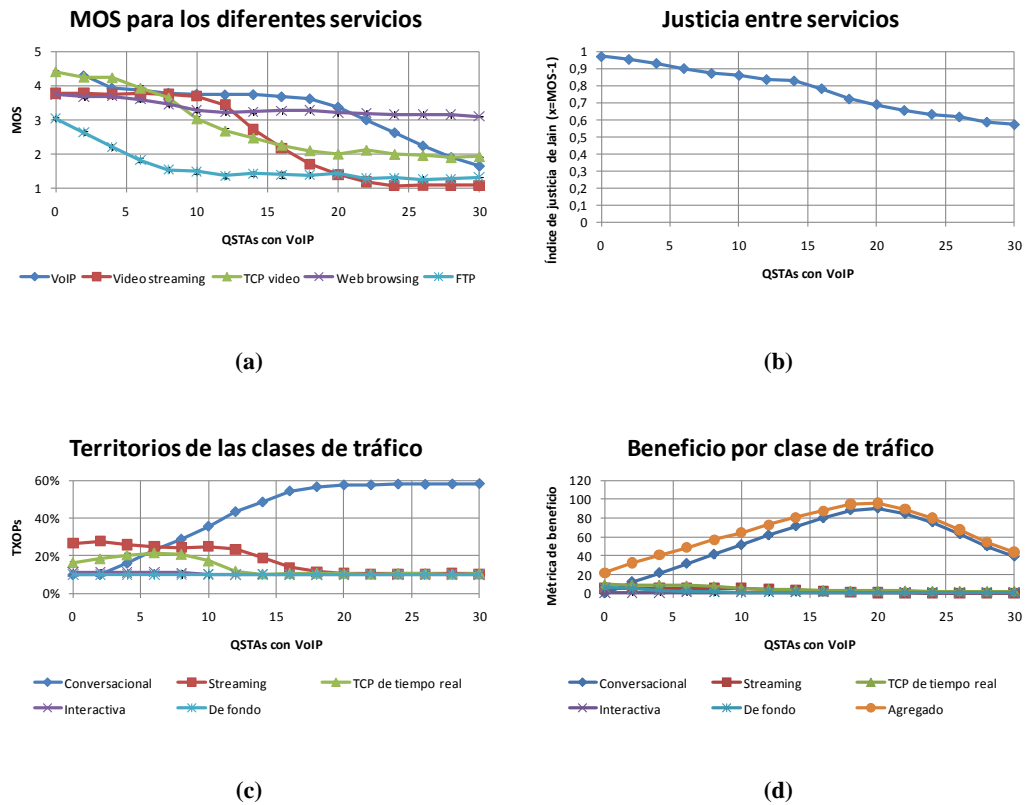


Figura 6.7. Resultados usando la política de reparto con derecho preferente para carga muy alta.

Para los tres escenarios considerados, los usuarios de *streaming* de vídeo percibieron una buena calidad subjetiva ($MOS \geq 3.5$) para cargas respectivas de hasta 16, 16 y 12 conversaciones de voz. Este valor es igual para las *cargas media* y *alta* porque el número de usuarios de *streaming* de vídeo es el mismo –uno– y su calidad no depende de servicios menos prioritarios.

De forma similar, los usuarios de vídeo sobre TCP experimentaron pocas pausas (aceptables en términos de MOS) debidas a la falta de datos en sus *buffers* con cualquier número de conversaciones para *carga baja*, hasta con 18 llamadas de voz para *carga alta* y sólo con 8 para *carga muy alta*.

Las sesiones de navegación web tuvieron un comportamiento casi ideal en todo el rango de llamadas de voz simulado (MOS cercanos a 3.5, si bien decreció ligeramente para *carga muy alta*). Esto se debe a la capacidad mínima garantizada para la clase de tráfico *interactivo*, suficiente para 5 sesiones ya que el tráfico HTTP generado es muy bajo (largos períodos de lectura de las páginas).

En este escenario, FTP sólo utiliza su capacidad mínima y las TXOPs sin paquetes a transmitir, por lo que su rendimiento decrece conforme la carga VoIP aumenta. FTP tuvo un rendimiento adecuado (MOS ≥ 3.5) hasta con 20 usuarios VoIP para *carga media*, hasta con 10 usuarios VoIP para *carga alta* y no consiguió un nivel adecuado con *carga muy alta* en ningún caso. Sin embargo, su MOS no bajó de 1.3 (aproximadamente 280 Kbps) gracias a su capacidad mínima.

Las subfiguras (d) muestran claramente que VoIP es el servicio más eficiente en términos de beneficio (ingresos por unidad de recurso), seguido del servicio vídeo sobre TCP, FTP, *streaming* de vídeo y navegación web con unas ganancias considerablemente inferiores. Inicialmente, el beneficio aumenta linealmente con el número de flujos de voz, dado que su MOS permanece constante y alto (en torno a 3.8). Cuando el territorio *conversacional* se satura y el rendimiento de VoIP disminuye, su beneficio también lo hace dado que los recursos son constantes –y el aumento de llamadas no compensa la reducción de MOS–. El máximo beneficio obtenido (para el caso de 20 llamadas) es 101, 99 y 96 respectivamente para las diferentes cargas.

Por último, las curvas de justicia (subfiguras (b)) indican que ésta disminuye conforme aumenta el tráfico VoIP para una carga dada, llegando a un mínimo de 0.7, 0.7 y 0.6 en los tres experimentos considerados. Estos resultados muestran que la injusticia crece conforme alguno de los servicios deja de tener una calidad aceptable, siendo éste hecho más notable para cargas altas. Estos niveles de injusticia se deben al esquema por prioridades utilizado en esta política.

Resultados de la política de reparto sin derecho preferente

En este apartado se exponen las mismas métricas que para el caso anterior, pero

considerando en este caso la política de *reparto sin derecho preferente*. Los resultados son muy dependientes de la evolución temporal de las fuentes de tráfico, ya que se sigue una aproximación de *el primero en llegar, el primero en ser servido* (*first come, first served*). Este es el motivo del comportamiento impredecible en la evolución de MOS (subfiguras (a)) y de los territorios (subfiguras (c)) una vez que escasean los recursos. En cualquier caso, los resultados para navegación web siempre son buenos (MOS superior a 3.5) debido a la capacidad mínima del territorio *interactivo*, de forma similar a los resultados obtenidos para la política de reparto *con derecho preferente*.

El valor máximo de beneficio agregado es de 91, 70 y 54 unidades para los distintos niveles de carga, correspondientes a 20, 14 y 14 llamadas de voz respectivamente, siendo

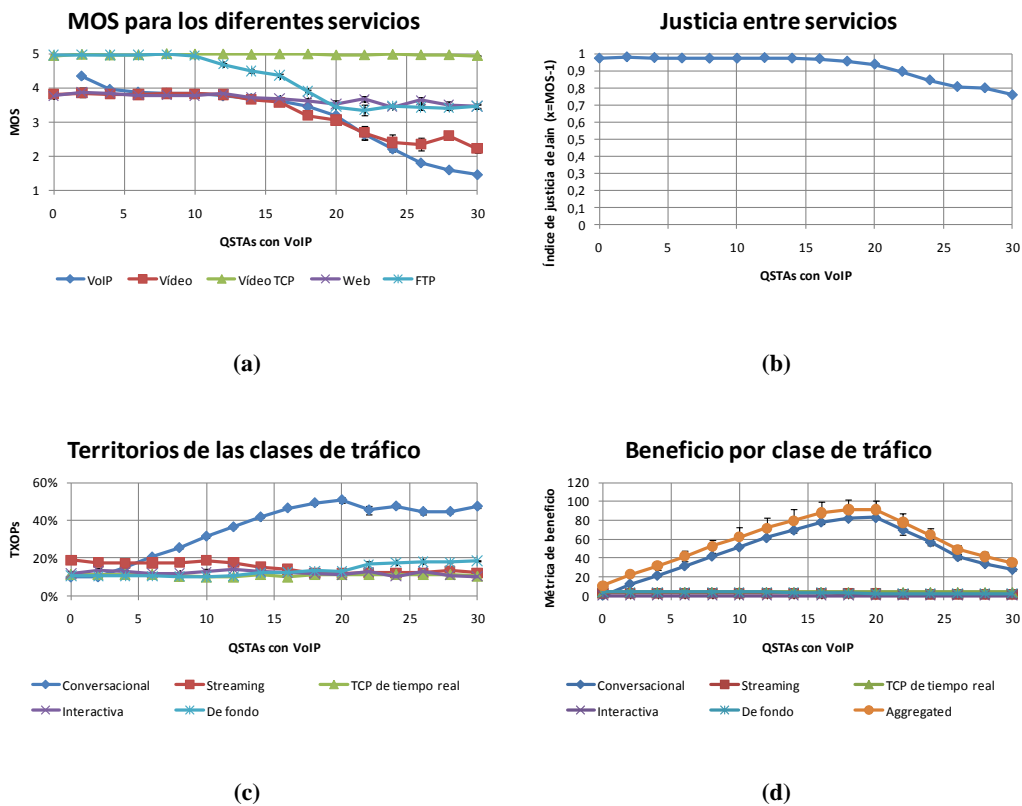


Figura 6.8. Resultados usando la política de *reparto sin derecho preferente* para carga media.

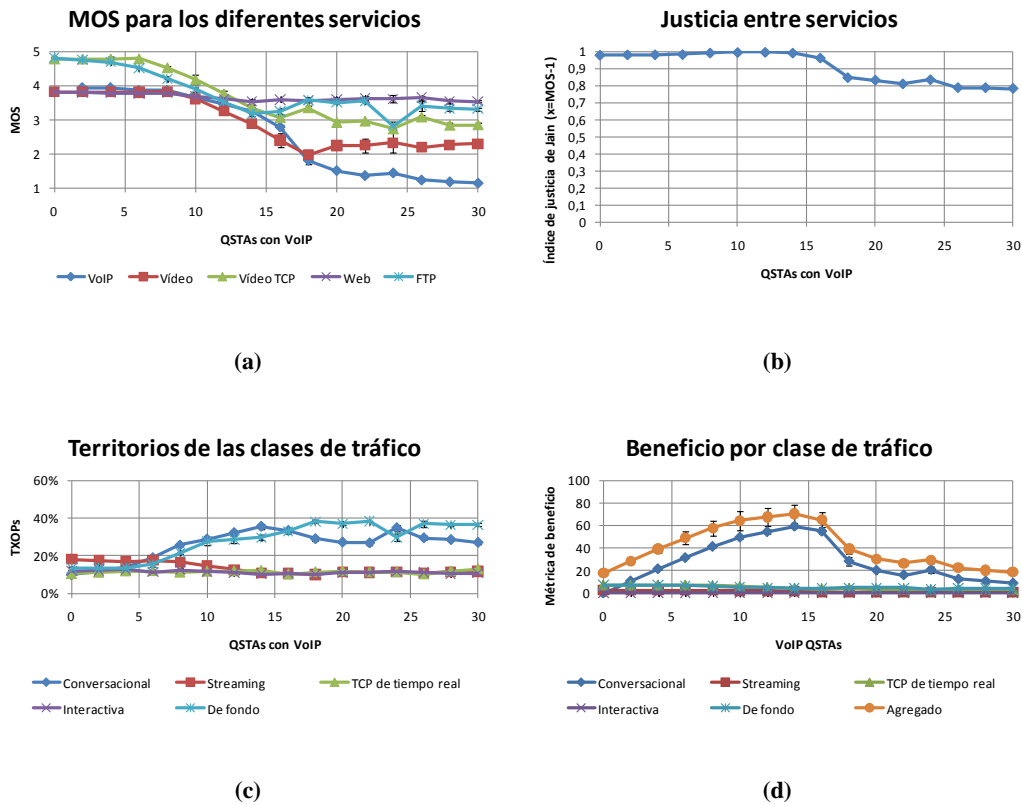


Figura 6.9. Resultados usando la política de reparto *sin derecho preferente* para carga alta.

este servicio el que más contribuye. El orden del resto de servicios en términos de beneficio es, en orden decreciente, FTP, vídeo sobre TCP (con un beneficio semejante), *streaming* de vídeo y finalmente el tráfico web (con un beneficio prácticamente nulo).

En términos de justicia, este esquema es más equitativo entre los servicios si se compara con la política de reparto *con derecho preferente*, ya que la asignación de servicios depende del orden en que lo soliciten y no de prioridades. El valor mínimo del índice (cuando el número de llamadas de voz es muy alto) es de 0.8 en los casos de *carga media* y *alta*, y alrededor de 0.6 ~ 0.7 para *carga muy alta*.

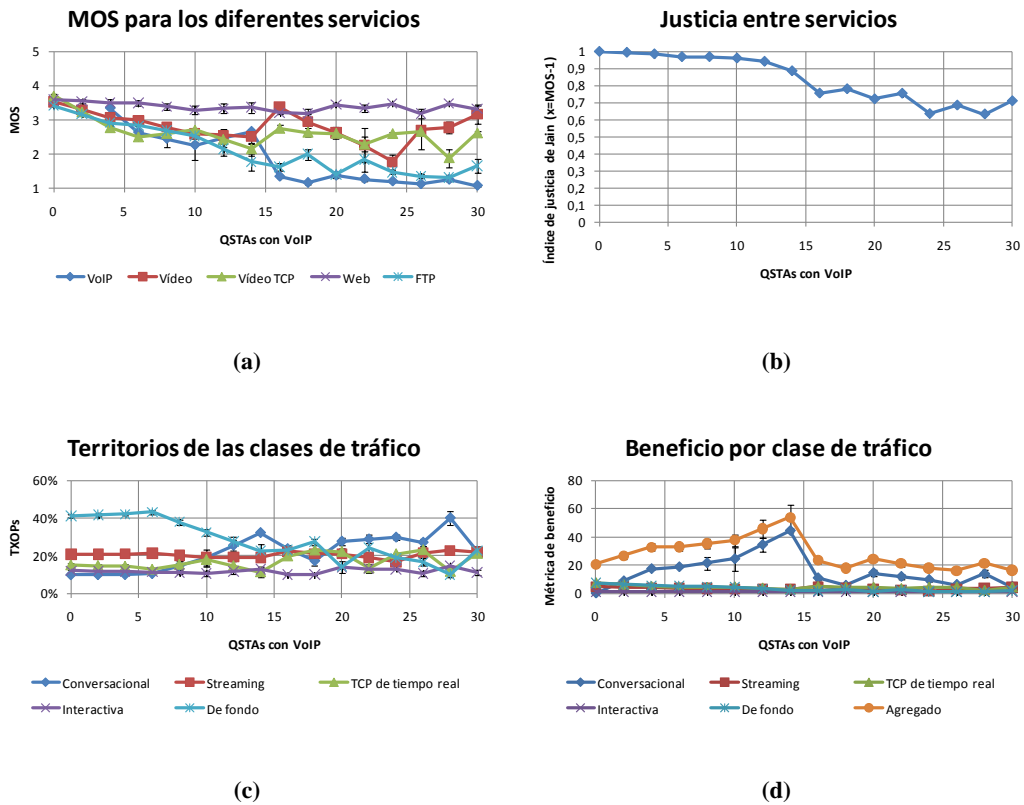


Figura 6.10. Resultados usando la política de reparto *sin derecho preferente* para carga muy alta.

Resultados de la política de reparto justo en términos de Calidad de Experiencia

En este caso, los recursos se distribuyen de forma que los distintos servicios tienen una calidad similar –en el caso de faltar recursos–. Como muestran las subfiguras (a), los valores de MOS de los servicios VoIP, *streaming* de vídeo y FTP son muy parecidos (hecho especialmente apreciable con *carga alta* y *muy alta*) cuando la carga aumenta y el MOS está por debajo de sus respectivos $UIM_i = 3.5$. De hecho, el tamaño de sus territorios (subfiguras (c)) varía para conseguir que sus valores de MOS se igualen. Sin embargo, los valores para vídeo sobre TCP y navegación web son superiores debido a sus capacidades mínimas, lo que garantiza un rendimiento mínimo para una carga de tráfico dada. De esta forma, el MOS del servicio de vídeo sobre TCP es siempre superior al

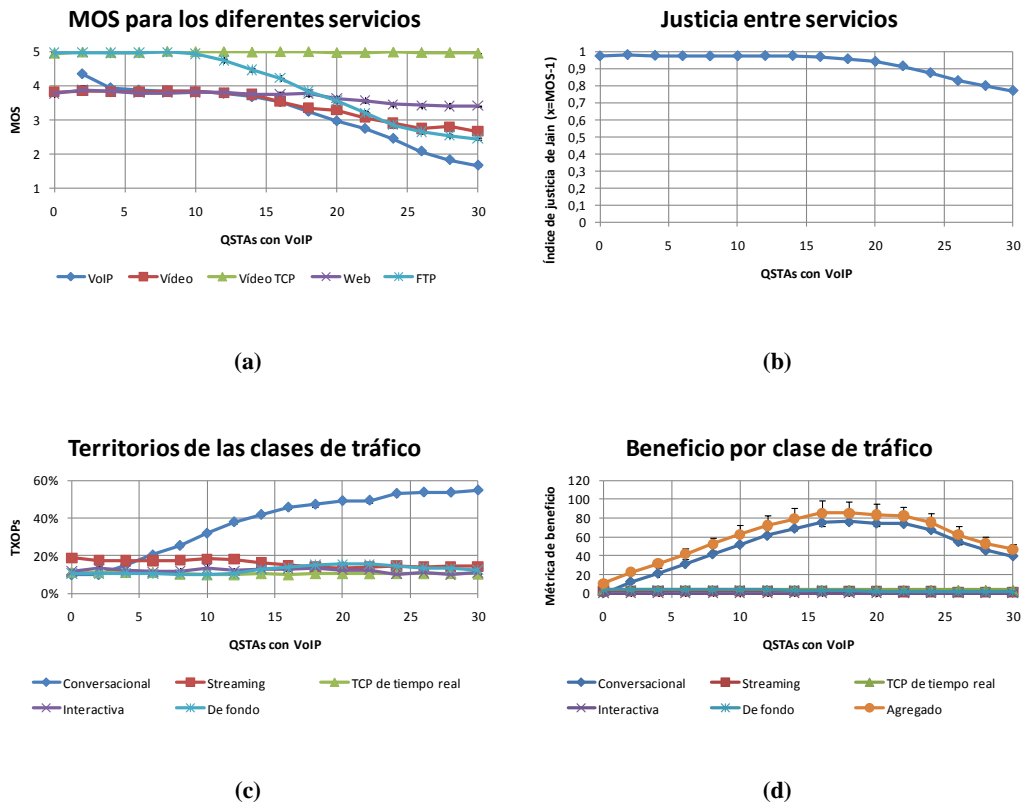


Figura 6.11. Resultados usando la política de reparto justo en términos de Calidad de Experiencia para carga media.

de estos servicios (*excelente* para *carga media*, *bueno* o *regular* para carga alta y llegando a *mediocre* para carga *muy alta*), mientras que los usuarios de tráfico web siempre experimentan una calidad dentro –o cerca– del rango aceptable (MOS superior a 3.2 en todos los casos).

Esta justicia en la gestión de los recursos se ve reflejada en las subfiguras (b), donde el valor mínimo del índice de Jain es aproximadamente de 0.8 para *cargas media* y *alta*. Sin embargo, este valor decrece hasta 0.6 para *carga muy alta*. Esto no es debido a la política utilizada sino a las capacidades mínimas de los territorios, que hacen que el tráfico web consiga la calidad requerida –no siendo

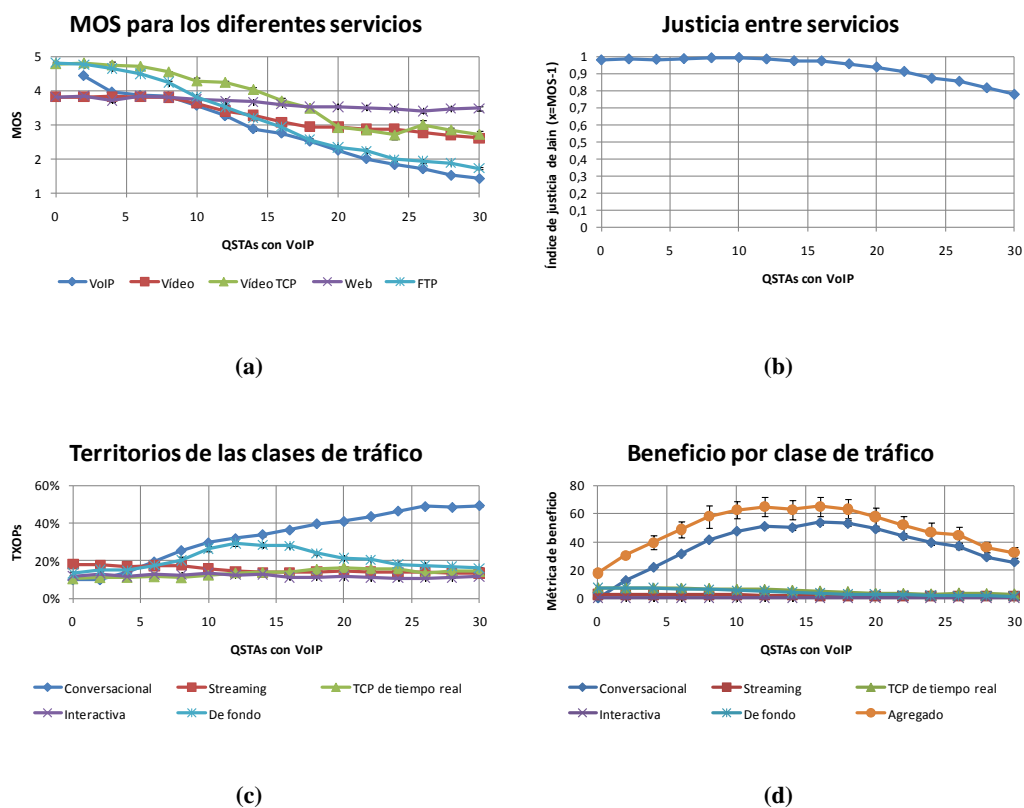


Figura 6.12. Resultados usando la política de reparto justo en términos de Calidad de Experiencia para carga alta.

así para los otros servicios—. Esa diferencia hace que disminuya la justicia entre servicios.

Sin embargo, la justicia conseguida –en general– con este esquema no es gratuita sino que tiene un precio. El beneficio global (subfiguras (d)) se ve reducido ya que esta política no da prioridad a los servicios conversacionales, que son los que tienen mayor eficacia en términos de beneficio. Su valor máximo es de 85, 65 y 43 unidades para cargas *media*, *alta* y *muy alta*, conseguidas con 16, 16 y 12 estaciones VoIP respectivamente.

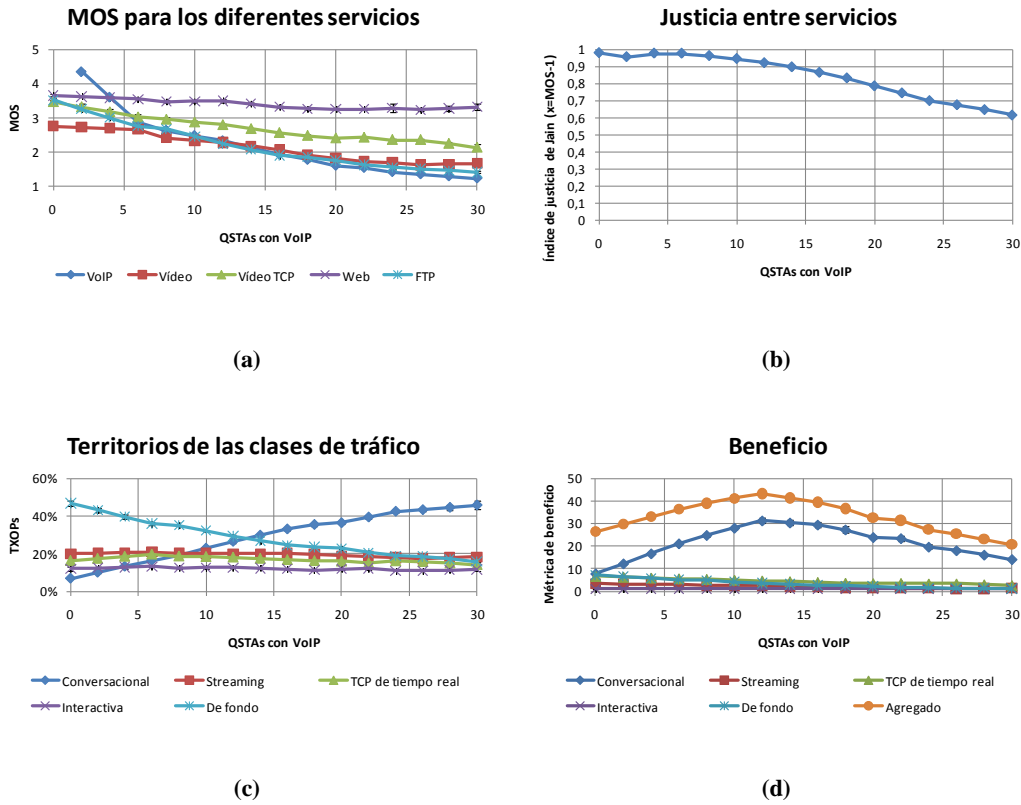


Figura 6.13. Resultados usando la política de reparto justo en términos de Calidad de Experiencia para carga muy alta.

Resultados de la política de reparto para aumento de beneficio

Los resultados obtenidos para la última política analizada se exponen en las siguientes figuras. Esta aproximación intenta aumentar los ingresos por parte del operador de la red.

Las variaciones en los territorios (subfiguras (c)) persiguen aumentar los ingresos por unidad de recurso, llevando a una priorización de aquellos servicios más eficientes en ingresos, i.e. primero VoIP, después vídeo sobre TCP, FTP, streaming de vídeo y finalmente navegación web. Los valores de ganancia son equivalentes a los obtenidos con

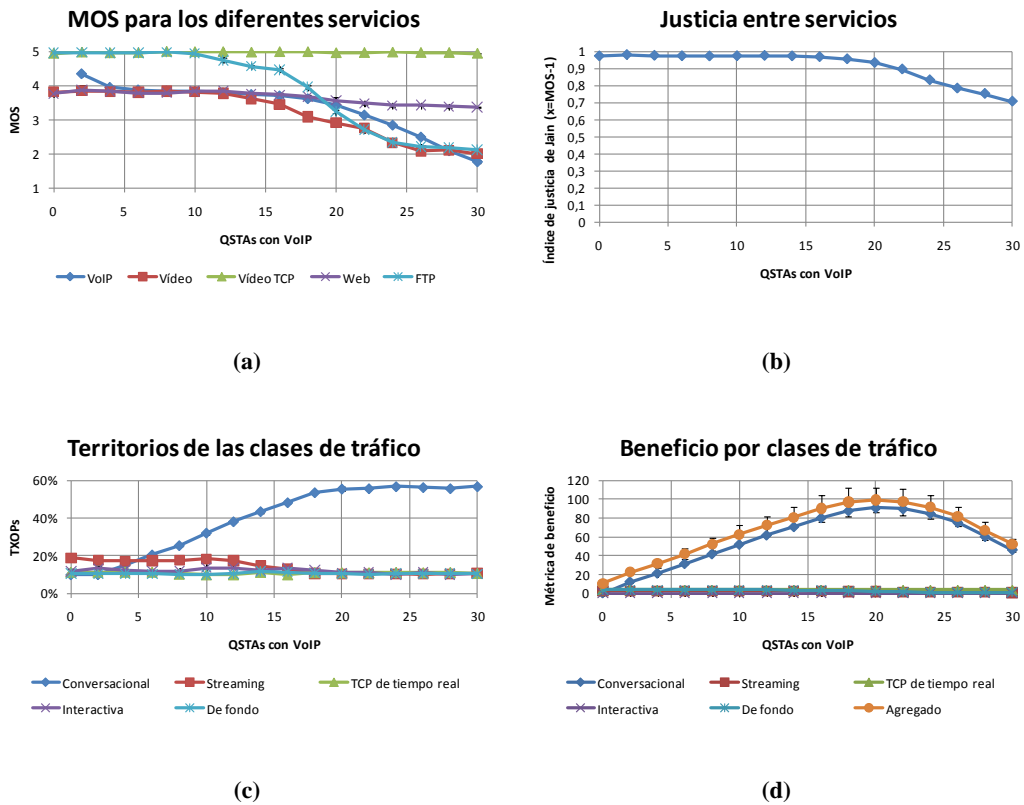


Figura 6.14. Resultados usando la política de *aumento de beneficio para carga media*.

la política *con derecho preferente* (subfiguras (d)). Esta semejanza se debe a que el servicio más prioritario coincide con el servicio que consigue mayores ganancias por unidad de recurso –VoIP– y que su beneficio es varios órdenes de magnitud mayor que el del resto de los servicios siendo un servicio dominante en el beneficio agregado. La introducción de otros servicios (con diferentes requisitos en cuanto a recursos) o el uso de diferentes factores de beneficio para cada CT conducirían a situaciones en las que esta política mejoraría el beneficio conseguido mientras que el esquema *con derecho preferente* no lo haría.

Debido a la política codiciosa empleada, se experimentan grandes variaciones en términos de QoE entre las diferentes clases de tráfico (subfiguras (a)), siguiendo la distribución de beneficio comentada. Así, por ejemplo, la calidad es muy inferior para los

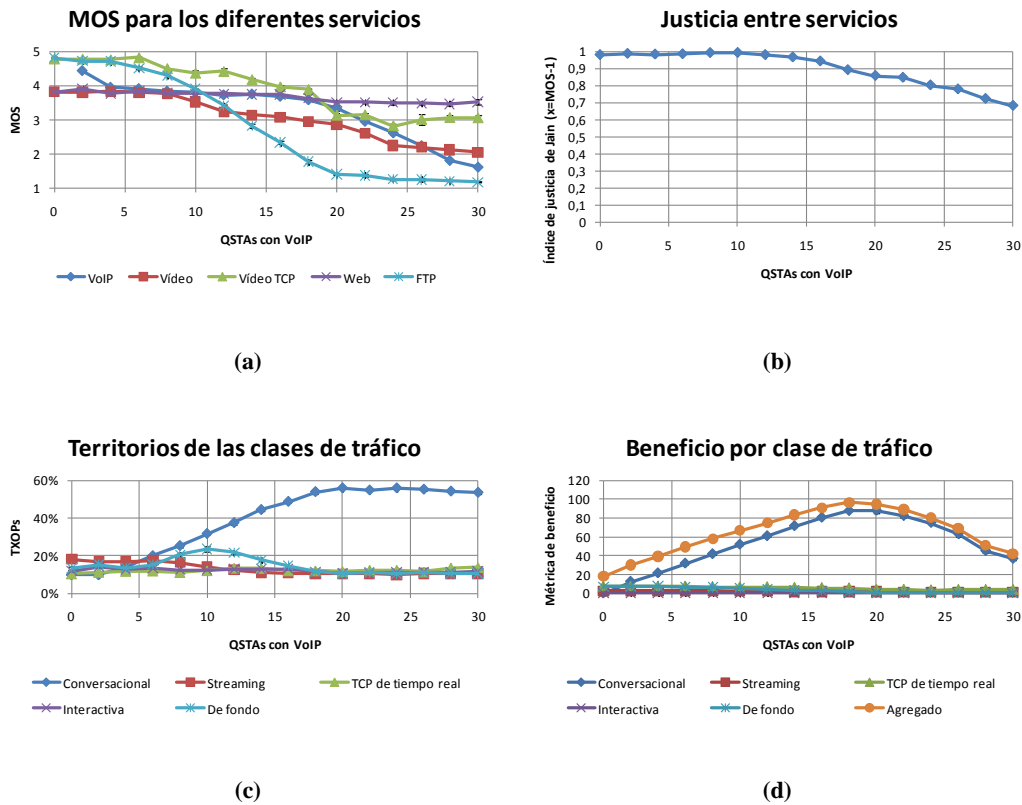


Figura 6.15. Resultados usando la política de *aumento de beneficio* para carga alta.

flujos de *streaming* de vídeo si se compara con VoIP.

Respecto a la justicia, debido a que se *prioriza* en función del beneficio, ésta es una política injusta entre los diferentes servicios. Los valores del índice de justicia son similares a los obtenidos con la política de reparto *con derecho preferente*, siendo sus mínimos 0.7, 0.7 y 0.6 para *cargas media, alta y muy alta* respectivamente.

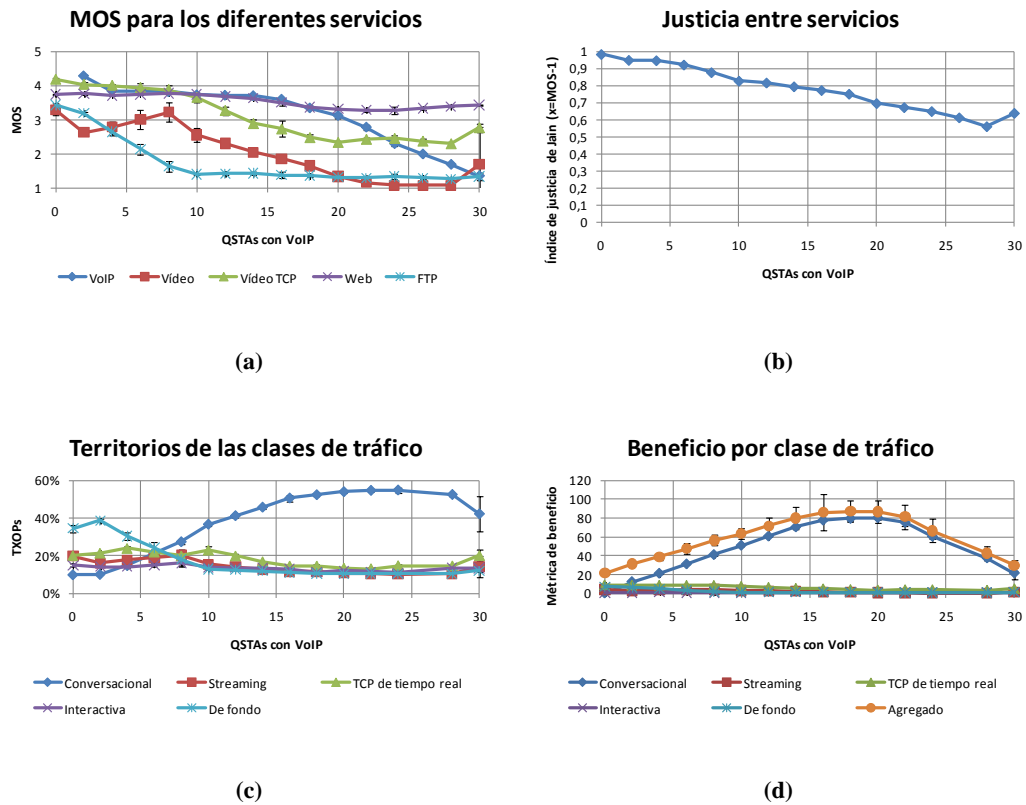


Figura 6.16. Resultados usando la política de *aumento de beneficio* para carga muy alta.

6.6 Comparativa entre las diferentes aproximaciones

Este apartado incluye una comparativa entre las diferentes aproximaciones planteadas para gestionar los recursos disponibles en el interfaz inalámbrico. Para ello, se presentan los principales resultados obtenidos para los diferentes escenarios simulados –con *carga media*, *carga alta* y *carga muy alta*– al utilizar:

- El planificador MSS (sin el Método de los Territorios)

6.6. Comparativa entre las diferentes aproximaciones | 135

- El planificador MSS conjuntamente con el Método de los Territorios y las políticas de:
 - Reparto con derecho preferente
 - Reparto sin derecho preferente
 - Reparto justo en términos de QoE
 - Aumento de beneficio

Las métricas mostradas, a modo de resumen de todos los resultados estudiados, son las siguientes:

- Número de usuarios VoIP satisfechos, i.e. con una calidad superior a la establecida como criterio en el Método de los Territorios (MOS superior a $UIM_i = 3.5$)
- Número de usuarios VoIP que permiten que todos los usuarios –de cualquier servicio– estén satisfechos
- Máximo beneficio obtenido (con la carga escogida y al variar el número de conversaciones de voz)
- Valor mínimo del índice de justicia de Jain (métrica $x = \overline{MOS}_i - 1$) para el rango de tráfico VoIP simulado (de 0 a 30 conversaciones de voz)

La Figura 6.17 incluye estos indicadores para los experimentos realizados con *carga media*. En este caso, dado que la carga no es excesivamente elevada, los resultados entre las diferentes aproximaciones no difieren demasiado. Así, el número de usuarios VoIP satisfechos que pueden estar a la vez en el sistema –con *carga media*– es de 16 para la disciplina MSS (sin utilizar el Método de los Territorios) y ligeramente superior al usar los territorios –con un máximo de 20 para la política de reparto con derecho preferente, esquema más favorable para los servicios *conversacionales*–.

En cuanto a la máxima carga soportada con todos los usuarios satisfechos, ésta se corresponde aproximadamente con la de 16 ~ 17 conversaciones de voz para todos los esquemas. Igualmente, el máximo beneficio es aproximadamente del mismo orden (entre 85 y 100 unidades).

La diferencia principal entre estas soluciones es cómo reacciona el sistema cuando está muy cargado. Así, en el punto de mayor carga simulado (30 usuarios VoIP) el índice

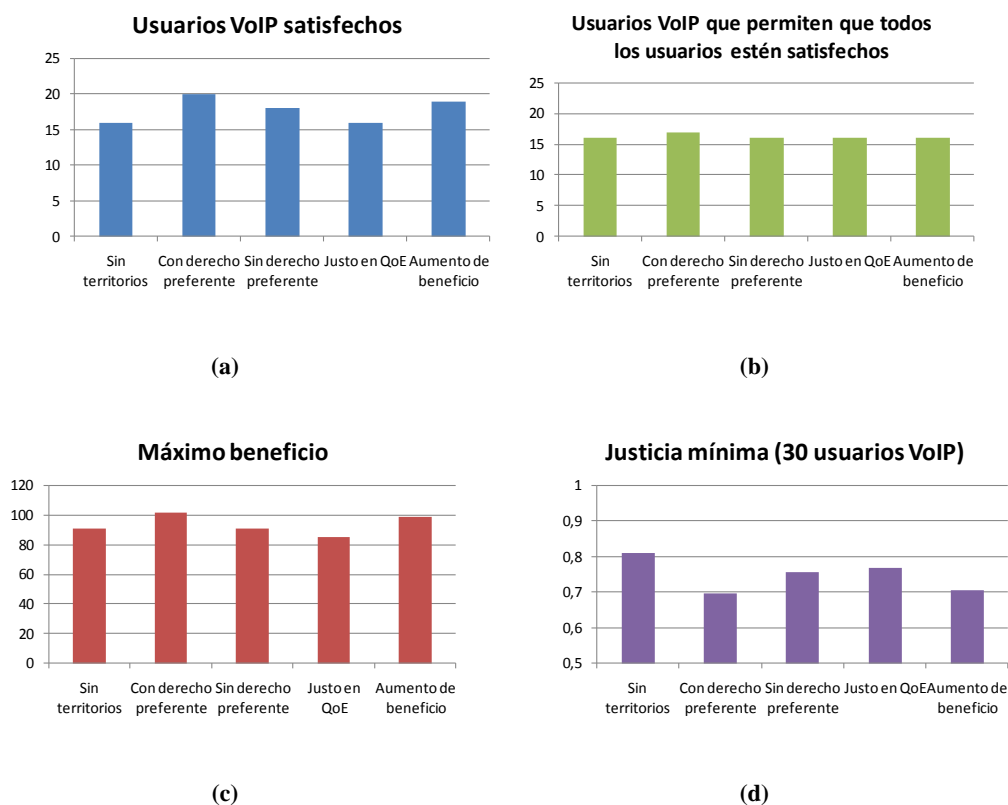


Figura 6.17. Comparativa entre las diferentes aproximaciones para la gestión de recursos en el caso de *carga media*.

de justicia revela qué aproximación es más igualitaria entre los diferentes servicios. Como se puede observar en la subfigura (d), el planificador MSS es el más justo entre servicios mientras que las políticas de reparto con derecho preferente y de aumento de beneficio son las que crean una mayor diferenciación entre servicios –debido a que se diseñaron con ese objetivo–.

La Figura 6.18 presenta la comparativa en el caso de *carga alta*. Al aumentar la carga, también aumentan las diferencias en los resultados de las diferentes soluciones. Así, el uso del planificador MSS sin la gestión basada en territorios hace que sólo se soporten 5 conversaciones con la calidad requerida, aumentando hasta 20 cuando se utiliza un reparto con derecho preferente. Como se comentó en la Sección 6.5.2, la

6.6. Comparativa entre las diferentes aproximaciones | 137

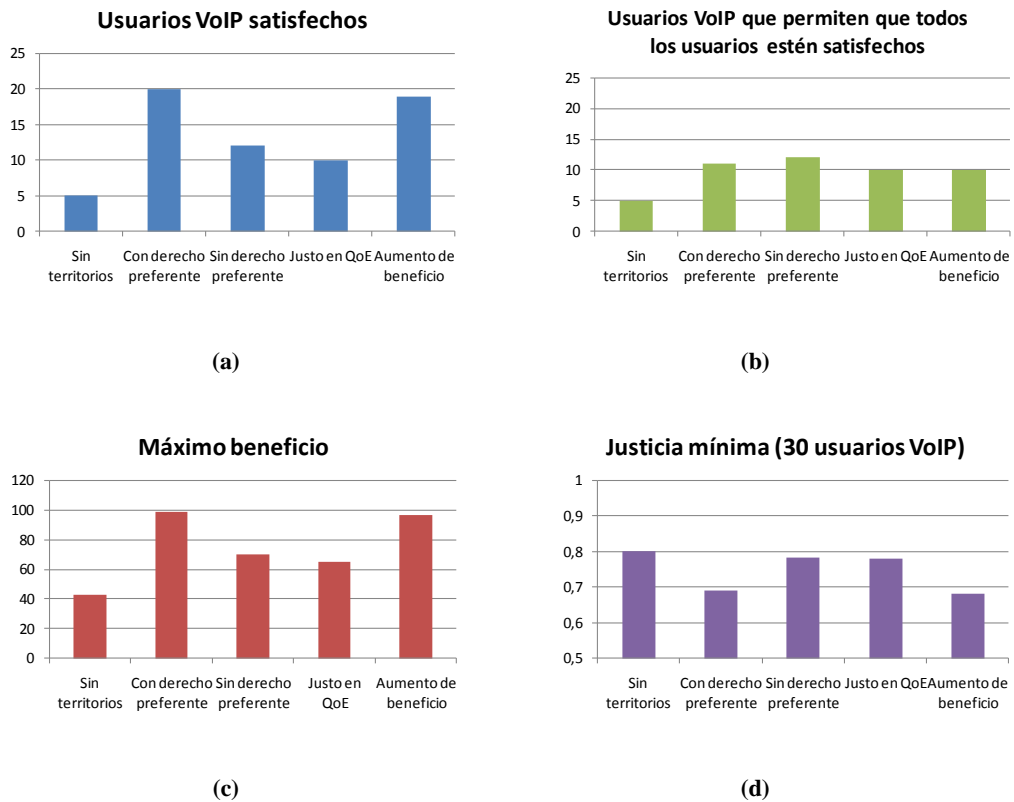


Figura 6.18. Comparativa entre las diferentes aproximaciones para la gestión de recursos en el caso de *carga alta*.

gestión que aumenta el beneficio también prioriza al servicio VoIP (el más eficiente en cuanto a ingresos), por lo que el resultado es similar. Las otras políticas consiguen entre 10 y 12 llamadas de voz con una QoE aceptable. Por tanto, sea cual fuere la política empleada, la capacidad para el servicio VoIP aumenta drásticamente al utilizar el Método de los Territorios.

Respecto a la máxima carga soportada –todos los usuarios de los diferentes servicios satisfechos–, ésta se corresponde a la de *carga alta* con 5 llamadas de voz para MSS y con entre 10 y 12 llamadas para el resto de esquemas. Nuevamente, el Método de los Territorios mejora el reparto de los recursos al introducir como criterio la QoE de los diferentes servicios y no asignar recursos innecesarios –que no mejoren la experiencia del

usuario–.

El máximo beneficio se consigue con los repartos con derecho preferente y aumento de beneficio –lógico por la definición de estas políticas–, con aproximadamente 100 unidades. Este valor se reduce a alrededor de 70 unidades para las otras políticas que utilizan el Método del Territorio. La propuesta que obtiene menos ganancias –40 unidades– es la del planificador MSS, por lo que el Método de los Territorios también incrementa los ingresos del administrador de la red respecto a no utilizarlo.

En términos de justicia, MSS es la mejor opción ya que no separa recursos para cada clase de tráfico –y además es una disciplina óptima respecto a la justicia entre los planificadores de su familia [54]–, consiguiéndose un índice de 0.8. Las políticas de reparto sin derecho preferente y justa en QoE obtienen un valor similar –ésta última diseñada con este objetivo–, aunque ligeramente inferior. Como cabía esperar, son las políticas de reparto con derecho preferente y de aumento de beneficio las más injustas, con un índice inferior a 0.7.

La Figura 6.19 sintetiza los resultados obtenidos para *carga muy alta*. Este escenario justificaba la propuesta del Método de los Territorios, ya que no se podían cursar llamadas de voz con la calidad requerida al usar sólo el planificador MSS (como se aprecia en la subfigura (a)). Sin embargo, el uso de los territorios y la priorización del tráfico VoIP permite hasta 19 llamadas con calidad suficiente al repartir con derecho preferente –su rendimiento es prácticamente independiente del resto de servicios con esta política–, 17 al mejorar el beneficio y 3 ~ 4 para las otras políticas.

En ninguno de los casos se consigue que todos los usuarios de los servicios considerados estén satisfechos, incluso sin conversaciones, debido a la alta carga de tráfico generada en este experimento.

En cuanto a las ganancias del administrador de la red, las distintas alternativas con *carga muy alta* aumentan las diferencias. Éstas se encuentran alrededor de 90 unidades para las disciplinas con derecho preferente y de aumento de ingresos –máximo entre las diferentes opciones–, reduciéndose a 53, 33 y 16 unidades para los esquemas sin derecho preferente, de justicia en QoE y sin utilizar territorios, respectivamente. Nuevamente, cualquier política de las definidas para el Método de los Territorios mejora las ganancias

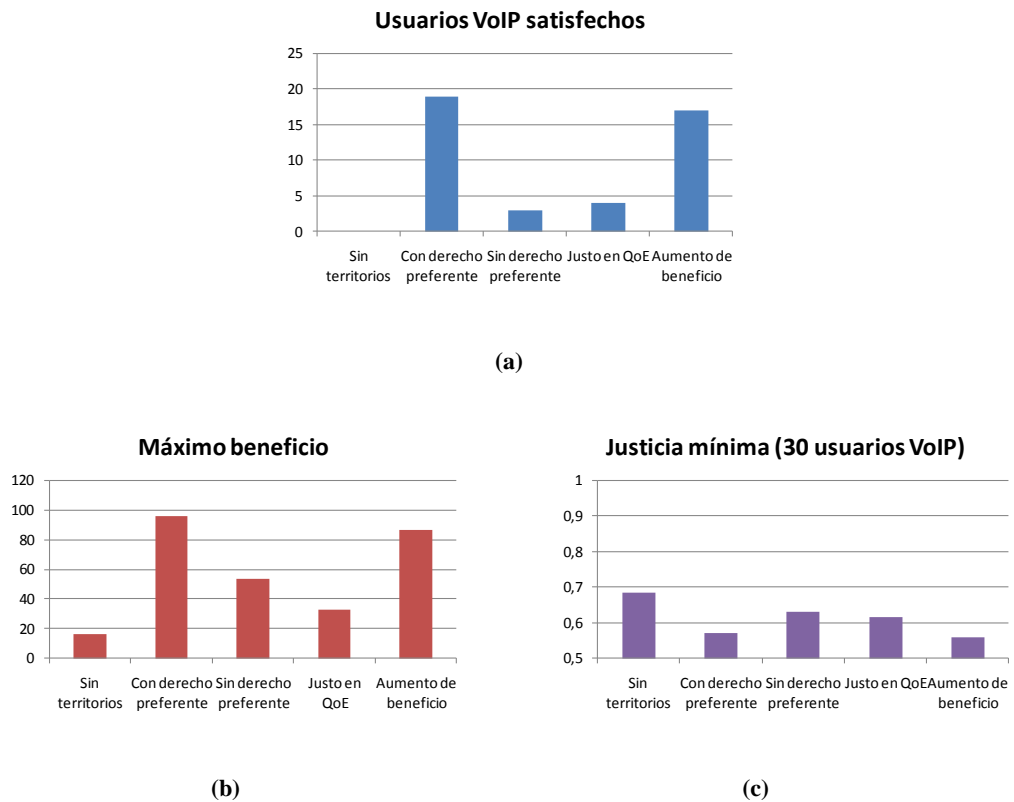


Figura 6.19. Comparativa entre las diferentes aproximaciones para la gestión de recursos en el caso de *carga muy alta*.

obtenidas respecto a no utilizar este esquema.

6.7 Conclusiones

En este capítulo se han presentado diferentes políticas para un arquitectura de gestión de recursos (el Método de los Territorios) que permitirían a un operador o administrador de red configurar cómo se comparte la capacidad inalámbrica en una red IEEE 802.11e que utilizase el mecanismo de acceso al medio HCCA.

140 | Políticas de gestión de recursos basadas en Calidad de Experiencia

Como se explicó en el apartado 5.1, el Método de los Territorios divide las oportunidades de transmisión en diferentes conjuntos. Estos conjuntos contienen los recursos para las diferentes clases de tráfico (*conversacional, streaming, TCP de tiempo real, interactiva y de fondo*), i.e. grupos de servicios con requisitos del mismo tipo. El tamaño de estos conjuntos o *territorios* está controlado por la Calidad de Experiencia del usuario final (en términos de MOS) y por la política elegida.

Se han propuesto cuatro políticas diferentes de reparto de recursos: *con derecho preferente, sin derecho preferente, justo en términos de Calidad de Experiencia y aumento de beneficio*. El análisis realizado muestra, en los escenarios simulados, las siguientes conclusiones. La política de *reparto con derecho preferente* permite la priorización de los servicios. La aproximación *sin derecho preferente* distribuye los recursos dependiendo de la evolución temporal de la demanda, sirviendo antes al primero que llega. Por tanto, la asignación de recursos no es conocida *a priori* y no sigue un criterio predefinido. Este no es el caso de la política *justa en términos de Calidad de Experiencia*, que permite al operador igualar la calidad subjetiva entre las distintas clases de tráfico. Por último, el algoritmo para *aumento de beneficio* incrementa las ganancias del operador de red (con factores que permiten ajustar los beneficios conseguidos por cada clase de tráfico). En cualquier caso, por encima de estas políticas se puede ajustar una capacidad dedicada para cada clase de tráfico garantizando un rendimiento mínimo para esos servicios.

Estas políticas se han comparado entre ellas y con el planificador MSS, en términos de calidad subjetiva, beneficio conseguido y justicia entre servicios. Esta comparativa muestra que las diferentes políticas propuestas para el Método de los Territorios mejoran el rendimiento respecto a utilizar únicamente el planificador MSS y consiguen los objetivos para los que fueron diseñadas.

Capítulo 7

TCP mejorado sobre HCCA

Una vez considerado el problema de la gestión de recursos en una red IEEE 802.11e con tráfico heterogéneo, en este capítulo se presentan las principales contribuciones relacionadas con el problema de la transmisión de datos sobre TCP.

En las propuestas realizadas en los capítulos anteriores, el envío de tráfico TCP se ha efectuado a través del mecanismo de acceso centralizado HCCA. Este mecanismo presenta varias ventajas sobre el acceso distribuido EDCA para la transferencia de este tipo de tráfico:

- 1) EDCA no permite garantizar ningún requisito de calidad de servicio. Aunque en general se considera que los servicios basados en TCP no requieren de este tipo exigencias, existen servicios donde estas garantías sí que son necesarias, como por ejemplo en la descarga progresiva de vídeo *on-line* –también denominada *streaming* de vídeo sobre TCP– e.g. YouTube [39].
- 2) La injusticia intrínseca entre flujos de dirección ascendente y descendente en redes IEEE 802.11 [61] sigue existiendo al utilizar EDCA [85] [86]

pero desaparece si se adopta HCCA. Esta injusticia se debe a que, en el acceso distribuido (DCF o EDCA), el punto de acceso es visto como una estación más a la hora de competir con el resto de estaciones. Al tener éste (en general) un mayor número de flujos –todos los de dirección descendente– que compiten a su vez entre sí, los turnos de transmisión para cada flujo descendente se reducen. Sin embargo, HCCA elimina esta injusticia dado que el planificador de paquetes –suponiendo que sigue una disciplina justa, como es el caso de MSS– no hace distinción entre flujos ascendentes y descendentes a la hora de asignar el siguiente turno de transmisión (TXOP).

- 3) El uso de un planificador de paquetes en HCCA aumenta la flexibilidad para desarrollar mecanismos y algoritmos más complejos, como la propuesta introducida en este capítulo.

De esta manera, el objetivo perseguido en las siguientes secciones será mejorar la eficacia del protocolo TCP –i.e. aumentar la tasa efectiva de transmisión– en redes inalámbricas con acceso HCCA.

7.1 Una variante de TCP partido para HCCA

En capítulos anteriores se han propuesto diferentes alternativas para permitir el uso compartido de un canal inalámbrico por parte de servicios con requisitos estrictos de tiempo y aplicaciones de tráfico elástico, como las basadas en el protocolo TCP.

Dado que a través del punto de acceso en una red inalámbrica puede haber múltiples conexiones TCP, se consideró que realizar algún tipo de coordinación entre los mecanismos de control de congestión de las diferentes conexiones podría mejorar el uso de los recursos en la red. De esta forma, aprovechando el conocimiento por parte del QAP del estado de las diferentes conexiones TCP, se podría mejorar su rendimiento en función de la calidad experimentada por los usuarios finales.

Sin embargo, esta aproximación presenta un problema: el control de congestión no

7.1. Una variante de TCP partido para HCCA | 143

depende exclusivamente del último salto (en este caso la red inalámbrica) sino de todo el camino entre el origen y el destino. Así, la dependencia con la parte cableada –diferente para cada conexión– y el desconocimiento sobre ésta, introducen una dificultad adicional en el problema abordado: la gestión coordinada de las conexiones TCP en HCCA.

A pesar de ello, en la literatura existen propuestas para independizar dos tramos con diferentes características en una conexión TCP. Una propuesta relevante es el *TCP partido* (*split-TCP*) [87], cuyo esquema utiliza un *proxy* de rendimiento mejorado (PEP, *Performance Enhancing Proxy*) en un punto intermedio entre el origen y el destino. De esta forma, la conexión TCP extremo a extremo inicial se divide en dos nuevas conexiones TCP, una entre el origen y el *proxy* y otra entre el *proxy* y el destino.

En el caso de un entorno heterogéneo, donde un QAP conecta una red cableada con una red inalámbrica, el punto intermedio surge de forma natural en el QAP –que actuaría como PEP–, dado que separa los entornos que *a priori* tienen características diferentes: el medio cableado –habitualmente con alta velocidad, baja probabilidad de pérdidas y alto retardo– y el medio inalámbrico –normalmente de una menor velocidad y mayor probabilidad de pérdidas (retransmisiones)–. De esta forma, se pasaría de un escenario como el mostrado en la Figura 7.1 a otro escenario como el expuesto en la Figura 7.2.

Es de esperar que, con esta nueva aproximación, cada conexión individual se pueda adaptar a las características del entorno inalámbrico sin estar influenciada por las particularidades del resto del camino (parte cableada).

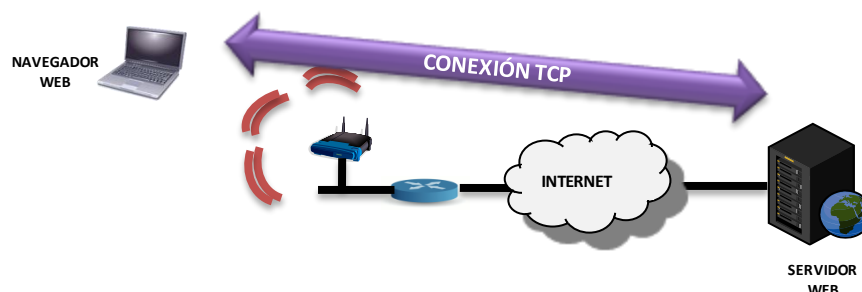


Figura 7.1. Escenario con una conexión TCP convencional.

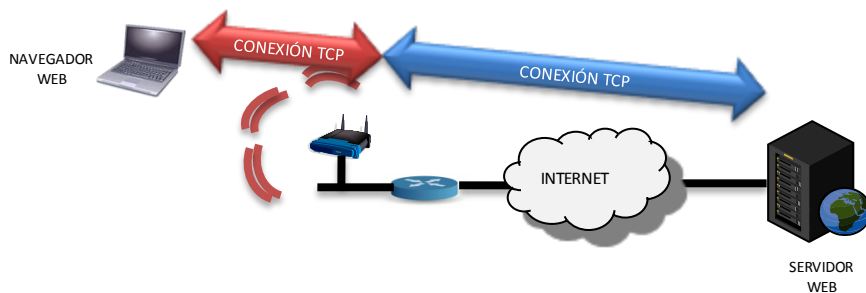


Figura 7.2. Escenario con una conexión TCP partida.

Una vez que se dispone de un conjunto de conexiones TCP que se extienden entre las diferentes estaciones a través del punto de acceso, se podría diseñar un mecanismo para el interfaz inalámbrico de gestión coordinada entre estas conexiones.

No obstante, al estudiar las diferentes alternativas para diseñar este mecanismo en un entorno IEEE 802.11e con acceso HCCA surge la siguiente pregunta: ¿es necesario utilizar el protocolo TCP en el interfaz inalámbrico? Esta pregunta, sencilla pero trascendental en la aproximación a seguir para el diseño a realizar, se puede contestar considerando las funcionalidades del protocolo TCP. Éstas se resumen a continuación:

- *Multiplexación.* Este mecanismo permite identificar los puntos finales (origen y destino) de la información; así, por medio de los puertos, un mismo equipo puede recibir mensajes enviados a diferentes aplicaciones.
- *Control de conexión.* TCP ofrece un servicio orientado a conexión. Por tanto, antes de enviar datos, se realiza un establecimiento de conexión entre el origen y el destino que permite sincronizar (establecer un estado común) entre ambas entidades, reservando los correspondientes recursos. De forma similar, al terminar la transmisión, para liberar recursos, se procede al cierre de la conexión.
- *Control de errores.* TCP adopta un procedimiento ARQ (*Automatic Repeat reQuest*) con confirmaciones positivas y acumulativas. Para ello de la cabecera de los segmentos TCP se usan los campos *comprobación*,

7.1. Una variante de TCP partido para HCCA | 145

secuencia, acuse y al bit A (Acknowledgment).

- *Control de flujo.* El transmisor es capaz de adaptar su velocidad para evitar que el receptor se sature, al recibir más datos de los que la aplicación puede procesar. Básicamente, se trata de un procedimiento de realimentación mediante el cual, el receptor anuncia –en cada segmento de datos enviado– la capacidad disponible (en bytes) en su *buffer* mediante el campo de *ventana ofertada*.
- *Control de congestión.* TCP es capaz de modificar el tamaño máximo de la ventana de transmisión, y en consecuencia su velocidad de transmisión, para adaptarse así a la dinámica de los recursos disponibles en la red (ancho de banda y *buffer*). El objetivo es evitar que los *routers* intermedios se congestionen (sinónimo de agotar sus recursos disponibles), ya que en ese caso se descartarían paquetes al dejar exhaustas sus colas.

A pesar de lo anterior, en un entorno de red IEEE 802.11e con mecanismo de acceso HCCA estas funcionalidades pueden ser proporcionadas sin tener que usar el protocolo TCP:

- *Multiplexación.* Esta funcionalidad sigue siendo necesaria, pero no requiere de un protocolo tan complejo como TCP. Por ejemplo, el protocolo UDP –mucho más sencillo– también incluye esta funcionalidad (mediante la identificación de los puertos origen y destino).
- *Control de conexión.* En las redes WLAN de infraestructura se realiza –en el nivel MAC– un proceso de negociación de las características del flujo de datos, siendo un procedimiento equivalente al establecimiento de conexión.
- *Control de errores.* La capa MAC de IEEE 802.11 realiza un control de errores basado en las retransmisiones de las tramas no confirmadas, utilizándose confirmaciones positivas –como en TCP–. De esta forma, este mecanismo no es necesario en el nivel de transporte. Aun en el caso de una probabilidad de error residual tras estas retransmisiones, el planificador del mecanismo HCCA puede incluir la retransmisión de las tramas no enviadas correctamente. Por ello, no son necesarios los campos involucrados ni los

mensajes de confirmación. De hecho, la eliminación de los mensajes de confirmación de una conexión TCP partida en una red WLAN ya fue propuesta en [88].

- *Control de flujo.* Si bien el control de flujo sigue siendo necesario, sólo requiere la inclusión de un campo de *ventana ofertada* en la cabecera.
- *Control de congestión.* Como se comentó, esta función permite adaptar la velocidad de transmisión al medio utilizado y evitar el agotamiento de los recursos en los nodos intermedios. En el caso de una red IEEE 802.11e con HCCA, no existen nodos intermedios entre las estaciones finales y el QAP. Además, la velocidad de transmisión viene gobernada por el planificador de paquetes por lo que no es necesaria una adaptación a la capacidad del canal. Por ello, este mecanismo –el más complejo de TCP– no es necesario en el salto inalámbrico de este tipo de redes.

Por estos motivos, sería suficiente utilizar un protocolo de transporte más sencillo, que incluyera los puertos utilizados y la ventana ofertada por el receptor. Así, se podría utilizar un protocolo similar a UDP que modificase su cabecera para incluir un campo de ventana ofertada. Además de las mejores prestaciones que cabe esperar al desaparecer el mecanismo de control de congestión, también es destacable la potencial mejora que supone la eliminación de los mensajes de confirmación en el sentido opuesto a la transmisión de datos.

Esta versión modificada de TCP partido, adaptada para su adopción en redes IEEE 802.11e con HCCA, ha sido denominada *TCP partido & UDP* (véase la Figura 7.3) aunque, como se ha comentado, utiliza una cabecera modificada respecto al protocolo UDP original (para incluir la ventana ofertada).

7.2 Evaluación de rendimiento

7.2.1 Condiciones experimentales

La topología de la red y el entorno de simulación, así como los parámetros generales utilizados son los descritos en la Sección 4.3.1

Para evaluar la mejora en el rendimiento debida al uso de una conexión TCP partida y la eliminación de TCP en la red inalámbrica, se han simulado algunos escenarios significativos. Estos escenarios han sido analizados comparando el protocolo TCP sin modificaciones (*TCP normal*), con la versión original de *TCP partido* y con la solución propuesta (*TCP partido & UDP*).

Estas variantes –*TCP partido* y *TCP partido & UDP*– han sido implementadas en NS-2, basadas (aunque con importantes modificaciones) en la implementación de un *proxy* de rendimiento mejorado para comunicaciones por satélite (*SatPEP* [89], cuya implementación está descrita en [90]).

Para descartar el impacto de otros efectos debidos al patrón de generación de tráfico, en el análisis realizado se ha considerado tráfico FTP. Además en algunos casos, para

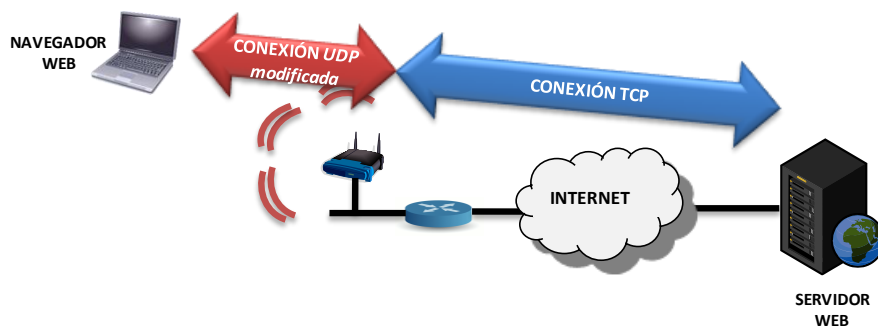


Figura 7.3. Escenario con una conexión TCP partido & UDP.

crear carga adicional (tráfico de *background*), por sencillez se ha utilizado tráfico UDP, ya que de esta forma se evitan todos los procedimientos presentes en una conexión TCP.

Al igual que en los capítulos anteriores, se ha utilizado el módulo *TCPLinux* [72] para reflejar el comportamiento real de TCP en las simulaciones, siendo *New Reno* [73] la variante elegida. Además, a diferencia de los capítulos anteriores, el rendimiento se estimará usando la velocidad de transmisión ya que, debido a la saturación del MOS en la función de utilidad de FTP (véase la Sección 3.2.5), hay muchos casos con la nota máxima (calidad *excelente*) y por tanto no se podrían comparar sus diferencias en términos de QoE.

7.2.2 Resultados

El primer experimento considerado tiene como objeto cuantificar la mejora que experimentaría un único usuario en el sistema. Debido a que la principal diferencia entre la variante propuesta y las existentes es la eliminación de TCP en el interfaz inalámbrico, se han modificado los principales parámetros que pueden afectar al comportamiento de su mecanismo de control de congestión, i.e. el tamaño de fichero y el tiempo de ida y vuelta (RTT, *Round-Trip Time*) en la parte cableada de la red.

Concretamente, se han realizado campañas de simulación para longitudes de fichero de 50KB, 500KB, 5MB e infinito (es decir, siempre hay datos a transmitir durante toda la simulación), pretendiendo con ello cubrir el rango habitual de datos a transmitir en una sesión.

De igual manera, se ha considerado que el enlace cableado (entre el servidor de FTP y el punto de acceso) tiene retardos *de ida* de 2, 50, 100 y 500 milisegundos, es decir, desde un enlace con un RTT casi despreciable (comparado con los valores típicos en Internet) hasta una red con un RTT elevado.

Los resultados obtenidos (en términos de *bit-rate* medio experimentado extremo a extremo) se muestran en las Figuras 7.4, 7.5, 7.6 y 7.7. En éstas, además de los resultados de simulación, se han añadido las velocidades que obtendría un único usuario utilizando:

- TCP normal en la red cableada –entre el servidor FTP y el usuario ocupando el lugar del punto de acceso– (etiquetado en las figuras como *TCP en cable*). Nótese que en las Figuras 7.4 y 7.8 esta curva no aparece debido a la escala utilizada –sus valores son muy superiores a los del resto de gráficas–.
- TCP normal en la red inalámbrica –suponiendo al servidor FTP en el QAP y estando el usuario en una QSTA– (etiquetado como *TCP en aire*)
- UDP en la red inalámbrica –caso similar al anterior, pero descargando el fichero a través del protocolo UDP– (etiquetado como *UDP en aire*)

El objetivo de añadir estos casos es establecer el límite máximo de la velocidad de transmisión en el caso de una conexión TCP partida –tanto si utiliza como si no UDP en el interfaz inalámbrico–, ya que su velocidad será inferior al mínimo de las velocidades en cada tramo [91]. Si bien la literatura sobre modelado y estimación del rendimiento de una conexión TCP es abundante (e.g. [92] [93] [94] [95] [96]), la evaluación de estas cotas superiores (*TCP en cable*, *TCP en aire* y *UDP en aire*) se ha realizado mediante simulación. Así, dado que la implementación de la variante de TCP utilizada (así como los valores de sus parámetros) para estimar estas cotas coincide con la usada en la modificación propuesta, sus resultados son directamente comparables.

En estas figuras se observa que, cuando el límite impuesto por *TCP en cable* (debido fundamentalmente al control de congestión de TCP) es superior al rendimiento máximo que puede ofrecer el medio inalámbrico (*TCP en aire* limitaría a la variante *TCP partido*; *UDP en aire* limitaría a *TCP partido & UDP*), la solución propuesta supera ampliamente tanto a *TCP normal* como a la versión original de *TCP partido*. En estos casos (e.g. cualquier tamaño de fichero para un retardo en el cable de 2 ms, o para tamaños grandes de ficheros en el resto de los casos) la ganancia (G) es aproximadamente:

$$G = \frac{7000 \text{ kbps} - 3700 \text{ kbps}}{3700 \text{ kbps}} \approx 90\%$$

Sin embargo, cuando el enlace cableado es el que restringe la velocidad de transmisión (debido al control de congestión de TCP, e.g. para ficheros de pequeño tamaño y retardos superiores a 2 ms), las opciones evaluadas consiguen una tasa equivalente. Esto es lógico ya que la parte inalámbrica se limita ahora a retransmitir lo

150 | TCP mejorado sobre HCCA

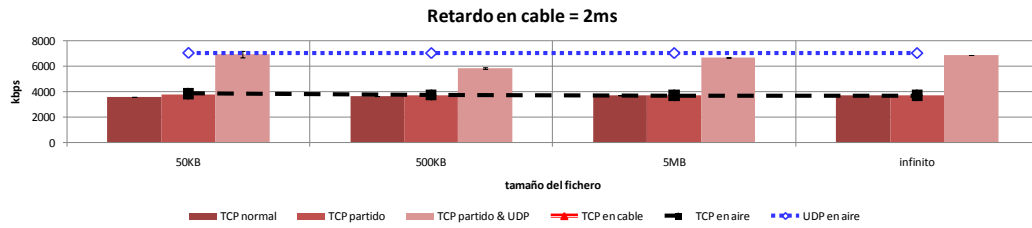


Figura 7.4. Resultados para 1 descarga FTP y un retardo en la red cableada de 2 ms.

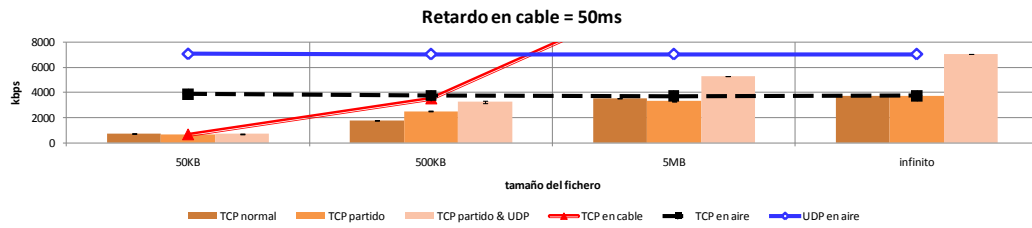


Figura 7.5. Resultados para 1 descarga FTP y un retardo en la red cableada de 50 ms.

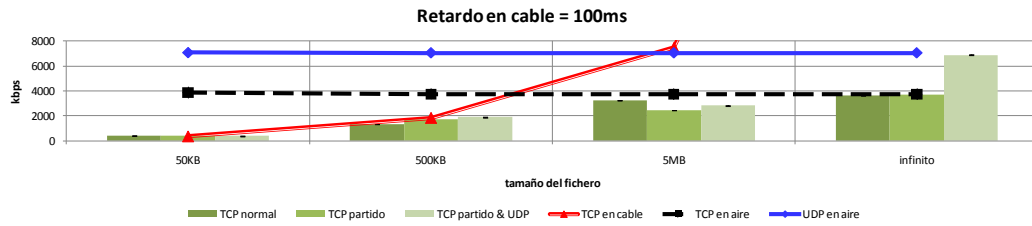


Figura 7.6. Resultados para 1 descarga FTP y un retardo en la red cableada de 100 ms.

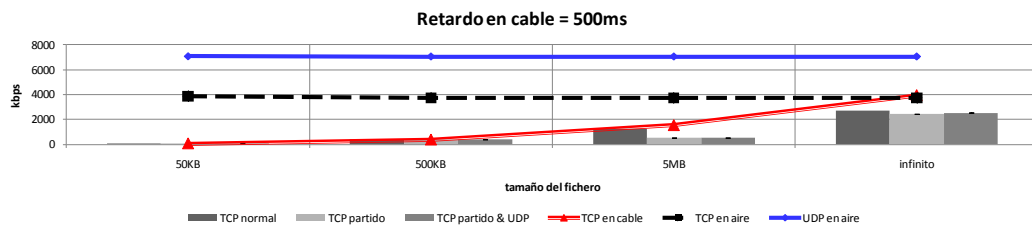


Figura 7.7. Resultados para 1 descarga FTP y un retardo en la red cableada de 500 ms.

que recibe –a menor velocidad– por el cable.

En cualquier caso, este experimento representa el caso peor respecto a la limitación debida a la parte cableada, ya que en escenarios con más de una QSTA –transmitiendo y/o recibiendo hacia/hasta un equipo en la red cableada– aumenta la probabilidad de que sea el interfaz inalámbrico el que restrinja la tasa binaria al compartirse entre más usuarios.

El siguiente conjunto de simulaciones considera que hay más carga en la red inalámbrica que la recibida por el usuario FTP. Por ello, se ha añadido tráfico (3Mbps) generado mediante UDP –para no incluir el efecto de una posible interacción entre fuentes TCP–.

El análisis de estos resultados (Figuras 7.8, 7.9, 7.10 y 7.11) confirma los razonamientos expuestos para el caso anterior. Cuando la parte de red cableada no sufre congestión, el rendimiento viene impuesto por el medio inalámbrico. Y, debido a la sustitución de TCP en este interfaz, la modificación propuesta sigue mejorando notablemente la velocidad de transmisión, en este caso G es aproximadamente:

$$G = \frac{3250 \text{ kbps} - 1700 \text{ kbps}}{1700 \text{ kbps}} \approx 90\%$$

Una vez evaluado el rendimiento cuando hay una única fuente de tráfico TCP –tanto en solitario como con tráfico adicional–, resta por comprobar el funcionamiento de *TCP partido & UDP* con múltiples fuentes de tráfico. Para ello, se ha simulado con 2, 4, 6, 8 y 10 descargas simultáneas –que perduran durante toda la simulación– utilizando diferentes retardos en el enlace por cable –2, 50, 100 y 500 ms– (véanse las Figuras 7.12, 7.13, 7.14 y 7.15). En estos casos y al igual que hasta ahora, para simplificar el análisis se supone que no hay errores de transmisión.

La compartición del canal inalámbrico entre varios usuarios reduce la posibilidad de que la parte con soporte físico –cable– limite el comportamiento del control de congestión de TCP, por lo que en este caso *TCP partido & UDP* siempre introduce una mejora en el rendimiento (incluso con ficheros pequeños y RTT elevados). La Figura 7.16 muestra la ganancia relativa, en términos de velocidad de transmisión, entre la propuesta y *TCP*

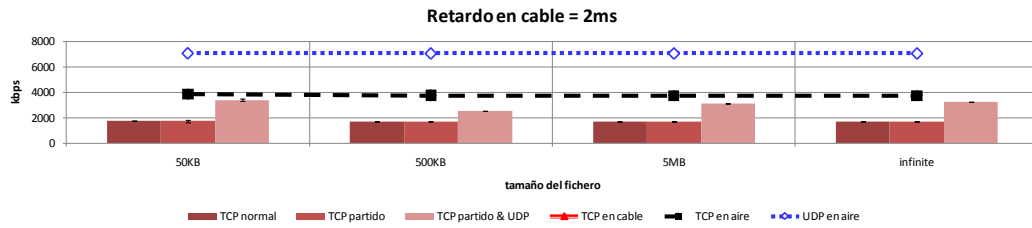


Figura 7.8. Resultados para 1 descarga FTP, 3 Mbps de tráfico UDP y un retardo (cable) de 2 ms.

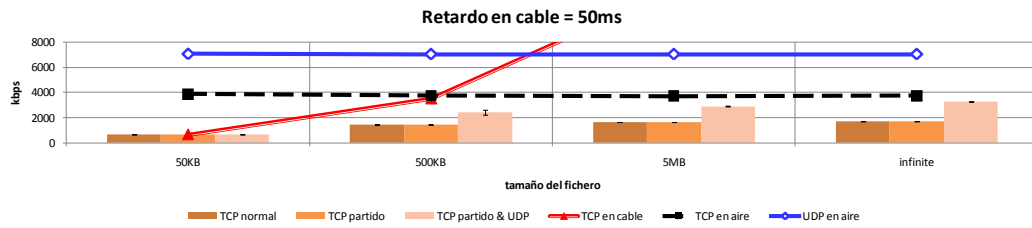


Figura 7.9. Resultados para 1 descarga FTP, 3 Mbps de tráfico UDP y un retardo (cable) de 50 ms.

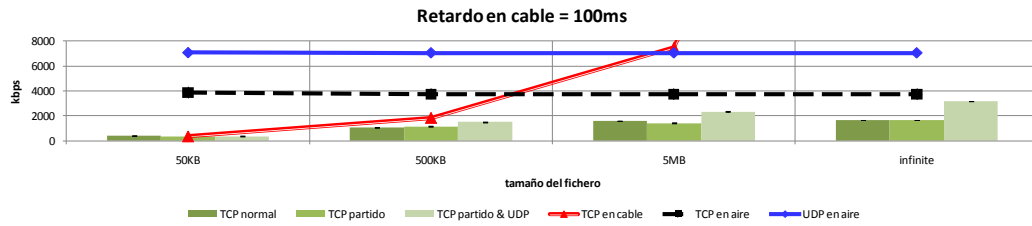


Figura 7.10. Resultados para 1 descarga FTP, 3 Mbps de tráfico UDP y un retardo (cable) de 100 ms.

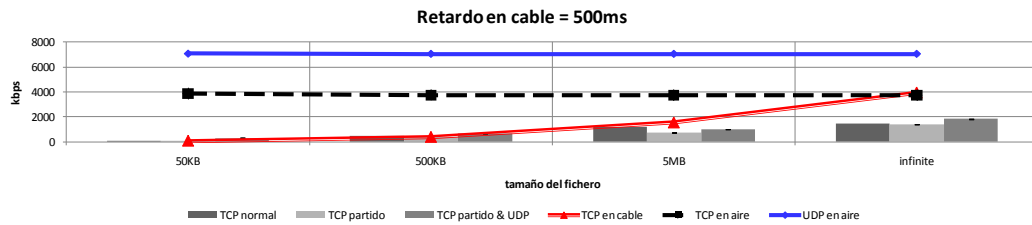


Figura 7.11. Resultados para 1 descarga FTP, 3 Mbps de tráfico UDP y un retardo (cable) de 500 ms.

normal. En esta figura se aprecia que la mejora es aproximadamente del 65% –con pocos usuarios–, salvo cuando el RTT es muy elevado (retardo de ida de 500 ms, i.e. $RTT = 1s$) en el que la ganancia está alrededor del 40%.

Cuando hay muchos usuarios (8 ó 10) se aprecia una reducción en la ganancia (véase la Figura 7.16), causada por el aumento del tráfico cursado agregado en el caso de *TCP normal* (aumenta de aproximadamente 4200 kbps con 2, 4 y 6 usuarios a unos 4500 kbps cuando hay 8 QSTAs, y alrededor de 5250 kbps cuando hay 10). Este crecimiento se debe a que los turnos no utilizados por un flujo TCP (e.g. cuando el control de congestión reduce la velocidad de transmisión al producirse desbordamientos en el *buffer* asociado a esa conexión en el QAP) son aprovechados por otro flujo.

Nótese que la disciplina de planificación utilizada en HCCA –MSS–, de forma similar al planificador de referencia [64], incluye en el QAP una cola por flujo de tráfico. Así, el rendimiento del control de congestión debido al almacenamiento de paquetes en el QAP es independiente entre conexiones (salvo por la disminución de ancho de banda disponible por flujo). Por ello, la reducción de tráfico cursado debida a la sincronización del control de congestión de múltiples flujos que comparten un *buffer* común [97] no ocurre en este escenario.

De esta manera, debido al aumento del tráfico cursado por *TCP normal* cuando hay muchos usuarios –en los escenarios simulados–, aunque UDP utiliza toda la capacidad del medio inalámbrico –el tráfico cursado agregado es siempre similar, aproximadamente 7000 kbps– la ganancia decrece en términos relativos.

Por otro lado, el descenso en la ganancia cuando el RTT es elevado ($2 \times 500ms$) se debe a la limitación introducida por el enlace cableado en el control de congestión –tal como ocurría con una fuente FTP, véase la Figura 7.7–.

Para completar el análisis, en el siguiente experimento se consideran diferentes probabilidades de error de trama en el interfaz inalámbrico. Dado que el objetivo es analizar las mejoras debidas al esquema propuesto y no realizar un estudio de las características de propagación en este tipo de entornos, se ha utilizado un modelo sencillo en el que la probabilidad de error se distribuye uniformemente.

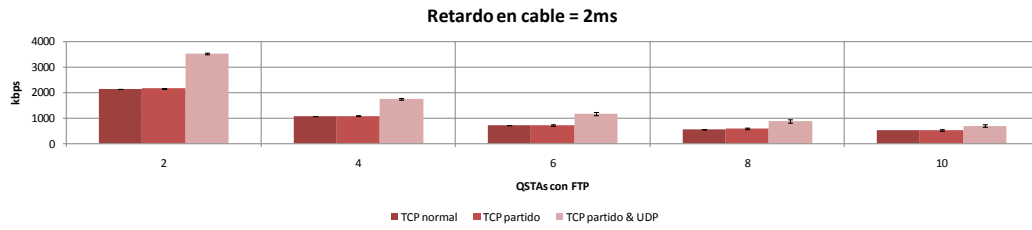


Figura 7.12. Resultados para varios usuarios FTP, un retardo (cable) de 2 ms.

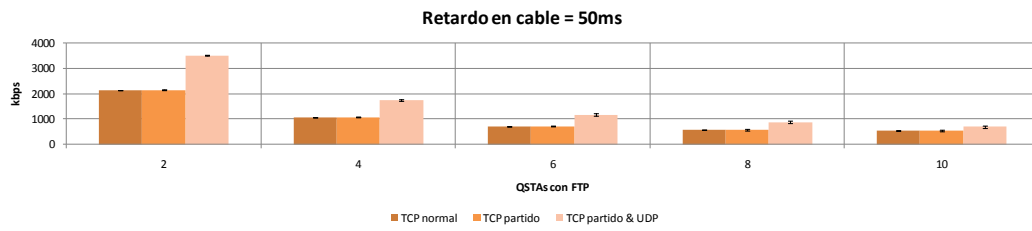


Figura 7.13. Resultados para varios usuarios FTP, un retardo (cable) de 50 ms.

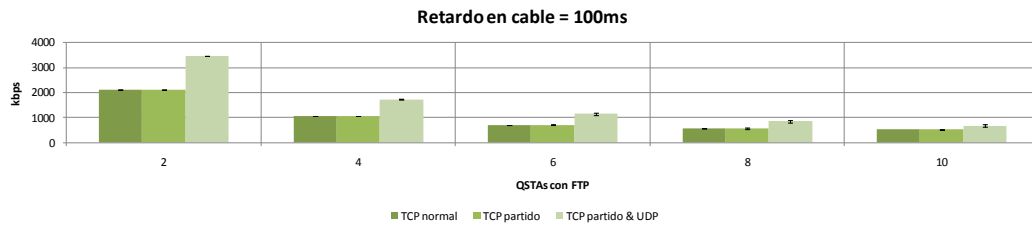


Figura 7.14. Resultados para varios usuarios FTP, un retardo (cable) de 100 ms.

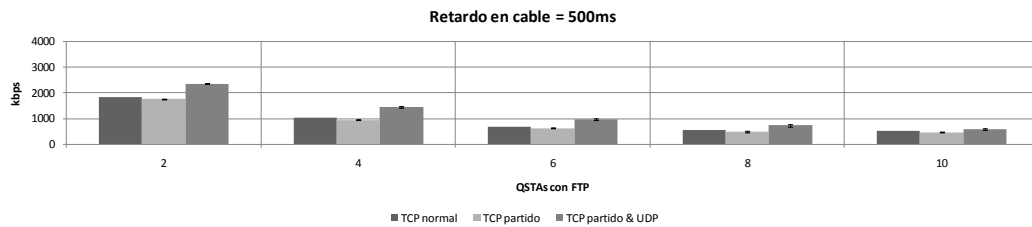


Figura 7.15. Resultados para varios usuarios FTP, un retardo (cable) de 500 ms.

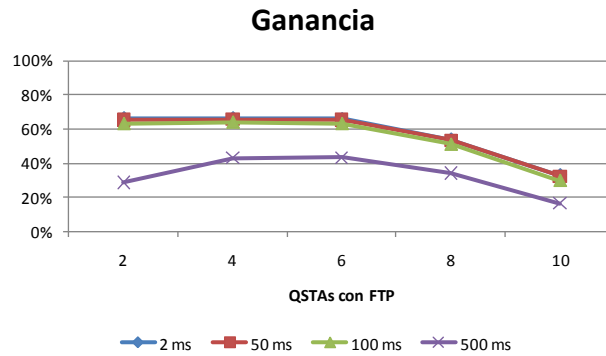


Figura 7.16. Ganancia para varios usuarios FTP.

Así, las siguientes figuras (Figuras 7.17 a 7.28) muestran los resultados obtenidos con la configuración anterior y diferentes probabilidades de error de trama: 10, 20 y 30% (valores razonables en entornos de interior [98]).

Nótese que, gracias a las retransmisiones en la capa MAC, la probabilidad real de error será muy inferior. Estas retransmisiones están gobernadas por el parámetro *Short Retry Limit* –cuyo valor es 7 por defecto–, que indica cuántas veces se puede retransmitir una trama con tamaño inferior al parámetro *RTSThreshold* –2432 bytes por defecto–. Sin embargo, aunque la probabilidad de error sea muy baja, las retransmisiones provocan que la tasa de transmisión efectiva se reduzca [99].

Como se puede observar, los resultados son similares a los anteriores salvo que, al aumentar la probabilidad de error, disminuye ligeramente el ancho de banda disponible.

Además, gracias a la independencia alcanzada –al partir la conexión TCP– entre los errores en el interfaz inalámbrico y el control de congestión de TCP en la parte cableada, también aumenta la ganancia entre utilizar *TCP normal* y *TCP partido & UDP*. Así, la Figura 7.16 mostraba una ganancia –en el caso sin pérdidas, con pocos usuarios y con RTT normales– $G \approx 65\%$ mientras que ésta crece al aumentar la probabilidad de error, llegando a ser alrededor de 80% cuando la probabilidad de pérdida de tramas es del 30% (véanse las Figuras 7.29, 7.30 y 7.31).

156 | TCP mejorado sobre HCCA

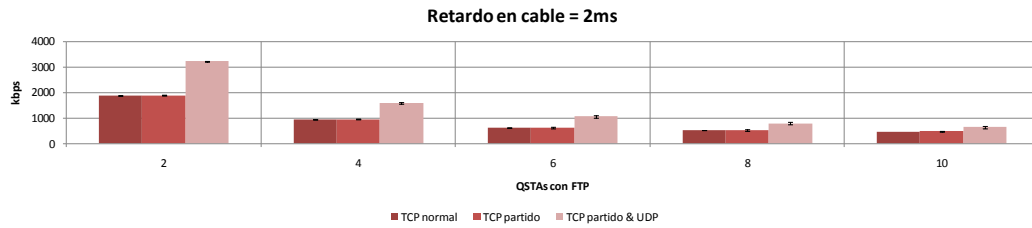


Figura 7.17. Resultados para varios usuarios FTP, un retardo (cable) de 2 ms y 10% de errores.

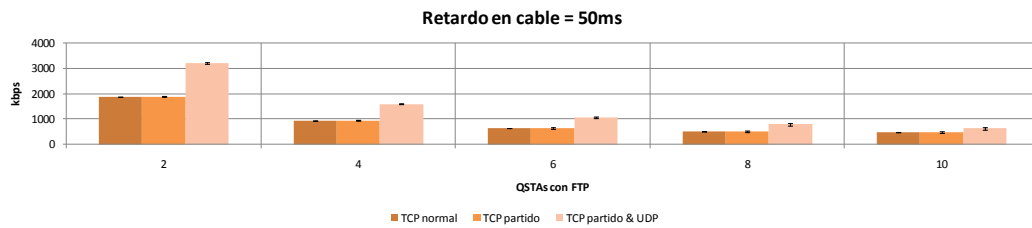


Figura 7.18. Resultados para varios usuarios FTP, un retardo (cable) de 50 ms y 10% de errores.

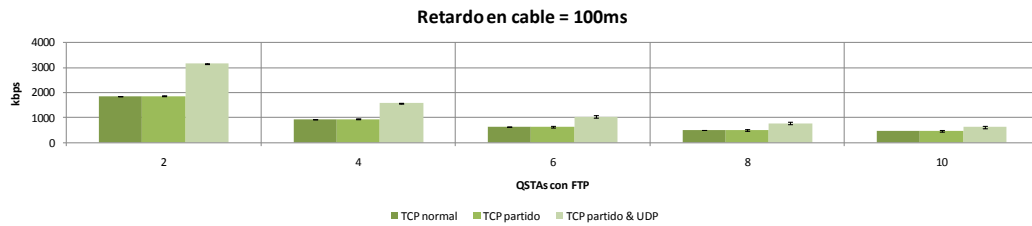


Figura 7.19. Resultados para varios usuarios FTP, un retardo (cable) de 100 ms y 10% de errores.

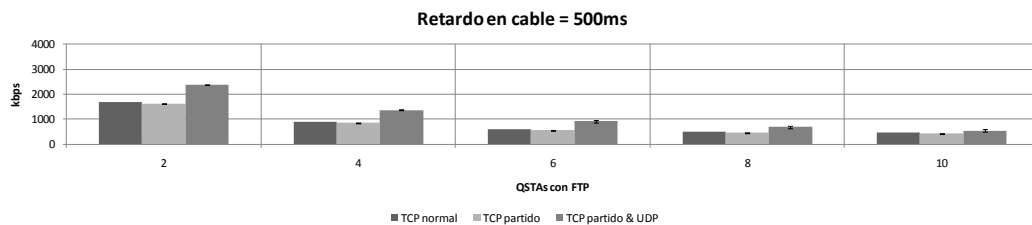


Figura 7.20. Resultados para varios usuarios FTP, un retardo (cable) de 500 ms y 10% de errores.

7.2. Evaluación de rendimiento | 157

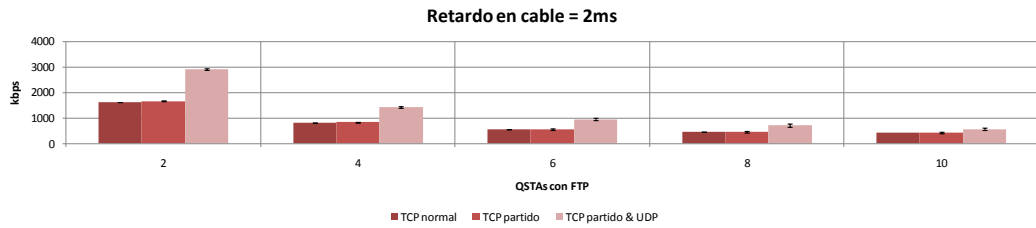


Figura 7.21. Resultados para varios usuarios FTP, un retardo (cable) de 500 ms y 20% de errores.

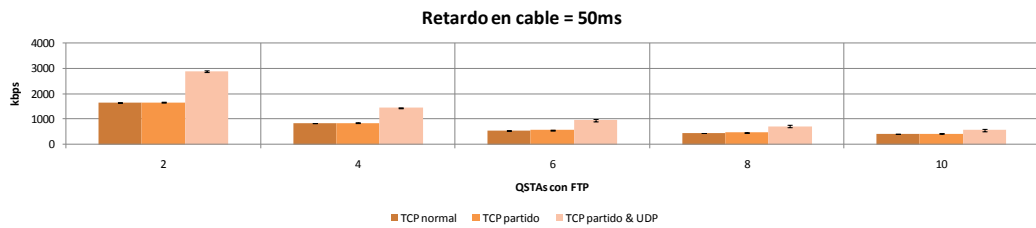


Figura 7.22. Resultados para varios usuarios FTP, un retardo (cable) de 500 ms y 20% de errores.

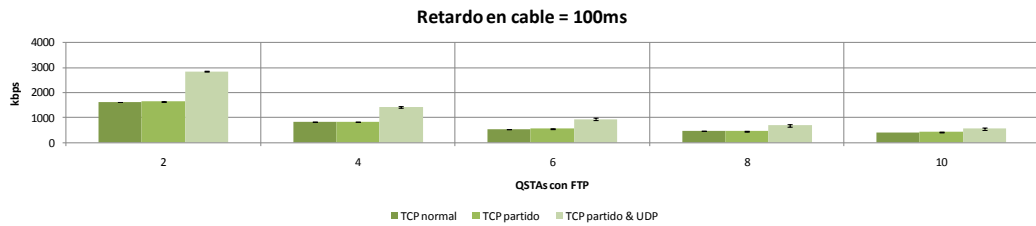


Figura 7.23. Resultados para varios usuarios FTP, un retardo (cable) de 500 ms y 20% de errores.

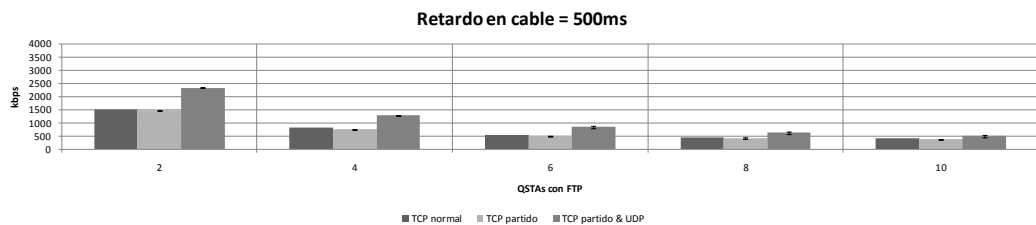


Figura 7.24. Resultados para varios usuarios FTP, un retardo (cable) de 500 ms y 20% de errores.

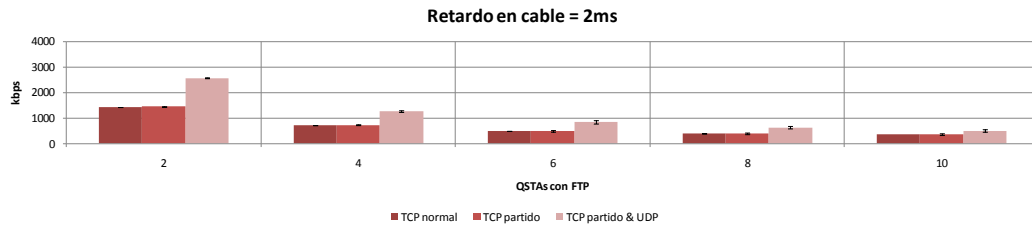


Figura 7.25. Resultados para varios usuarios FTP, un retardo (cable) de 2 ms y 30% de errores.

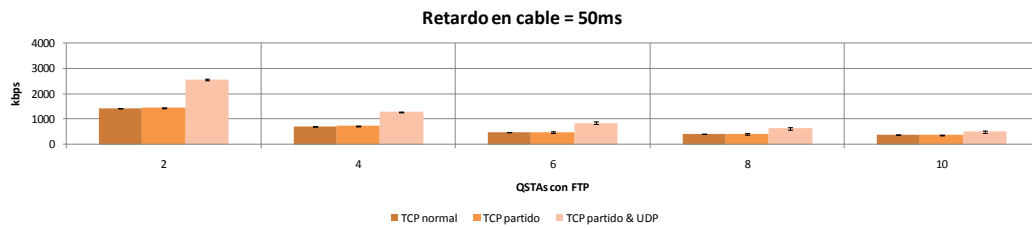


Figura 7.26. Resultados para varios usuarios FTP, un retardo (cable) de 50 ms y 30% de errores.

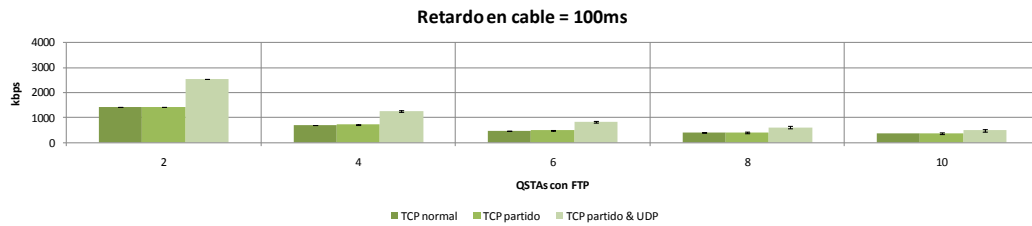


Figura 7.27. Resultados para varios usuarios FTP, un retardo (cable) de 100 ms y 30% de errores.

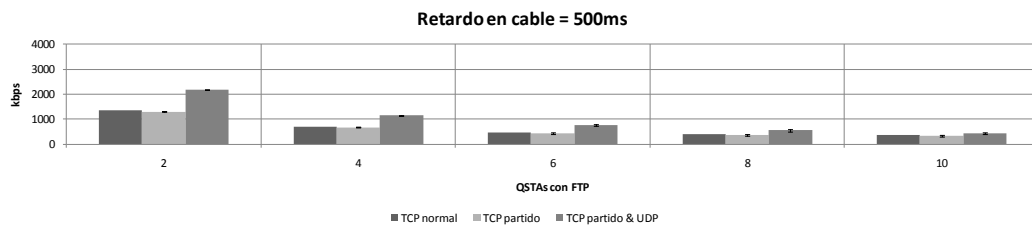


Figura 7.28. Resultados para varios usuarios FTP, un retardo (cable) de 500 ms y 30% de errores.

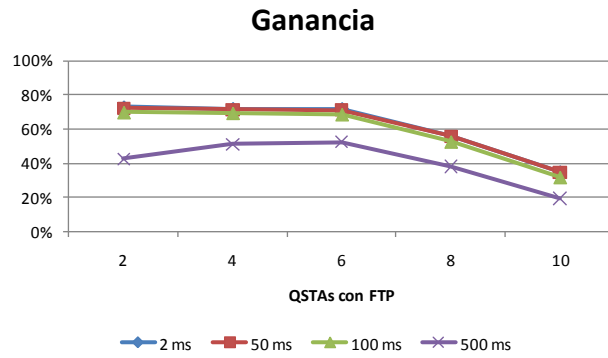


Figura 7.29. Ganancia para varios usuarios FTP y 10% de errores.

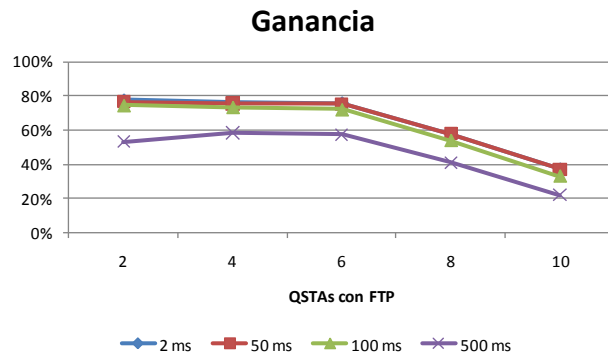


Figura 7.30. Ganancia para varios usuarios FTP y 20% de errores.

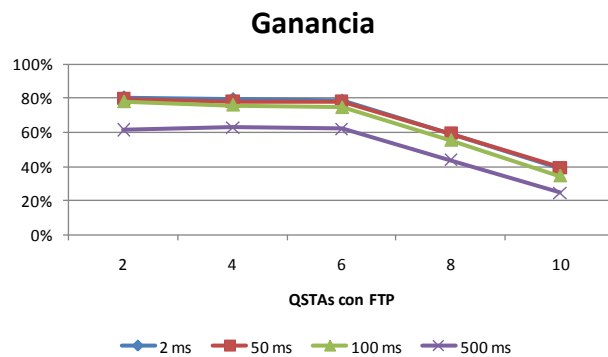


Figura 7.31. Ganancia para varios usuarios FTP y 30% de errores.

7.3 Conclusiones

En este capítulo se ha propuesto un esquema para mejorar el rendimiento de una conexión TCP que atraviese una red WLAN con soporte del mecanismo HCCA. El uso de HCCA como mecanismo de acceso para las aplicaciones basadas en TCP está justificado por las siguientes razones:

- 1) EDCA no permite garantizar requisitos de QoS, que pueden ser convenientes para servicios como vídeo sobre TCP.
- 2) HCCA elimina la injusticia intrínseca de las redes IEEE 802.11 entre los flujos ascendentes y descendentes.
- 3) HCCA permite un mayor control para desarrollar mecanismos y algoritmos más complejos, como la propuesta realizada en este capítulo.

Para el diseño realizado, primero se han separado los dos entornos –cableado e inalámbrico– a través de un *proxy* de rendimiento mejorado (PEP) existente en la literatura y denominado *TCP partido* (o *split TCP*).

Esta separación permite desarrollar la propuesta descrita, basada en que las funcionalidades de TCP son redundantes con las desarrolladas por la capa MAC de 802.11e con HCCA. Así, se elimina la capa TCP en las comunicaciones entre el punto de acceso y las estaciones, siendo ésta sustituida por un protocolo de transporte más sencillo (como puede ser UDP con una ligera modificación en su cabecera, para incluir la ventana ofertada del control de flujo).

El rendimiento de esta versión modificada de *TCP partido* –denominada *TCP partido & UDP*– ha sido evaluado y comparado con el uso normal del protocolo TCP y con la versión original de *TCP partido* en algunos escenarios relevantes. En particular se estudian los siguientes casos:

- Un único usuario FTP en la red inalámbrica, con diferentes tamaños de ficheros (50KB, 500KB, 5MB, ∞) y diferentes valores de *Round-Trip Time*

(retardos de ida de 2, 50, 100 y 500 ms).

- Un usuario FTP (diferentes tamaños de ficheros y diferentes RTT) y tráfico adicional generado con UDP (carga de 3 Mbps).
- Varios usuarios FTP (2, 4, 6, 8, 10) transmitiendo ininterrumpidamente, con diferentes valores de RTT.
- Varios usuarios FTP (siempre transmitiendo y con diferentes RTT) con distintas probabilidades de error de trama (10, 20, 30%) en el interfaz inalámbrico.

Los resultados muestran que, en los escenarios simulados, la solución propuesta mejora notablemente el rendimiento conseguido respecto al uso normal de TCP, siempre que no sea el enlace por cable el que limite la velocidad de transmisión debido al mecanismo de control de congestión de TCP.

Capítulo 8

Conclusiones

Este capítulo tiene como objeto presentar las conclusiones fruto de la realización de la tesis, dando una visión general del trabajo realizado.

En el presente trabajo se ha abordado el reto de mejorar la calidad de experiencia sobre redes IEEE 802.11, contribuyendo finalmente a varios aspectos. Para ello, se ha optado por utilizar ésta –la Calidad de Experiencia– como criterio clave para la toma de decisiones respecto a la gestión de los recursos radio. En concreto, se ha desarrollado una plataforma flexible –basada en el mecanismo de acceso coordinado HCCA– que, además de considerar la QoE, permite introducir otros criterios que el administrador de red puede configurar a su conveniencia. Además, el uso de este mecanismo ha permitido modificar la capa de transporte para las aplicaciones basadas en TCP, mejorándose de forma notable el rendimiento de las mismas.

Antes de explicar y analizar las soluciones propuestas, se ha descrito el funcionamiento de los principales estándares utilizados en redes inalámbricas de área local, i.e. el estándar IEEE 802.11 y sus variantes en la capa física. Además, se han descrito detalladamente los diferentes modos de acceso definidos tanto por la norma original (DCF y PCF) como por la corrección 802.11e (EDCA y HCCA), haciendo

164 | Conclusiones

énfasis en la falta de soporte de QoS de algunos y las garantías de otros. Las diferentes técnicas encontradas en la literatura para el cumplimiento de requisitos de QoS en este tipo de redes están limitadas por su ámbito de aplicación, no habiendo encontrado ninguna que satisfaga de forma eficiente los requisitos exigidos en un entorno de tráfico heterogéneo con diferentes tipos de servicios, siendo éste el principal problema a resolver en esta tesis.

Posteriormente se ha presentado el concepto de *Calidad de Experiencia* que se utilizará en los diversos mecanismos diseñados en la tesis. Se ha considerado como hipótesis de partida que la incorporación de criterios de QoE en estos algoritmos mejorará de forma significativa la calidad percibida por los diferentes usuarios de la red en escenarios en los que el tráfico proviene de diferentes servicios. Para simplificar la gestión de los servicios considerados, se ha realizado una clasificación de los mismos introduciendo el concepto de *Clase de Tráfico*, agrupando servicios con el mismo tipo de requisitos de QoS. Esta clasificación considera las siguientes clases: *conversacional*, *streaming*, *TCP de tiempo real*, *interactiva* y *de fondo*.

Debido a la necesidad de estimar de forma automática –y en tiempo real– la calidad percibida por los usuarios, se ha presentado el concepto de *función de utilidad*, definida como aquella que permite mapear indicadores objetivos del rendimiento (QoS) en la calidad subjetiva que experimentaría un usuario (QoE). Se han especificado funciones de utilidad para el conjunto de servicios elegidos, algunas basadas en la bibliografía existente (VoIP, *streaming* de vídeo, FTP) y otras diseñadas expresamente (tráfico web y *streaming* de vídeo sobre TCP). Basándose en estas funciones de utilidad, se ha propuesto una metodología unificada para evaluar el rendimiento de las diferentes mejoras diseñadas, utilizando la Nota de Opinión Media (MOS) como métrica común para los diferentes servicios.

La primera solución diseñada propone la adopción de un planificador de paquetes desarrollado como una variante del Planificador–Conformador Monolítico (MSS), para su uso con el mecanismo de acceso HCCA. Este algoritmo de planificación fue elegido por la idoneidad de sus características, ya que es capaz de garantizar tanto la velocidad de transmisión negociada, como un retardo acotado en la transmisión de los paquetes. Su rendimiento se ha analizado en términos de calidad experimentada por los usuarios y de justicia entre los diferentes servicios, incluyendo una comparativa con algunas disciplinas

de planificación destacadas de la literatura actual. Los resultados obtenidos permiten concluir que, en los escenarios estudiados, MSS es más eficiente (en términos de QoE y de justicia) que los planificadores de referencia considerados.

La propuesta más importante de esta tesis es el diseño de una arquitectura para la gestión de recursos radio denominada el Método de los Territorios. Este mecanismo actúa de recubrimiento sobre la asignación de turnos realizada por el planificador de paquetes utilizado en HCCA, con los objetivos de garantizar la QoE de los servicios más prioritarios y de mejorar la utilización del ancho de banda disponible. Para ello, basa sus decisiones en la calidad experimentada por los usuarios de estos servicios, aumentando o disminuyendo los recursos asignados para los mismos en función del MOS percibido. Debido al esquema adoptado de *asignación con derecho preferente*, las clases más prioritarias pueden *atribuirse* recursos de otras clases con menos prioridad. Sin embargo, gracias a la definición de una capacidad dedicada para cada clase de tráfico se garantiza un rendimiento mínimo para cada tipo de servicio.

El rendimiento de esta solución ha sido evaluado y analizado tanto para condiciones estáticas –sin variaciones en el número de usuarios durante la simulación– como dinámicas. En el primer caso se ha mostrado su correcto funcionamiento para diferentes niveles de carga –*media, alta y muy alta*–, comprobándose la priorización de servicios y su rendimiento mínimo gracias a la capacidad dedicada. En el segundo caso se ha examinado el comportamiento dinámico de la solución propuesta en situaciones cambiantes, modificándose –en el escenario simulado– el tamaño de los territorios de forma adecuada según la carga por servicio. Este análisis ha permitido evidenciar la rápida respuesta ante crecimientos de tráfico y la reacción más gradual en el caso de decrementos (evitándose posibles inestabilidades de QoE y territorios).

Basándose en esta arquitectura (el Método de los Territorios), se han presentado diferentes políticas de gestión de recursos que permitirían a un operador o administrador de red configurar cómo se comparte la capacidad inalámbrica en una red IEEE 802.11e que utilizase el mecanismo de acceso al medio HCCA. En concreto, se han propuesto cuatro políticas diferentes de reparto de recursos:

- *Con derecho preferente.*
- *Sin derecho preferente.*

166 | Conclusiones

- *Justo en términos de Calidad de Experiencia.*
- *Aumento de beneficio.*

El análisis realizado muestra, en los escenarios simulados, las siguientes conclusiones.

La política de *reparto con derecho preferente* permite la priorización de los servicios. La aproximación *sin derecho preferente* distribuye los recursos dependiendo de la evolución temporal de la demanda, sirviendo antes al primero que llega. Por tanto, la asignación de recursos no es conocida *a priori* y no sigue un criterio predefinido. Este no es el caso de la política *justa en términos de Calidad de Experiencia*, que permite al operador igualar la calidad subjetiva entre las distintas clases de tráfico. Por último, el algoritmo para *aumento de beneficio* aumenta las ganancias del operador de red (con factores que permiten ajustar los beneficios conseguidos por cada clase de tráfico). En cualquier caso, por encima de estas políticas se puede ajustar una capacidad dedicada para cada clase de tráfico garantizando un rendimiento mínimo para esos servicios.

Estas políticas se han comparado entre ellas y con el planificador MSS, en términos de calidad subjetiva, beneficio conseguido y justicia entre servicios. Esta comparativa muestra que las diferentes políticas propuestas para el Método de los Territorios mejoran notablemente el rendimiento respecto a utilizar únicamente el planificador MSS y consiguen los objetivos para los que fueron diseñadas.

La última mejora presentada consiste en un esquema para mejorar el rendimiento de una conexión TCP que incluya una red IEEE 802.11 con soporte de HCCA. El uso de HCCA como mecanismo de acceso para las aplicaciones basadas en TCP está justificado por las siguientes razones:

- 1) EDCA no permite garantizar requisitos de QoS, que pueden ser convenientes para servicios como vídeo sobre TCP.
- 2) HCCA elimina la injusticia intrínseca de las redes IEEE 802.11 entre los flujos ascendentes y descendentes.
- 3) HCCA permite un mayor control para desarrollar mecanismos y

algoritmos más complejos, e.g. la propuesta realizada para mejorar TCP en este tipo de entornos.

Para esta propuesta se han separado los dos entornos –cableado e inalámbrico– a través de un *proxy* de rendimiento mejorado (PEP) existente en la literatura y denominado *TCP partido* (o *split TCP*). Esta separación permite que el esquema desarrollado elimine las funcionalidades de TCP que son redundantes con las desarrolladas por la capa MAC de una red 802.11e con HCCA. Por ello, se utilizará un protocolo de transporte más sencillo –e.g. UDP con una ligera modificación en su cabecera– en las comunicaciones entre el punto de acceso y las estaciones.

El rendimiento de esta versión modificada de *TCP partido* –denominada *TCP partido & UDP*– ha sido evaluado y comparado con el uso normal del protocolo TCP y con la versión original de *TCP partido* en algunos casos considerados como relevantes. Los resultados muestran que, en los escenarios simulados, la solución propuesta mejora notablemente el rendimiento conseguido usando una implementación normal de TCP, siempre que no sea el enlace por cable el que limite la velocidad de transmisión por el mecanismo de control de congestión de TCP.

Capítulo 9

Conclusions

This chapter aims at summarizing the main conclusions drawn as a result of the work presented in the different chapters.

In the present thesis, the challenge of improving the quality of experience over IEEE 802.11 networks has been addressed. To reach this goal, the first design decision has been the adoption of QoE as main criterion for managing the scarce radio resources. More precisely, a flexible platform has been developed –based on the HCCA coordinated medium access– which, in addition to consider QoE, allows the network administrator to configure other criteria at his/her convenience. Moreover, the usage of this mechanism permits the modification of the transport layer for TCP-based applications, improving noticeably their performance.

Before explaining and analyzing the proposed solutions, the main standards of wireless local area networks, i.e. IEEE 802.11 and its amendments related to the physical layer, have been described. In addition, the different medium access mechanisms defined by the legacy standard (DCF and PCF) and by the 802.11e amendment (EDCA and HCCA) have been described, emphasizing on the QoS support aspect. The different techniques found in the literature for guaranteeing QoS requirements in this kind of

170 | Conclusions

networks are limited by their scope, not finding any proposal that satisfies the requirements in a scenario with a heterogeneous traffic mix with different services. This is the main objective of this thesis.

Subsequently, the Quality of Experience concept has been presented. It will be used in the diverse mechanisms designed during this thesis, assuming –as initial hypothesis– that the incorporation of QoE criteria in these algorithms will drastically improve the perceived quality of users with different services. To simplify the management of the considered services, they are classified into *Traffic Classes*, i.e. set of services with similar type of requirements. The following classes have been considered: *conversational, streaming, real-time TCP, interactive* and *background*.

The first designed solution proposes the adoption of a packet scheduler, which is a variant of the Monolithic Shaper-Scheduler (MSS), for its usage on HCCA. This scheduling algorithm was selected because of its suitable characteristics, since it is able to guarantee both data rate and bounded delay. Its performance has been analyzed by means of simulations, in terms of end-user's perceived quality and fairness between services, including a comparison with some reference schedulers from the literature. The analysis derived from the results show that, in the simulated scenarios, MSS is more efficient (in terms of both QoE and fairness) than the other considered schedulers.

The main contribution of this thesis is the design of an architecture for radio resource management, designated the Territory Method. This mechanism is built on top of the HCCA packet scheduler, aiming at guaranteeing QoE for high priority services and increasing the utilization of the available bandwidth. For this purpose, its decisions are based on the perceived quality for those services, increasing or reducing the reserved resources depending on the average MOS. Since the adopted scheme is *preemptive*, the highest priority classes can acquire resources from other lower priority classes. However, thanks to the definition of a dedicated capacity for each traffic class, a minimum performance is guaranteed for each type of service.

The performance of this solution has been evaluated and analyzed for both static –no changes in the number of users during the simulation– and dynamic simulations. The first set of experiments have demonstrated the proper behavior of the Territory Method for different load levels –*medium, high* and *very high*–, achieving the prioritization of

services and their minimum performance (thanks to the minimum dedicated capacity). In the second case, the dynamic operation of the solution has been tested under changing conditions, varying –in the simulated scenario– the territories sizes according to the load per service. This analysis proved the fast reaction under load increments and a smoother reaction in the case of decrements (preventing possible instabilities of QoE and territories).

Based on this general framework (the Territory Method), several policies have been designed that allow a network administrator to configure how the scarce wireless capacity is shared under an IEEE 802.11e network using the HCCA medium access mechanism. More precisely, four policies have been proposed: *preemptive*, *non-preemptive*, *MOS-driven fairness* and *revenue improvement*. The conducted analysis has shown, in the simulated scenarios, the following conclusions. The *Preemptive* policy permits the prioritization of the services. The *Non-preemptive* approach distributes the resources depending on the temporal evolution of the demand, in a *first come, first served* fashion. Therefore, the resource allocation is not known *a priori* and it does not follow a predefined criterion. This is not the case of the *MOS-driven Fairness* policy, which allows the operator to equalize the subjective quality between traffic classes. Finally, the *Revenue Improvement* algorithm increases the profit of the network operator (with factors to tune the incomes from each service). On top of these policies, the adjustment of a dedicated capacity for each traffic class guarantees a minimum performance for those services.

These policies have been compared to the MSS scheduler, in terms of subjective quality, obtained revenue and fairness between services. This comparison has shown that the different policies proposed for the Territory Method considerably improve the performance with respect to only using the MSS scheduler. They also achieve the goals they were designed for.

The last proposal consists in a scheme to improve the performance of TCP connections, assuming they cross a wireless network with HCCA support. The usage of HCCA for TCP-based applications is justified because of the following reasons:

- 1) EDCA cannot guarantee QoS requirements, which might be convenient for some services like progressive download video (over TCP).

172 | Conclusions

- 2) HCCA removes the intrinsic unfairness of uplink and downlink flows in IEEE 802.11 networks.
- 3) HCCA permits a higher control for developing more complex mechanisms and algorithms, like the proposed solution for improving TCP in these environments.

For this proposal, first both environments –wired and wireless– have been separated by using a performance enhancing proxy (PEP) available from the current literature (*split TCP*). This division facilitates the removal of TCP functionalities that are redundant with the ones available from the MAC layer of an 802.11e network using HCCA. For that reason, a simpler transport protocol has been proposed –UDP with a small header modification– for the communications between the access point and the stations.

The performance of this modified version of *split TCP* –called *split TCP & UDP*– has been evaluated and compared to the normal usage of the TCP protocol and the original version of *split TCP* in some relevant scenarios. The results show that, in these scenarios, the proposed solution outperforms both normal TCP and split TCP connections (as long as the wired link is not limiting the data rate due to the congestion control mechanism of TCP).

Apéndice A

Thesis summary

In this chapter, the main contributions of the thesis are summarized in English.

A.1 Introduction and problem statement

Since James Clerck Maxwell first described electromagnetic waves a century and a half ago, wireless communications have noticeably evolved, especially in the last decades. The term *wireless* first appears when, in 1896, Guglielmo Marconi developed the first wireless telegraph system (later, in 1909, he shared the Nobel Prize for Physics with Karl Ferdinand Braun “in recognition of their contributions to the development of wireless telegraphy”).

The main application that promoted the development of wireless technologies is voice communication, with an increasing demand for mass-market: the citizen-band radio

craze, the walkie-talkie market, the connection among radios and the public telephone system, and the incredible proliferation of cellular phones.

However, wireless communications are not limited to the voice service but include data transmission, mainly due to the development of the Internet. Since its origins, Internet has experienced a dizzying growth in both number of users and offered services. These are not limited to web browsing and electronic mail, but also new applications have appeared with requirements that were not considered in the original design of the Internet, as video streaming (e.g. IPTV) or again voice communications (VoIP). This growth of data communications has highly influenced the development of wireless technologies.

Nowadays there exists a high number of wireless technologies that enable high speed data transmission, e.g. Ultra-Wide Band, WiMax, HSPA or LTE. In the local area environment, IEEE 802.11 [1] (also known as WLAN or Wi-Fi) is the predominant standard for wireless communications. These networks are widely deployed in different scenarios such as offices, hotspots in airports, hotels, coffee shops, at home, etcetera.

Similarly to the Internet, the IEEE 802.11 standard was not originally conceived to support Quality of Service (QoS) requirements –as those required for multimedia services– but it is based on a *best effort* scheme. The legacy standard included two medium access mechanisms, DCF (Distributed Coordination Function) –mandatory distributed access mode– and PCF (Point Coordination Function) –optional scheme for coordinated access–. Even though PCF was proposed to support real time services, it cannot guarantee either bandwidth or maximum delay since the transmission time of a mobile station is not bounded.

Even though there exist a number of solutions based on the 802.11e amendment [2] –e.g. [3] [4] [5] [6] [7] [8] [9] [10] [11] [12] [13]– to carry real time traffic in these networks with the required guarantees, none of them –to the best of the author’s knowledge– is able to simultaneous and efficiently support real time and elastic services (e.g. applications based on TCP), typical Internet traffic. This will be one of the fundamental objectives of the work developed in this thesis.

On the other hand, several works [14] [15] [16] [17] [18] [19] suggest that significant

improvements on the system performance can be obtained by incorporating the Quality of Experience (QoE) –end-user’s satisfaction based on his/her subjective perception– in the design of algorithms and procedures for resource management. This initial hypothesis will be considered as the basis for some of the main contributions in this work.

A.1.1 Dissertation’s goals

The overall objective in this thesis is to provide new mechanisms with the aim of improving the QoE experienced by the end-user in a wireless environment –based on the IEEE 802.11 standard–. More precisely, the designed proposals shall consider scenarios with traffic mixes generated by different services with diverse requirements, such as real time and elastic traffic applications.

The particular objective of this thesis can be broken down into the following targets:

- The provision of metrics that objectively quantifies the quality of experience perceived by end-users, for the most relevant services in wireless networks.
- The development of mechanism that allows an IEEE 802.11 network to guarantee strict QoS requirements.
- The design of algorithms and procedures for radio resource management for improving the end-user’s experienced quality in an environment with heterogeneous traffic mixes.
- The improvement of the TCP protocol for wireless environments (IEEE 802.11e with HCCA), in order to better exploit the capacity of the physical medium.

A.2 IEEE 802.11e and HCCA

The WLAN standard (IEEE 802.11 [1]) does not define any access mechanism to support strict real time requirements. To eliminate this lack of QoS support, IEEE 802.11e [2] defines a new coordination function called Hybrid Coordination Function (HCF). HCF combines previous legacy access schemes (Distributed Coordination Function and Point Coordination Function) with enhanced QoS mechanisms to achieve service differentiation. For that purpose, HCF provides both distributed and centrally controlled channel access mechanisms similar to DCF and PCF in the original standard. The distributed, contention-based channel access mechanism of HCF is called Enhanced Distributed Channel Access (EDCA), and the centrally controlled, contention-free channel access mechanism is called HCF Controlled Channel Access (HCCA).

EDCA provides differentiated, distributed medium access by using different priorities for different types of data traffic, defining four Access Categories (ACs). Service differentiation is introduced by making use of a different set of parameters for each AC to contend for the medium. Therefore, relative priorities are provisioned by configuring the time –with a random value within a defined range– to access the channel once it is detected idle. In that sense, EDCA provides relative but not guaranteed QoS support since a higher priority access category has a higher probability to access the channel, but eventually it may occur (with a smaller probability) that a lower priority AC can acquire the wireless medium. After the contention, a QoS Station (QSTA) is allowed to transmit multiple frames within an EDCA transmission opportunity (referred as to TXOP), whose maximum duration depends on the particular AC.

HCCA mechanism is designed for parametric QoS support, utilizing a hybrid coordinator (HC) to manage the allocation of wireless medium data transfer bandwidth. The HC has higher priority access than wireless stations to be able to schedule transmission opportunities to these stations in both contention and contention free periods. A QSTA, based on its requirements, requests the HC for TXOPs, both for its own transmissions as well as for transmissions from the *QoS access point* (QAP) to itself. The hybrid coordinator, which is usually collocated at the QAP, either accepts or rejects the request based on the parameters supplied by the QSTA. For uplink transmissions, the HC polls the QSTA based on its requirements. For downlink transfers, the QAP directly

obtains TXOPs from the collocated HC and delivers the queued frames to the QSTA.

In HCCA, each QSTA may establish up to eight Traffic Streams (TSs), which are characterized by Traffic Specifications (TSPEC) negotiated between the QSTA and the QAP. The TSPECs include a number of parameters [2], most of them optional, which are used for traffic and requirements description. Some of the main TSPEC fields for a TS_i are *Mean Data Rate* (R_i), *Nominal Service Data Unit (SDU) Size* (N_i), *Maximum SDU Size* (M_i), *Minimum PHY Rate* (F_i), *Delay Bound* (D_i) and *Maximum Service Interval* (MSI_i). The characterization of the traffic stream includes the first five parameters, whereas the MSI_i is a service requirement. Specifically, the delay bound indicates the maximum lifetime of an SDU (before dropping) and the maximum service interval specifies the maximum time that can elapse between two subsequent polls at the same QSTA.

The QAP initiates Controlled Access Phases (CAPs) with a higher priority than contention-based access mechanisms (i.e. DCF or EDCA) by waiting a *PCF interframe space* (PIFS) period before accessing the medium. A CAP consists of one or several TXOPs, during which the QAP may either transmit data frames to QSTAs or receive from QSTAs by sending first a polling frame. This poll specifies the maximum time a QSTA can utilize the wireless medium (including interframe spaces and acknowledgments); hence it provides QoS support for strict real time applications. The QAP can start CAPs (for HCCA TXOPs) in both contention and contention-free periods, whereas EDCA TXOPs are allowed only in the contention period (see Figure A.1).

Requirements for HCCA schedulers

The scheduling of TXOPs shall be done according to the negotiated TSPEC

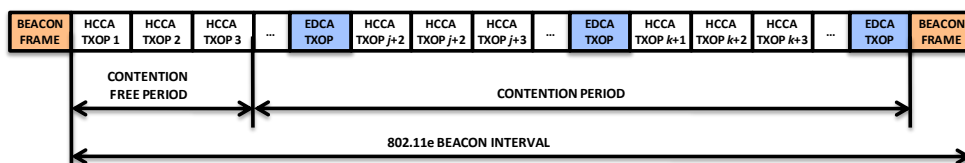


Figure A.1. IEEE 802.11e beacon interval.

parameters of the admitted TS. Therefore, as specified in the IEEE 802.11e standard, the cumulative duration $W_i(t_1, t_2)$ of the assigned TXOPs for TS_i , for any time interval $[t_1, t_2]$ greater than a minimum Specification Interval (which is advertised by the QAP), shall satisfy the following equation:

$$W_i(t_1, t_2) \geq \frac{t_2 - t_1 - \Delta_i}{\frac{N_i}{R_i}} \cdot tx(N_i, \Gamma_i) \quad (\text{A.1})$$

where $tx(N_i, \Gamma_i)$ is the time to transmit a Service Data Unit (SDU) of nominal size N_i at the physical rate Γ_i , R_i is the *Mean Data Rate*, and Δ_i is equal to MSI_i if it is specified or D_i (delay bound) otherwise [2]. The nominal transmission time for TS_i , including MAC control frames and interframe spaces, can be computed as:

$$tx(N_i, \Gamma_i) = \text{SIFS} + \left(\frac{h_{PHY} + h_{MAC} + N_i}{\Gamma_i} \right) + \text{SIFS} + \left(\frac{h_{PHY} + h_{ACK}}{\Gamma_{BASIC}} \right) \quad (\text{A.2})$$

where SIFS is the *short interframe space* period, h_{PHY} , h_{MAC} and h_{ACK} are the sizes of the physical header, MAC header and acknowledgment frame respectively, and Γ_{BASIC} is the physical transmission rate for the acknowledgment frames.

HCCA reference scheduler

The 802.11e standard does not specify how to schedule the TXOPs. However, a sample scheduler is described and it can be used as a reference. This scheduler first determines a global *Service Interval* (SI), which is smaller than the MSI of all the admitted streams and multiple of the beacon interval (which includes both contention and contention-free periods, see Figure A.1). Any admitted stream will be able to get a TXOP at the end of each SI, providing at least one TXOP within each TS *Maximum Service Interval* time limit. The scheduler then determines the TXOP duration needed for each stream by considering the number of packets that may arrive within an SI.

$$TXOP_i = \max\left(\frac{nMSDU_i \times N_i}{\Gamma_i} + O, \frac{M_i}{\Gamma_i} + O\right) \quad (\text{A.3})$$

where $TXOP_i$ is the duration of the transmission opportunity for TS_i , N_i is the nominal MSDU size, Γ_i is the minimum physical transmission rate, M_i is the maximum allowable

size of MSDU, O denotes the overhead (which includes interframe spaces, acknowledgments and polling frames) in time units and $nMSDU_i$ is the number of MSDUs arrived in the QSTA at mean data rate λ_i during the SI period, which can be calculated from:

$$nMSDU_i = \left\lceil \frac{SI \times \lambda_i}{N_i} \right\rceil \quad (\text{A.4})$$

An admission control algorithm is also suggested for the sample scheduler. A new flow is admitted if the following expression is satisfied:

$$\frac{TXOP_{k+1}}{SI} + \sum_{i=1}^k \frac{TXOP_i}{SI} \leq \frac{T_{beacon} - T_{contention\ period}}{T_{beacon}} \quad (\text{A.5})$$

where k is the number of existing flows, T_{beacon} is the duration of a beacon period and $T_{contention\ period}$ is the time used for EDCA access.

A.2.1 Related works

The work on this thesis is based on HCCA since, as it was commented, it is the only medium access mechanism for IEEE 802.11 networks that guarantee strict real-time requirements for the carried flows. In addition, HCCA can be implemented in a standard 802.11e card by means of software drivers [28].

In this section, in order to establish the context of this work, a survey of traffic scheduling in 802.11e networks for the HCCA mechanism found in the literature is provided. The different proposals have been categorized into two main classes: scheduling algorithms that comply with the requirement of the standard (A.1), and other schedulers that improve the efficiency for CBR and/or VBR traffic without checking the previous condition. In general terms, within these two categories, other relevant works have considered scheduling disciplines based either on the deadline of packets [3] [4] [5] [6] [7] or on queue length estimations [8] [9] [10] [11] [12] to select the traffic flow for data transfer.

This survey will focus on the first category, since the proposals in this thesis are intended to guarantee both the delay and bandwidth requirements, therefore complying the standard specifications. To the best of the author's knowledge, only four algorithms have been demonstrated to provide the required guarantees for each flow: the *sample* scheduler (previously described), the SETT-EDD scheme [3], the RTH [4] and the WTTP [13] schedulers. In addition to these algorithms, the scheduler proposed in [5] also fulfills the previous conditions since, as it will be explained later, all the *latency-rate* schedulers, by definition, provides a guaranteed service time to each flow (see Theorem 1 in Section A.4).

The Scheduling based on Estimated Transmission Time uses the Earliest Due Date (SETT-EDD [3]) policy. This scheduler estimates the minimum and maximum departure times (using the new required *minimum service interval (mSI)* parameter and the *maximum service interval (MSI)*). Using the arrival time of the packets (which is estimated for QSTA queues), SDUs are served using the Earliest Due Date (EDD) algorithm, which performs better than the *reference* scheduler especially with VBR traffic. However, SETT-EDD does not provide bounded QoS parameters such as bandwidth or delay.

Cicconetti *et al* propose a Real-Time HCCA Scheduler (RTH) [4], which is based on the Earliest Deadline First (EDF) scheduling algorithm with the Stack Resource Policy (SRP). This scheduler can be found in the literature for real-time periodic task scheduling in multiprogramming environments. It is composed by three blocks: an *admission control* algorithm, an offline *timetable computation* and an online *enforcement procedure*. Since the periodical timetable is calculated when a new flow is admitted, this scheduler is intended only for CBR traffic and therefore it cannot efficiently handle elastic traffic.

The Wireless Timed Token Protocol (WTTP) [13] provides CBR streams with a fixed capacity, whereas a minimum reserved rate is allocated for VBR flows. This approach allows the VBR traffic flows to utilize the *unused* bandwidth. WTTP is based on the Timed Token Protocol (TTP) used for FDDI networks, which is basically a round robin algorithm. The round robin list is composed only by the backlogged flows, so the flows are added and removed based on their queue sizes (piggybacked from QSTAs to QAP). This scheduler guarantees that each TS is granted TXOPs at a sufficient rate. However, it does not guarantee that all packets will be transmitted within their delay

bound.

Foronda *et al* [5] utilize an admission control and a modified *latency-rate* scheduler with a token bucket shaper. Although it has some similarities with one of the approaches designed during this thesis, several drawbacks have been identified. First, the scheduling is performed periodically, which reduces its flexibility while increasing the latency of the server (higher than a SI period). Despite the latency may be acceptable for 802.11a networks due to its high data rate, it will lead to an extremely high delay for 802.11b or slower physical interfaces. As it will be exposed later, the scheduler proposed in this work is executed on a packet basis, which reduces the latency. Second, the mean data rate has to be specified for each traffic class, which is suitable for real-time traffic (CBR and certain VBR cases) but not for TCP applications such as web browsing or FTP downloading, since these applications will try to increase their throughput based on the TCP congestion control mechanism.

A.3 Quality of Experience

Quality of Experience is defined as the “*the overall acceptability of an application or service, as perceived subjectively by the end-user*” [29]. Hence, QoE includes end-to-end system effects (client, terminal, network, service infrastructure, etc.) and it can be influenced by user’s expectations regarding the service or due to the environment.

On the other hand, *Quality of Service* is defined as the “*totality of characteristics of a telecommunications service that bear on its ability to satisfy stated and implied needs of the user of the service*” [30].

Both concepts imply multiple factors –sometimes overlapped– but, in general terms, it is assumed that QoE is related to the end-user’s subjective perception whereas QoS is an objective concept determined by observable network performance metrics.

As it was commented in Section A.1, the main contributions of this work (see Section A.1.1) are based on the following hypothesis: the introduction of QoE criteria in

the design of Radio Resource Management algorithms, compared to the traditional strategies that only consider QoS parameters, can significantly increase the global perceived quality –which considers different traffic types– in a IEEE 802.11 network.

For this purpose, it is necessary to estimate the quality perceived by end-users when utilizing any of the different considered services. Thus, a common metric is required –so the performance of different services can be compared– and its calculation must be done in real time –in order to accordingly react to the varying network conditions–.

The Mean Opinion Score (MOS) is proposed as the unifying metric for the different services [31]. This scale (from 1 –bad– to 5 –excellent–) is not only used for voice but also for data applications [32], video-telephony [33] and IPTV [34].

In order to simplify the RRM algorithms –considering individual services–, a *traffic class* (TC) defines a set of services with similar QoS requirements (extended from UMTS [35]). The following classification was adopted, which includes their fundamental characteristics for QoS:

- **Conversational:** This class includes audio (VoIP) and video (videoconference) traffic that demands interactivity. They follow a conversational pattern –requiring stringent and low delay– and shall preserve time relation (variation) between packets –i.e. *jitter*–. The traffic is bidirectional between humans, therefore the tolerance limits –for delay and jitter– are determined by their perception mechanisms –hearing and vision apparatus–.
- **Streaming:** It comprises real-time audio and video delivery not requiring interactivity –i.e. not conversational–, even though the destination is also a human. Transportation is unidirectional.

This scheme is characterized by jitter, which shall be bounded. In this case the requirement for delay is not strict since it can perform a time alignment of data –by means of a playout buffer–. Due to this alignment, the requirements for jitter are also less demanding.

- **TCP real-time:** This traffic class is identical to the previous one, but for

services that use TCP as transport protocol.

- **Interactive:** In this case, the traffic is generated by an end-user –which can be a human or a machine– who/which is requesting information to remote equipment (e.g. a server). Some typical examples are web browsing (with human-machine interaction) or the access to web services (e.g. SOAP [37] or REST [38] protocols between machines)

Interactive traffic follows a client-server model, so the round-trip time is one of its main characteristics. Another requirement is the reliability of packet transmission (without errors and using flow control).

- **Background:** This class is defined for unidirectional traffic without delay requirements but demanding reliability. Typical examples are background file transfer (e.g. FTP), electronic mail delivery and the short message service (SMS).

For the present work, the following services were considered: Voice over IP (VoIP) for *Conversational*, RTP/UDP video streaming for *Streaming*, TCP video streaming (e.g., YouTube video broadcasting) for *TCP real-time*, web browsing for *Interactive* and FTP for *Background*.

The end-user experience depends on the specific traffic service. Next, it is shown how to compute an estimate of a subjective user perception using objective quality metrics for each service, by means of utility functions. These functions map objective network metrics (such as delay or throughput) into subjective quality (MOS).

Utility function for VoIP

The MOS estimation for VoIP is based on the E-model [22], whose formula is in Eq. A.6. For the performance evaluation, an indoor wireless scenario was simulated using G.711 codec. The parameters of Table A.1 were assumed.

$$MOS = \begin{cases} 1.0 & \text{if } R < 0 \\ 4.5 & \text{if } R > 100 \\ 1 + 3.5 \cdot 10^{-3}R + 7 \cdot 10^{-6}R \cdot (R - 60) \cdot (100 - R) & \text{otherwise} \end{cases} \quad (\text{A.6})$$

Parameter	Value	Description
A	5	Advantage factor
R_0	93.2	Basic signal-to-noise ratio
I_s	0.0	Simultaneous impairment factor
I_e	0.0	Equipment impairment factor
B_{pl}	0.043	Packet-loss robustness factor

Table A.1. E-Model parameters for indoor with G.711 codec.

where

$$R = R_0 - I_s - I_d - I_{e,eff} + A$$

$$I_d = 0.024 \cdot D + \max\{0, 0.11 \cdot (D - 177.3)\}$$

$$I_{e,eff} = I_e + \frac{(95 - I_e) \times lp}{lp + B_{pl}}$$

lp is the loss probability and D is the delay in milliseconds.

Utility function for Video streaming

The subjective quality evaluation for video streaming is based on the EvalVid framework [44], which evaluates the differences between the original and the resulting images after network transmission. However, due to the image reconstruction process, it cannot be used for online algorithms but only for offline quality estimations.

In [23], two learning models were described for *online* video quality prediction using EvalVid. Based on the temporal (movement) and spatial (edges, brightness) features, videos are classified into three groups: *slight movement* (SM), *gentle walking* (GW) and *rapid movement* (RM). The video quality, based on the measured Frame Rate (FR, fps), Sending Bit Rate (SBR, kbps) and Packet Error Rate (PER, between 0 and 1), is given by:

$$MOS = \frac{a_1 + a_2 FR + a_3 \ln(SBR)}{1 + a_4 PER + a_5 (PER)^2} \quad (\text{A.7})$$

The a_i coefficients for SM, GW and RM are shown in Table A.2.

<i>Coefficient</i>	<i>SM</i>	<i>GW</i>	<i>RM</i>
a_1	4.5796	3.4757	3.0946
a_2	-0.0065	0.0022	-0.0065
a_3	0.0573	0.0407	0.1464
a_4	2.2073	2.4984	10.0437
a_5	7.1773	-3.7433	0.6865

Table A.2. a_i definitions for video streaming MOS.

Utility function for TCP video

TCP video is buffered at the receiver according to a dual-threshold strategy [45]. The first threshold (*BTT*, *buffered time threshold*) defines the minimum buffer size (in time units) before playout starts. The second threshold defines the buffer size at which the receiver signals the server to stop sending data.

Due to the TCP reliable delivery, subjective quality of TCP video streaming can be mostly affected by buffer underruns, which interrupt the playout process. Since buffer underruns are caused by short-term variations, an *instantaneous* quality metric estimate cannot be obtained from user surveys that last tens or hundreds of seconds. Therefore, a simplified artificial MOS function has been designed which depends only on the instantaneous buffer conditions. The criteria for designing the function are:

- The incurred delay due to initial buffering is usually expected by the users and accordingly it should not decrease the MOS.
- Buffer underruns shall abruptly reduce MOS to 1 since they halt the playout, which is annoying for the end-user.
- When rebuffering after an underrun, MOS should not increase rapidly. This reflects the fact that after a video interruption, it takes some time before the user return to be satisfied again. Besides, this approach avoids unwanted MOS oscillations.
- If the buffer level gets below the first threshold, then the MOS should decrease, albeit not abruptly, so as to capture the likelihood that the buffer will underrun.

Figure A.2 shows an example of the desired behavior. The MOS function for TCP video streaming is given in Eq. A.8 and A.9.

$$MOS[k] = \begin{cases} 5 & \text{if } BT \geq BTT \\ 1 + (5 - 1) \times BT/BTT & \text{if } 0 < BT < BTT \\ 1 & \text{if } BT \leq 0 \end{cases} \quad (\text{A.8})$$

$$AVGMOS[k] = \begin{cases} MOS[k] & \text{if } MOS[k] < AVGMOS[k - 1] \\ \alpha AVGMOS[k - 1] + (1 - \alpha)MOS[k] & \text{otherwise} \end{cases} \quad (\text{A.9})$$

where BT is the *buffered time*, i.e. playout time corresponding to the packets in the receiver buffer.

The α parameter allows the network administrator to adjust the end-user impairment after a buffer underrun. The *recovery time* (RT) is defined as the time elapsing from a complete rebuffering until the user is totally satisfied again (i.e. $MOS = 5 - \varepsilon$). This time depends on α as follows:

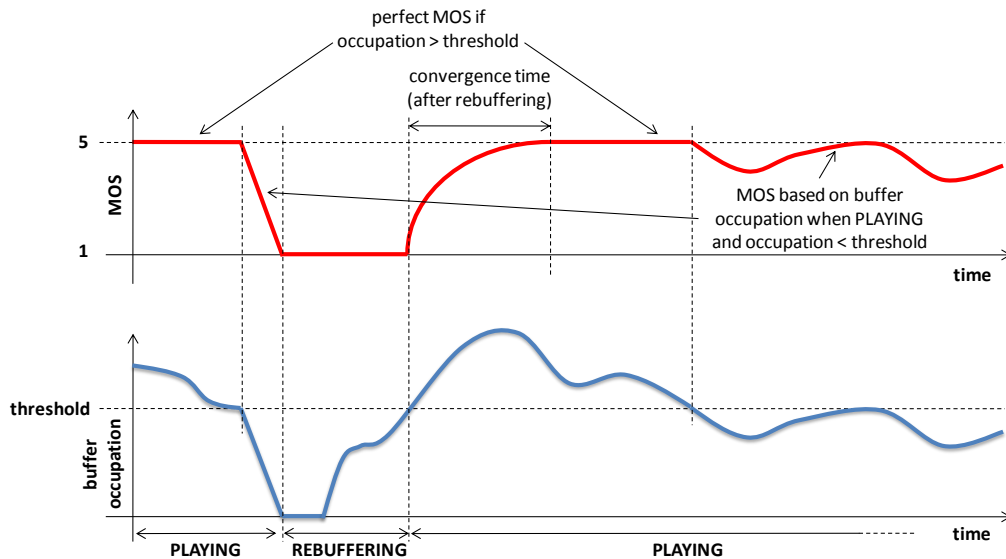


Figure A.2. Artificial MOS evolution for TCP video streaming.

$$\alpha = \left(\frac{RT}{MUP} \right) \sqrt{\frac{\varepsilon}{4}} \quad (\text{A.10})$$

where MUP is the *MOS Update Period* (time interval for collecting objective metrics required for estimating the QoE) in seconds. As an example, $\alpha=0.48$ implies $CT = 5s$, with $MUP = 1s$ and $\varepsilon = 0.1$.

Utility function for *Web browsing*

The MOS function was obtained after a survey [25] [16] during which 52 volunteers were asked to rate the download of 14 different web pages at different application data rates (from few kbps to 2 Mbps, more details about this experiment can be found in Section 3.2.4). Eq. A.11 shows the relationship between throughput and MOS, assuming an average web page size of 130 Kbytes [50].

$$MOS = 5 - \frac{578}{1 + \left(\frac{THR + 541.1}{45.98} \right)^2} \quad (\text{A.11})$$

where THR is the throughput in kbps.

Utility function for *FTP*

In [24], Y. Sato *et al* studied user's subjective assessment of communication quality in bulk data transmission. They surveyed 91 users, considering typical file sizes and FTTH-like bandwidths (max. 25 Mbps).

For this work, the maximum throughput was scaled to take into account IEEE 802.11e's scarcer bandwidth. Based on an IEEE 802.11b physical layer, it was defined 1 Mbps as the target throughput for the FTP service, i.e. the one above which the end-user is highly satisfied. Accordingly, the results were scaled by a factor of 25. With these assumptions, the MOS function is expressed as follows:

$$MOS = 6.5 \times THR - 0,54 \quad (\text{A.12})$$

where THR is the throughput measured in Mbps.

A.4 Radio resource management strategies for HCCA

A.4.1 A variant of the Monolithic Shaper-Scheduler for HCCA

Original Monolithic Shaper-Scheduler (MSS)

The reference Monolithic Shaper-Scheduler is described in [54]. It was selected due to its interesting features, since it supports minimum bandwidth and delay guarantees, and at the same time it has optimal latency among all GPS-related schedulers.

The MSS supports both shaped and unshaped traffic streams. If the flow is shaped, a dual-leaky-bucket regulator is utilized as traffic shaper. A dual-leaky-bucket consists of two components: a Sustainable-Bit-Rate (SBR) leaky bucket and a Peak-Bit-Rate (PBR) leaky bucket. The SBR leaky bucket constrains the long-term behavior of the flow (i.e. average throughput), whereas the PBR leaky bucket imposes a bound on the transmission rate during bursts (i.e. peak throughput). Both buckets are defined by their sizes and their token generation rates.

After a packet arrival, the flow is tagged with a start (*eligible*) time and a finish (*deadline*) time. The packet is held in the scheduler until its eligible time expires. The scheduler orders the transmission of eligible packets according to their deadlines, implementing the *Smallest Eligible Finishing time First* (SEFF) packet selection policy. The eligible time and the deadline of the packets of a flow with a guaranteed rate R_i are computed as the start and finish times for transmitting the packet in an ideal network in which the flow has dedicated links of capacity R_i . In particular, the eligible time of a packet is calculated as the maximum between the packet arrival time and the deadline of the previous packet of the flow. The deadline of a packet is computed as the sum between the packet's arrival time and the time it takes to transmit the packet in the ideal system, i.e. $\frac{N_i}{\Gamma_i}$, where N_i represents the length of the k^{th} packet of the flow i and Γ_i is the minimum physical transmission rate (assumed 1 Mbps for IEEE 802.11b networks). In order to check the time instant when the flow becomes eligible, upon the arrival of its k^{th} packet of length N_i at time $t_{i,a}^k$, it receives a new timestamp F_i^k based on the following rule:

A.4. Radio resource management strategies for HCCA | 189

$$F_i^k = \max(t_{i,a}^k, F_i^{k-1}) + \frac{N_i}{R_i} \quad (\text{A.13})$$

The operation of the MSS consists of four basic activities: *test for virtual compliance*, *flow enqueue*, *flow dequeue* and *flow selection*. The test for virtual compliance evaluates the possibility of violating one of the SBR or PBR token bucket profiles. The flow enqueue activity updates the flow timestamp. Both the test for virtual compliance and the flow enqueue activity are executed when a new packet reaches the head of the queue (new packet to an empty queue or after packet transmission from a queue with more packets). The flow dequeue activity checks whether the flow becomes *eligible* or not. If the flow is not eligible, a recalibration of the timestamp takes place in order to avoid an uncontrollable growth of the distance between the timestamp and the real-time reference. Finally, the flow selection procedure identifies the flow of the next packet to be transmitted. Both the flow dequeue activity and the flow selection procedure are performed after a packet transmission. Figures A.3 and A.4 include the flowcharts of these operations, and a more detailed explanation may be found in [54].

The MSS is a *Latency-Rate (LR) server* [59], since it is a modification of the *Virtual Clock* scheduling discipline [60] with the same bandwidth and latency guarantees. An *LR server* is defined as a scheduler whose offered service is at least equal to its reserved rate (over every time interval starting after its latency Θ). Since the MSS is a *Latency-Rate server* with a dual-leaky-bucket regulator, the *dual leaky-bucket-shaped traffic* analysis of [59] can be applied.

Therefore, the maximum delay D_i of a packet from session i in a MSS, whose traffic is shaped by a dual-leaky-bucket, is bounded by the following expression [59]:

$$D_i \leq \left(\frac{\rho_{i,p} - R_i}{\rho_{i,p} - \rho_{i,s}} \right) \left(\frac{B_{i,s}}{R_i} \right) + \Theta_i^{\text{MSS}} - \frac{N_i}{\rho_{i,s}} \quad (\text{A.14})$$

where $\rho_{i,s}$ and $\rho_{i,p}$ are the token rates of the SBR and PBR buckets for the i flow respectively, Θ_i^{MSS} is the latency of the scheduler and $B_{i,s}$ is the *Sustainable Bit Rate* (SBR) bucket size. The latency of the MSS, which is optimal among all GPS-related schedulers, is limited [54] by:

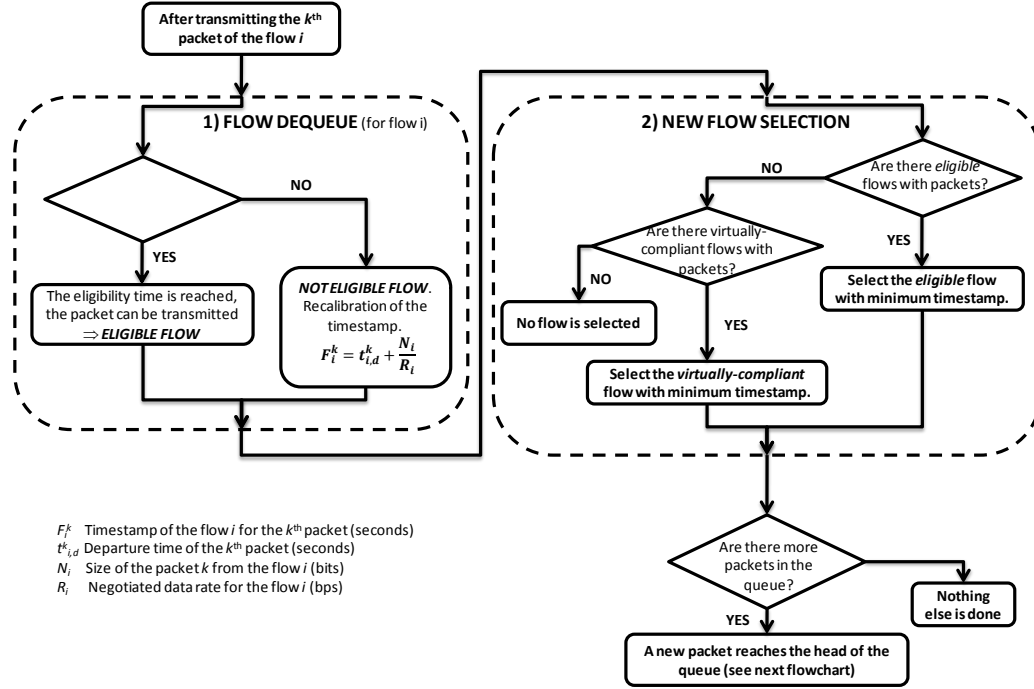


Figure A.3. Flowchart of the flow dequeue activity and the flow selection procedure.

$$\Theta_i^{MSS} \leq \frac{N_i}{R_i} + \frac{N_{max}}{\Gamma_i} \quad (\text{A.15})$$

where N_{max} is the maximum nominal SDU size of all the flows.

As it is derived from Eq. A.14, the delay bound is influenced by the size of the SBR bucket. Francini *et al* [54] provide the minimum *Sustainable Bit Rate* (SBR) bucket size, $B_{i,s}$, and the minimum *Peak Bit Rate* (PBR) bucket size, $B_{i,p}$, for which the MSS guarantee no overflow (expressed in bits instead of seconds):

$$B_{i,s}^{min} = \left(2N_i \left(\frac{1}{R_i} - \frac{1}{\Gamma_i} \right) + \frac{N_{max}}{\Gamma_i} \right) \cdot \rho_{i,s} \quad (\text{A.16})$$

$$B_{i,p}^{min} = \left(2N_i \left(\frac{1}{\rho_{i,p}} - \frac{1}{\Gamma_i} \right) + \frac{R_i}{\rho_{i,p}} \cdot \frac{\Gamma_i - \rho_{i,p}}{\Gamma_i - R_i} \cdot \frac{N_{max}}{\Gamma_i} \right) \cdot \rho_{i,p} \quad (\text{A.17})$$

A.4. Radio resource management strategies for HCCA | 191

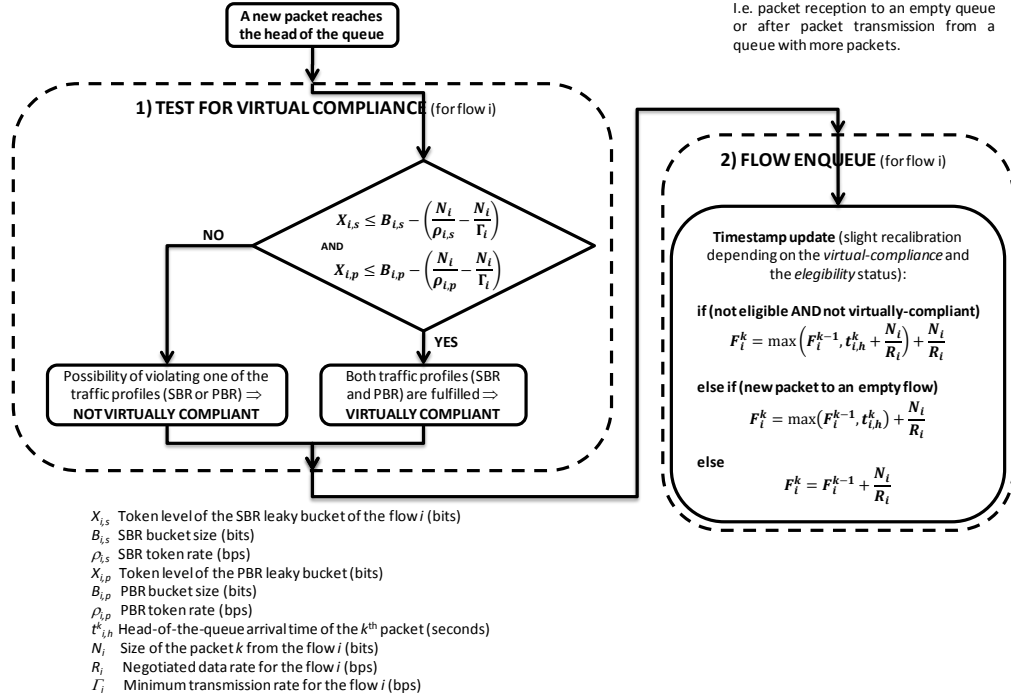


Figure A.4. Flowchart of the test for virtual compliance and the flow enqueue activity.

Since, in the MSS, the guaranteed service rate R_i is equal to the SBR token bucket rate $\rho_{i,s}$, and the latency Θ_i^{MSS} is bounded by Eq. A.15, assuming the minimum SBR bucket size (Eq. A.16) the expression of the delay bound (Eq. A.14) can be simplified as:

$$D_i \leq 2N_i \left(\frac{1}{R_i} - \frac{1}{\Gamma_i} \right) + \frac{2 \times N_{\max}}{\Gamma_i} \quad (\text{A.18})$$

Theorem 1: The MSS scheduler satisfies the requirement imposed by the 802.11e standard for HCCA schedulers (Eq. A.1), assigning Δ_i (*maximum service interval or delay bound*, depending on which is specified in TSPEC) a value higher than the MSS latency (Θ_i^{MSS}).

Proof: From the definition of a *latency-rate* server [59], the total service W_i^{MSS} (in transmitted bits) provided by the MSS to the session i during a busy period (started at τ)

after time t can be expressed as:

$$W_i^{MSS}(\tau, t) \geq \max\left(0, R_i(t - \tau - \Theta_i^{MSS})\right) \quad (\text{A.19})$$

Considering the physical transmission rate Γ_i in the previous equation, the total service can be calculated in time units. Assuming, without loss of generality, that $t > \tau + \Theta_i^{MSS}$ and that the physical transmission rate Γ_i incorporates the MAC control frames and interframe spaces, it can be derived that:

$$W_i^{MSS}(\tau, t) \geq \frac{R_i}{\Gamma_i}(t - \tau - \Theta_i^{MSS}) = \frac{N_i/\Gamma_i}{N_i/R_i}(t - \tau - \Theta_i^{MSS}) = \frac{t - \tau - \Theta_i^{MSS}}{\frac{N_i}{R_i}} \cdot tx(N_i, \Gamma_i) \quad (\text{A.20})$$

This expression is the same that Eq. A.1, identifying t as t_l , τ as t_2 and specifying $\Delta_i \geq \Theta_i^{MSS}$, so it demonstrates that the MSS verifies the requirement for HCCA schedulers.

Adaption of MSS scheduler to 802.11e HCCA

In the original Monolithic Shaper-Scheduler, the server maintains all the required information to select the proper flow after a packet transmission. This is possible because the server enqueues all the packets from all the traffic flows before selecting which flow shall transmit.

However, in IEEE 802.11e networks, the scheduler (which is collocated at the QAP) only enqueues packets sent *to* QSTAs, but not *from* QSTAs. The packets of QSTAs originated flows are enqueued in the QSTAs themselves, so the scheduler has no information about these flows *a priori*.

For each flow, the required data are:

- Enqueued packet availability
- Its timestamp
- Its eligibility flag

A.4. Radio resource management strategies for HCCA | 193

- Its virtual-compliance status

This information can be easily computed and maintained at the scheduler for downlink flows, since their packets are received and forwarded by the QAP. However, MSS has no *a priori* information the packets generated by the QSTAs, so the information for uplink flows shall be communicated to the access point.

In order to know if there are packets waiting to be transmitted, the optional queue size piggybacking mechanism (Queue Size subfield present in the QoS data frames [2]) can be used. The remaining information shall be added to the MAC header of the 802.11e data and/or acknowledgments frames. Since the required information takes only few bytes (depending on the required precision for timestamps), the introduced overhead is negligible and therefore it can be ignored in the simulations.

With this information for both uplink and downlink flows, the QAP is able to operate. If the selected flow is downlink, the QAP will send its packets directly. However, if it is an uplink flow, then the QAP will poll the QSTA before packet transmission. It could be noticed that, since all the flows (uplink or downlink) are treated similarly in the selection process, the intrinsic unfairness of 802.11 networks between uplink and downlink streams [61] is resolved.

Additionally, in Appendix B it is calculated the extra delay introduced by the IEEE 802.11e MAC layer due to headers and contention procedures, which is negligible (less than one millisecond). However, it could be added to the result of A.18 in order to have a more accurate estimation of the delay bound.

A.4.2 Resource management for heterogeneous traffic mix based on QoE: the Territory Method

In this section a dynamic resource allocation algorithm –the *territory method*– is presented. Here the objective is to improve the resource utilization while maintaining end-user satisfaction. In addition, it aims at prioritizing real-time services and at improving the performance of TCP applications.

This method shares the common pool of transmission opportunities (TXOPs) into different territories (see Figure A.5), one per each TC. Each territory has a customizable *minimum capacity (MC)*, which can be non null to avoid starvation. A higher *reserved capacity (RC)* can be harvested by booking from a pool of the spare resources. The RC fluctuates according to the end-user MOS, as shown in Figure A.6. This resource allocation algorithm was designed so that the average MOS of certain TC (average of MOS values from the different flows belonging to that TC) is expected to stay between an *upper MOS threshold (UMT)* and a *lower MOS threshold (LMT)*. If the estimated MOS is smaller than the LMT, the end-user perception is worse than required and RC will increase in order to keep the MOS within the required range. On the contrary, if it is above the UMT, the end-user experience is better than required and hence part of the RC can be returned. Both MOS measurements and territory updates are performed at a customizable *MOS Filtering Period (MFP)*. At the AP, a scheduler (the Monolithic Shaper Scheduler, MSS [54]) arbitrates the TCs according to these territories. First, the

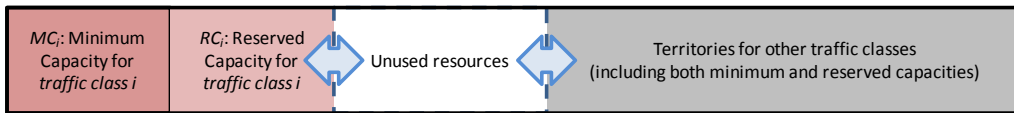


Figure A.5. Example of territories definition for a generic traffic class.

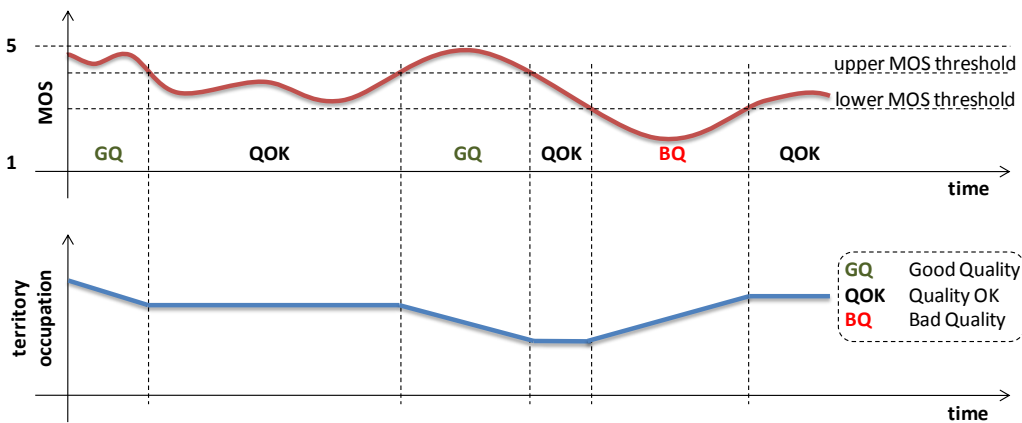


Figure A.6. Example of the relationship between subjective quality and the territory evolution for a generic traffic class.

A.4. Radio resource management strategies for HCCA | 195

territory method checks against a round-robin list which TC owns the current TXOP. Then, the scheduler selects a flow belonging to that TC for packet transmission. As long as the scheduler enforces a fair queueing, the actual scheduling policy has a minor impact on the results.

This method utilizes a preemptive resource sharing policy: TCs are sorted by priority, and a higher priority TC could preempt the RC of a lower priority TC (although without eating into its MC). Accordingly, real-time services (*conversational* and *streaming*) are favored at the expenses of non real-time services (*interactive* and *background*). Its flowchart is exposed in Section 0 (see Figure A.7).

The algorithm for updating the territory should take into account the different dynamics of the various TCs (e.g., web traffic is bursty and intermittent, hence requiring fast territory modifications, whereas FTP is continuous, which calls for smoother variations). For this reason, the following decisions were taken.

First, an exponential filtering is used when measuring a connection's average MOS, so as to avoid oscillations in the territory due to sheer instantaneous variations in the MOS. The averaging process is shown in Eq. A.21.

$$AVGMOS_i[k] = \alpha \cdot AVGMOS_i[k-1] + (1 - \alpha) \cdot MOS_i[k] \quad (\text{A.21})$$

where $AVGMOS_i[k]$ is the average MOS of the traffic class i for the k -th MOS filtering period, α is the averaging weight and $MOS_i[k]$ is the MOS sample of the k -th period.

In addition, each TC can select its own averaging period depending on its traffic characteristics. For example, services with stable traffic generation (e.g. FTP) could have larger periods than services with bursty traffic (e.g. web browsing). In the latter case the traffic pattern varies more quickly, and long periods could make territories not to react properly to those changes.

Furthermore, the epochs for adjusting the territories can be customized. Different *Territory Update Periods* (TUP) are considered, one per TC, with the objective of considering their different properties. Several MOS samples should be considered for resizing the territories: too few samples might lead to fluctuations in the territories due to

MOS peaks, hence leading to unstable performance of the services; too many samples might lead to unresponsiveness, again causing a suboptimal performance. Five MOS samples were selected for averaging by default, i.e. the *territory update periods* are equal to 5s for all TCs excluding interactive, whose period is 1s.

Another critical issue regards how to increment/decrement the territories when LMT/UMT violations are detected. Again, responsiveness has to be traded for stability. Selecting the same step for both increasing and decreasing e.g. (e.g. 1 out of 40 TXOPs), leads to unresponsiveness against e.g. a rapid traffic growth. A better result can be obtained by keeping the *decrease step* (DS) and the *increase step* (IS) different. The IS was designed for being proportional to the distance to the *good quality* range (MOS between LMT and UMT), i.e. it is larger if the MOS is very far from the requirements and smaller if it is near the UMT.

Eq. A.22 illustrates how IS is calculated for a territory update period. The maximum step has been set to 20% of the resources. For the DS a fixed value of 2.5% (1 TXOP out of 40) was chosen, to avoid that a drastic decrease leads to performance degradation.

$$IS = \left\lceil \max_step \times \frac{\text{current_MOS_diff}}{\max_MOS_diff} \right\rceil \quad (\text{A.22})$$

where

$$\max_MOS_diff = \left(\frac{UMT + LMT}{2} - 1 \right)$$

$$\text{current_MOS_diff} = \left(\frac{UMT + LMT}{2} - AVGMOS \right)$$

A preliminary study about PFM_i , PAT_i and α_i parameters is included in Appendix D to justify the chosen values.

A.4.3 Resource sharing policies for HCCA

Different policies for resource sharing in IEEE 802.11e networks were designed based on the territory method framework. These policies are described next.

Preemptive. This is the original policy in the territory method depicted in the previous subsection. Its objective is to ensure that the performance of real-time services is not affected by non real-time traffic. For that purpose, higher priority traffic classes can preempt resources from the reserved capacity of lower priority TCs. Therefore, conversational services can always utilize all the transmission opportunities except those belonging to the minimum capacity of other traffic classes.

Figure A.7 describes the flowchart of this approach. For each TC i , the MOS is computed by averaging the MOS of the single constitutive flows, which are sampled at fixed intervals.

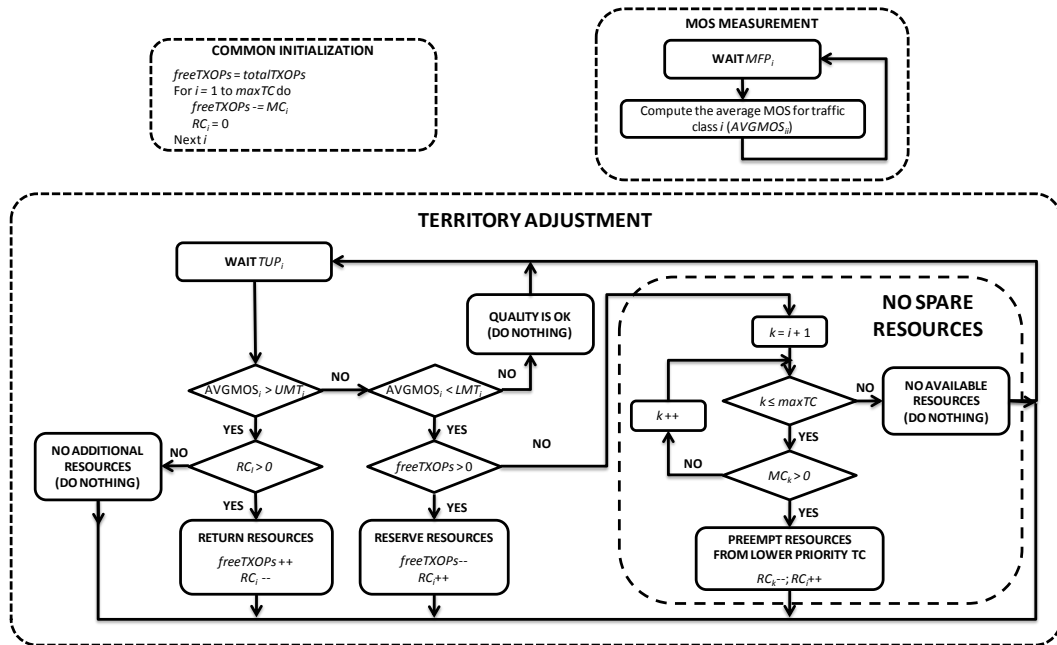


Figure A.7. Flowchart of the territory method using the preemptive policy.

Non-preemptive. This approach assumes that no traffic class has priority over other TCs. Therefore, no spare resources can be pre-empted once another TC has got hold of them. However, a certain service is guaranteed for each traffic class by means of its minimum capacity, which can be customized by the administrator.

This policy allocates resources based on the current needs of each TC. If the global MOS of a TC (average MOS of all flows belonging to that traffic class) is below its LMT, its territory will grow if there are spare resources. Similarly, if the average MOS is above the UMT part of the reserved resources will be released.

Thus, it is a *first come, first served* policy without priorities between traffic classes. Its flowchart is similar to Figure A.7 but substituting the “*no spare resources*” box with Figure A.8.

MOS-driven Fairness. Neither of the previous schemes can ensure inter-TC fairness. A *MOS-driven Fairness* policy behaves similarly to the *Non-preemptive* policy as long as the MOS of all the different TCs is good. However, it will try to equalize the MOS among TCs when there are not enough resources. In that situation, the TCs with the lowest MOS can preempt some RC from the TCs with the highest quality. As for the previous policy, the “*no spare resources*” box has to be replaced by Figure A.9 in the flowchart.

Revenue Improvement. Similarly to the *MOS-driven Fairness* policy, the *Revenue Improvement* approach only operates when the MOS of a TC is below its threshold. In that case, TCs are ranked by Revenue Efficiency (RE), i.e. income per resource unit (Eq. A.23). The latter is computed by dividing the overall MOS of TC i by the amount of resources used to support it. The result is weighted by a TC-specific revenue weight W_i , which can be selected by a network administrator to reflect his revenue policies. A higher revenue TC can preempt spare resources from a lower revenue TC (see Figure A.10).

$$RE_i = W_i \cdot \left(\frac{AVGMOS_{i-1}}{RC_i} \right) \quad (\text{A.23})$$

A.4. Radio resource management strategies for HCCA | 199

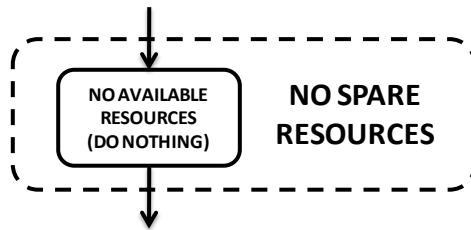


Figure A.8. Flowchart for the *non-preemptive* policy.

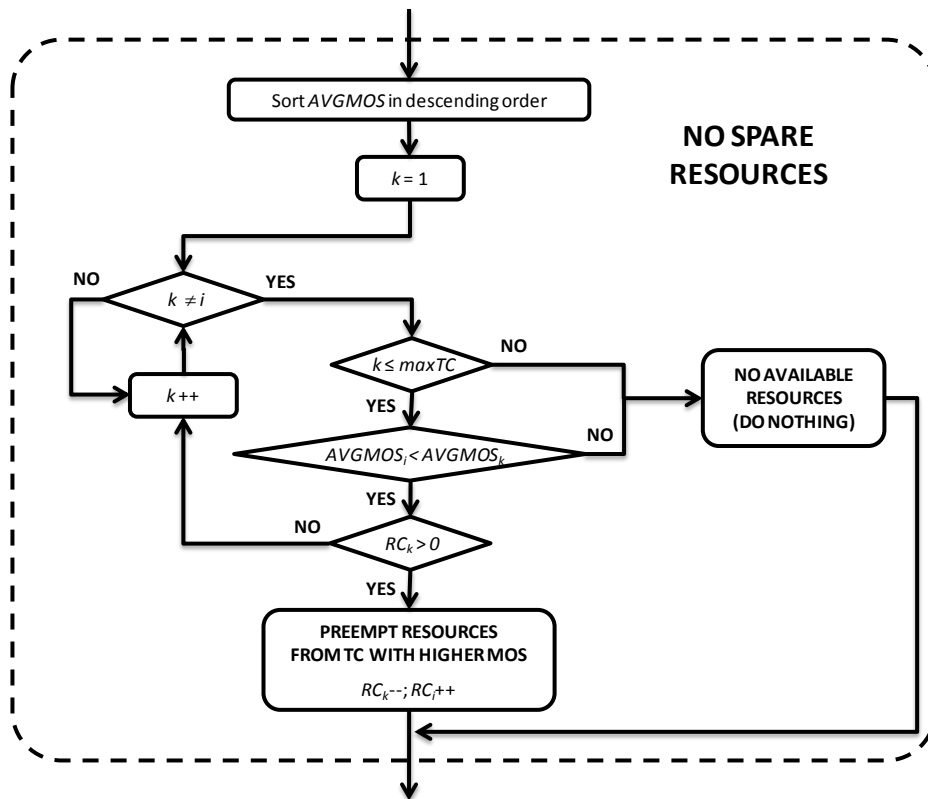


Figure A.9. Flowchart for the *MOS-driven fairness* policy.

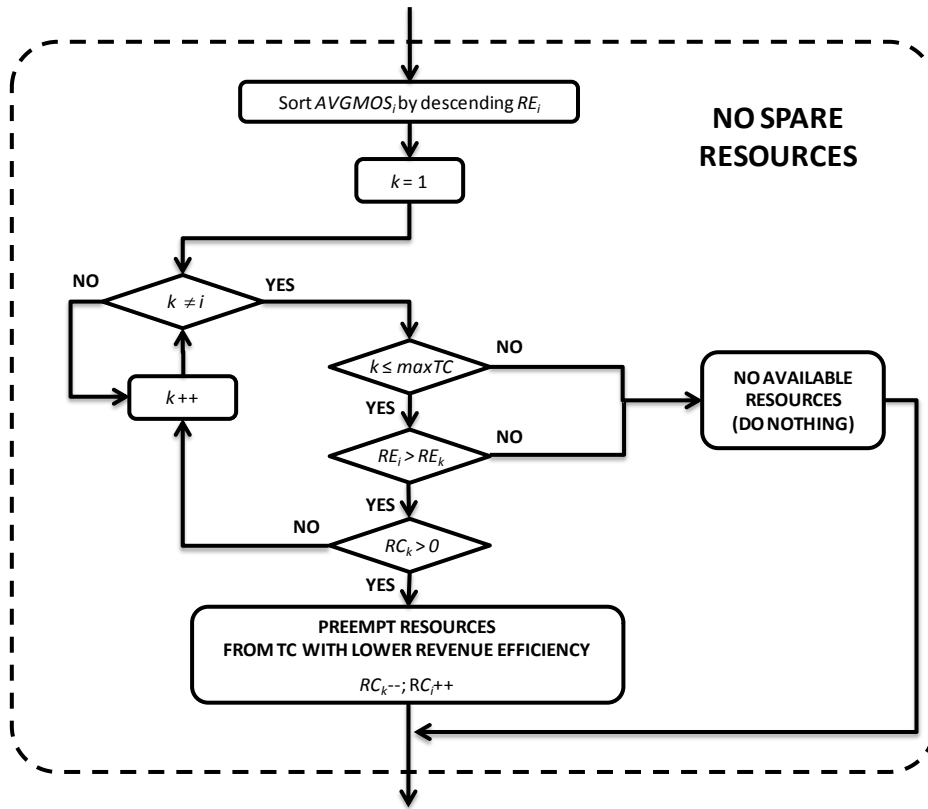


Figure A.10. Flowchart for the *revenue improvement* policy.

A.4.4 Main results of the proposed solutions for resource allocation

This section includes a comparison between the different approaches that have been proposed for resource management in the wireless medium. For that purpose, the main results obtained in the different simulated experiments –with *medium load*, *high load* and *very high load*– are presented. These results are shown for the following schemes:

- The MSS scheduler (without the Territory Method)

A.4. Radio resource management strategies for HCCA | 201

- The MSS scheduler used in conjunction with the Territory Method and one of the next resource sharing policies:
 - Preemptive
 - Non-preemptive
 - MOS-driven fairness
 - Revenue improvement

To summarize the performance evaluation, the following metrics are exposed:

- Number of satisfied VoIP users, i.e. with a quality higher than $UIM_i = 3.5$ (in a MOS scale)
- Number of VoIP users for which all the users –of any service– are satisfied (for a given traffic mix)
- Maximum revenue obtained (for the chosen traffic load and for any number of voice conversations)
- Minimum value of the Jain’s fairness index (metric $x = \overline{MOS}_i - 1$) for the simulated range of VoIP traffic (from 0 to 30 conversations)

Figure A.11 includes these indicators for the experiments that use *medium load*. In this case, since the load is not too high, the results for the different approaches are similar. The number of satisfied VoIP users that are supported by the system –with *medium load*– is 16 for the MSS discipline (without the Territory Method) and slightly higher when using territories –with a maximum of 20 when using the preemptive policy, scheme that was designed to prioritize conversational services–.

Regarding the maximum supported load for which all the users are satisfied, it is approximately 16 ~ 17 voice conversations for all the schemes. Similarly, the incomes for the proposed policies are in the same order of magnitude, ranging the maximum values from 85 to 100 revenue units.

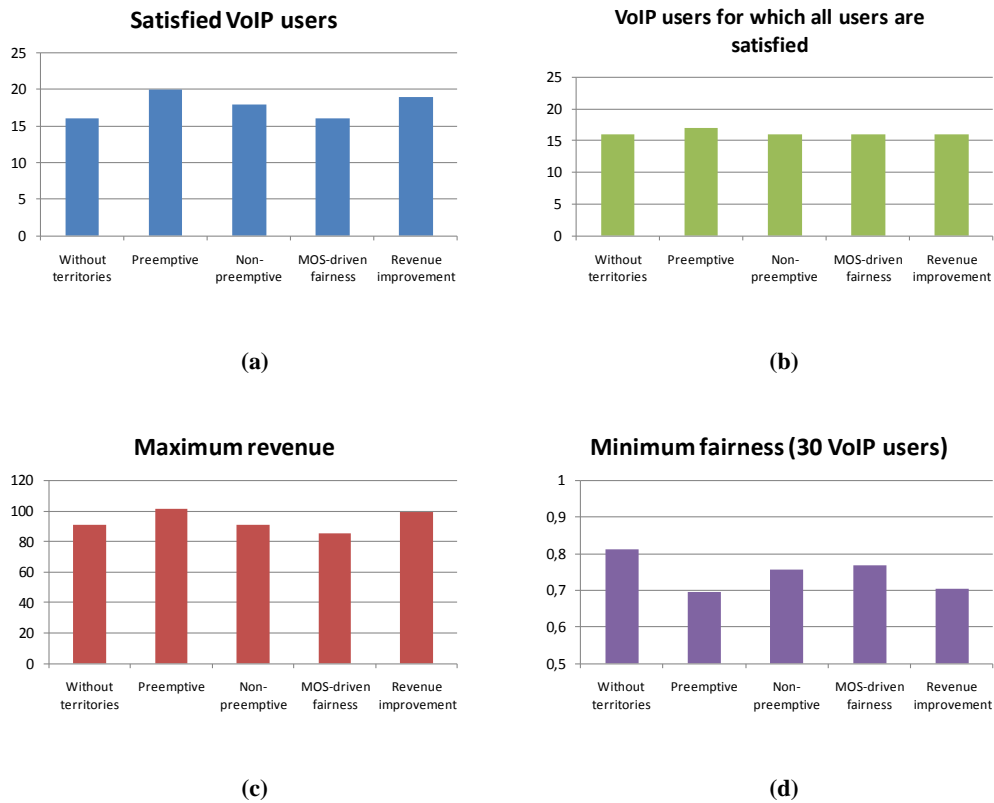


Figure A.11. Comparison among the different RRM approaches in the case with *medium load*.

The main difference between these solutions is how the system reacts when it is high loaded. The fairness index in the point with highest load (30 VoIP users) reveals which solution is the most egalitarian for the different services. As it can be observed in the subfigure (d), the MSS scheduler is the fairest one, whereas the preemptive and revenue improvement policies create the highest differentiation among services –since they were designed with this objective–.

Figure A.12 shows the comparison in the case with *high load*. As the load increases, the differences between the proposals also do. For this reason, the MSS scheduler only supports 5 conversations with the required quality, increasing to 20 when using the preemptive policy. As it was commented in Section 0, the revenue improvement policy

A.4. Radio resource management strategies for HCCA | 203

also prioritizes the VoIP service (the most efficient in terms of incomes), so it obtains similar results. The other policies achieve an acceptable QoE for 10 ~ 12 voice calls. Thus, using any of the proposed policies for the Territory Method, the capacity for VoIP service is significantly higher than without using this mechanism.

With respect to the maximum load that can be supported—all users for the different services satisfied—, it corresponds to *high load* with 5 voice calls for MSS and with 10 ~ 12 calls for the remaining schemes. As before, the Territory Method improves the resource sharing due to the introduction of QoE as a criterion so that no unnecessary resources are assigned—which do not increase the end-user’s experienced quality—.

The maximum revenue is achieved using the preemptive and revenue improvement

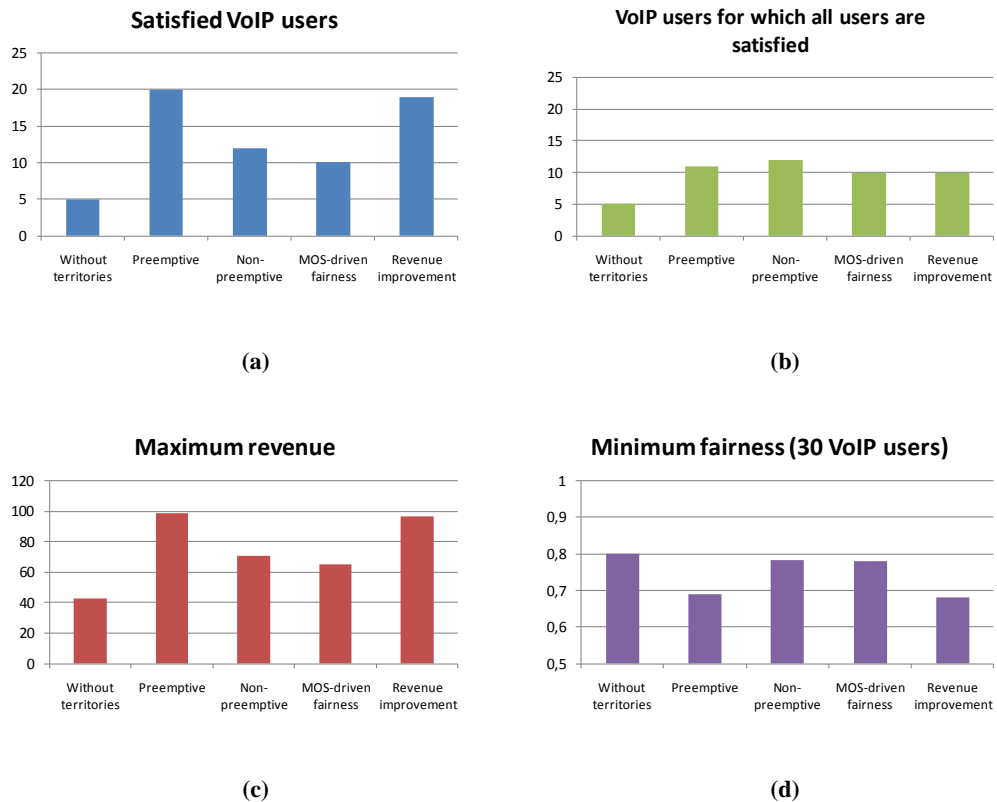


Figure A.12. Comparison among the different RRM approaches in the case with *high load*.

policies –expectable due to their definitions– with approximately 100 units. This value is reduced to around 70 units for the other Territory Method policies. The proposal with lowest incomes –40 units– is the MSS scheduler, hence the Territory Method also increases the network administrator’s benefits.

In terms of fairness, MSS is the best option since it does not split the resources for the different traffic classes –and this is the fairest discipline within the latency-rate schedulers [54]–, achieving an index equal to 0.8. The non-preemptive and MOS-driven fairness policies obtain a similar but slightly lower value –the latter was designed with this objective–. As it could be expected, the preemptive and revenue improvement policies are the most unfairness approaches, with an index smaller than 0.7.

Figure A.13 synthesizes the results obtained for *very high load*. This scenario justified the Territory Method proposal, since no voice call could be carried with the required quality by only using the MSS scheduler (as it can be observed in the subfigure (a)). However, the use of the territories and the prioritization of VoIP traffic permits 19 calls with enough quality by using the preemptive resource sharing approach –its performance is practically independent of the remaining services using this policy–, 17 when increasing the revenue and 3 ~ 4 for the other policies.

In this case it is not possible to satisfy all the users with any of these solutions, even without conversations, due to the high traffic load generated in the experiment.

Regarding the network administrator’s profit, the differences between the alternatives increase with *very high load*. The revenue is around 90 units for both the preemptive and revenue improvement disciplines –maximum benefit among the different options–, decreasing to 53, 33 and 16 units for the non-preemptive, MOS-driven fairness and without the Territory Method schemes respectively. Again, all the defined policies for the Territory Method improve the results compared to not using territories.

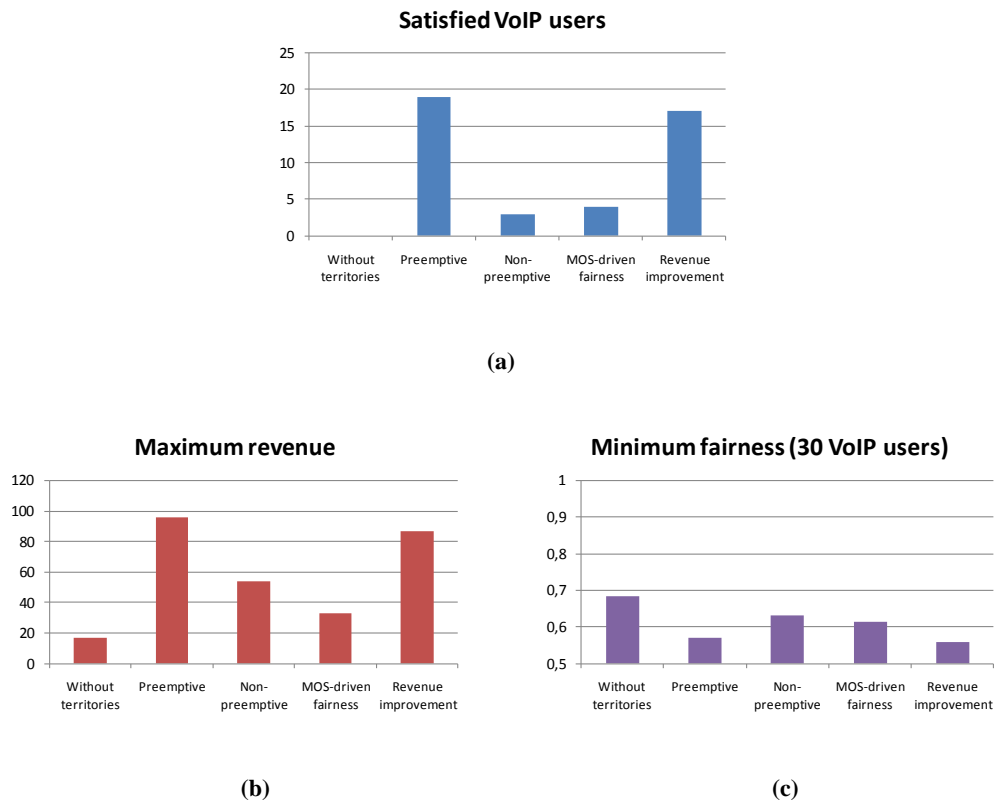


Figure A.13. Comparison among the different RRM approaches in the case with *high load*.

A.5 Improved TCP over HCCA

In the previous proposals, TCP flows were transmitted using the HCCA mechanism. The usage of HCCA –instead of EDCA– for these kinds of flows has several advantages:

- First, EDCA cannot guarantee QoS requirements, which might be appropriate for some TCP services, e.g. progressive download video like YouTube.

- Second, the intrinsic unfairness between uplink and downlink flows in IEEE 802.11 networks [61] also exists when using EDCA [85] [86], but it disappears with HCCA. The reason for this unfairness is that the QAP competes with the same success probability than other QSTAs (assuming the same Access Category, which depends on the particular flow) in the distributed channel access (DCF or EDCA). Since the QAP has (in general) a higher number of flows –all the downlink flows–, they have a smaller transmission probability. However, using HCCA each flow –independently of its direction and assuming a fair scheduling discipline– has the same probability of being selected.
- Finally, the usage of a HCCA packet scheduler allows the designer to develop more complex algorithms and mechanisms to achieve the pursued goals. This is the case of the TCP enhancement explained in this section.

Since multiple TCP connections can share the wireless medium, a coordinated congestion control mechanism could be devised. Hence, since the QAP has some knowledge about the TCP connections status, their performance could be improved by considering their end-users' experience.

However, this approach has a critical problem: the congestion control does not only depend on the last hop (the wireless network) but on the end-to-end path. Therefore, since the problems in the wired part are unknown, the coordination for managing multiple TCP flows can be very complex –or even unfeasible–.

For that reason, a different approach has been considered. It makes both environments –wired and wireless– independent by means of a split TCP connection [87]. Thus, this scheme utilizes a *Performance Enhanced Proxy* (PEP) in an intermediate point between the source and the destination. In this way, the original end-to-end connection is split into two new TCP connections –one between the source and the PEP and another between the PEP and the destination–.

In a heterogeneous network, where a QAP is connected to a wired network and to a wireless network, the intuitive point for the proxy is the QAP since it connects two mediums with different properties: the wire –high data rate, low loss probability and high

delay– and the wireless –usually lower data rate and high loss probability–. Thus, the scenario would change from Figure A.14 to Figure A.15.

After this split, each individual connection can be adjusted to the characteristics of its environment without being influenced by the other medium.

Removal of TCP in the wireless medium

The previous approach was proposed and analyzed in [87]. Based on this scheme, one of the contributions of this thesis consists in removing TCP in the wireless medium, as it is explained next.

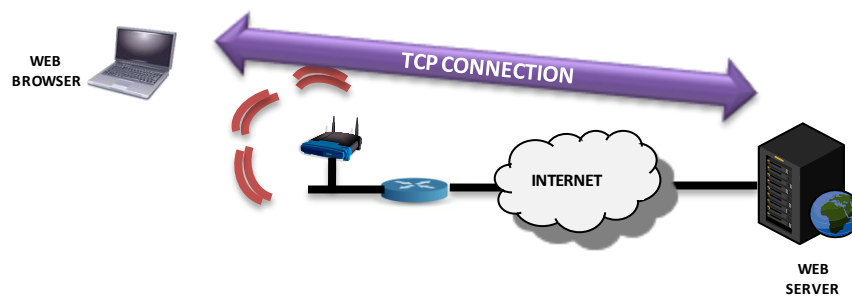


Figure A.14. Scenario with a *normal TCP* connection.

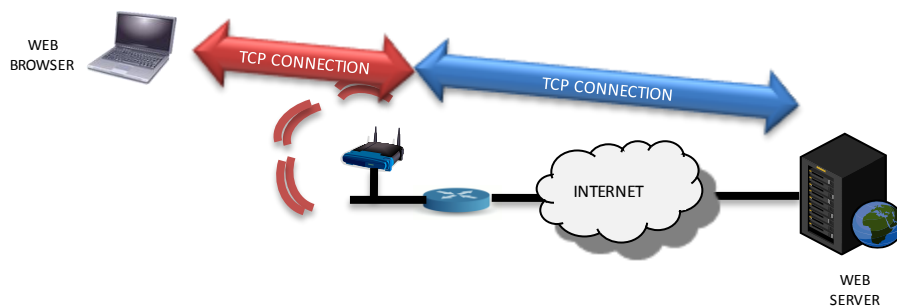


Figure A.15. Scenario with a *split TCP* connection.

Using a split TCP connection, the following question arises: is it necessary to have TCP in the wireless interface? To answer this question, it is checked whether the main functionalities of TCP are needed:

- **Connection control.** Connection establishment and finish is done using a 3-way handshake. A similar function is already done in the MAC layer in the admission control.
- **Multiplexing.** By means of ports, an equipment can receive messages sent to different applications. While this is a mandatory feature, other simpler protocols –e.g. UDP– also includes it.
- **Error control.** This function is already done in the IEEE 802.11 MAC layer. In addition, the scheduler can take into account if retransmissions are needed to remove any residual error probability. Therefore TCP acknowledgment messages are not needed and can be substituted by MAC acknowledgments.
- **Flow control.** This feature is mandatory to avoid the receiver not being able to process incoming data. However, it only requires that packet headers include the receiver’s advertised window.
- **Congestion control.** TCP is able to modify its data rate so it adjusts to the capacity of the end-to-end path, avoiding intermediate routers being congested. However, in the case of a WLAN with HCCA, the scheduler controls the data rate for each flow. Moreover, there are no intermediate routers between the QAP and the QSTAs. Thus, this functionality is not needed as long as there is a scheduler in the wireless part.

For these reasons, a simpler transport protocol would be enough –just including the used ports and the receiver’s advertised window–. Hence, a protocol similar to UDP but including the advertised window could be used.

This modified version of split TCP is designated *split TCP & UDP*, even though it uses a modified header with respect to the original UDP protocol. This scheme is presented in Figure A.16.

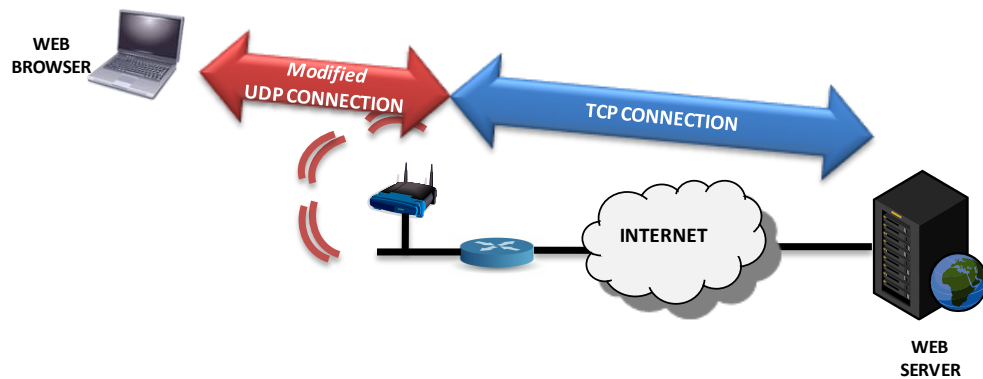


Figure A.16. Scenario with a *split TCP & UDP* connection.

A.5.1 Main results for improved TCP over HCCA

These TCP variants –split TCP and *split TCP & UDP*– have been implemented in NS-2, based (with significant changes) in the implementation of a Performance Enhancing Proxy described in [90]. As before, *TCPLinux* [72] has been used to reflect the real behavior of TCP in the simulations, with *New Reno* [73] as the selected TCP variant.

In order to discard other effects due to the traffic generation pattern, FTP has been chosen as the TCP service. UDP has been utilized –since it is not affected by changes in TCP– to create, in some cases, extra load in the network. In addition, the performance evaluation has focused on data rate since, due to MOS saturation on the utility function for FTP (see Section A.3), there are many cases with the highest mark and therefore their differences –in terms of QoE– could not be compared.

The first experiment aims at quantifying the achieved gain when only one user is transmitting in the wireless medium. The simulation campaign has included different file sizes (50KB, 500KB, 5MB and infinite –always transmitting–) and different *Round-Trip Times* (the one-way delay is 2, 50, 100 and 500 ms).

The results show that *split TCP & UDP* noticeably improves data transmission

–approximately 90%– when the TCP congestion control mechanism in the wired link is not limiting (i.e. very high RTT or very short files). Figure A.17 is provided as an example (please notice that *TCP in wired link* is outside the plot range); please refer to Section 7.2.2 for more details.

The second experiment considers an extra load of 3Mbps (carried over UDP in order not to interact with TCP sources). This extra load reduces the bandwidth for the FTP user (compare Figures 7.4 - 7.7 to Figures 7.8 - 7.11). However, the gain is similar to the previous case (see Figure A.18).

The performance of this solution has also been obtained with several FTP users (2, 4, 6, 8 and 10) and with several frame error probabilities (0, 10, 20 and 30%). In these simulations, the QSTAs were always receiving FTP traffic.

Figures A.19 to A.22 show the performance gain –in terms of throughput– for these experiments. As before, the proposed solution achieves a considerable gain as long as the

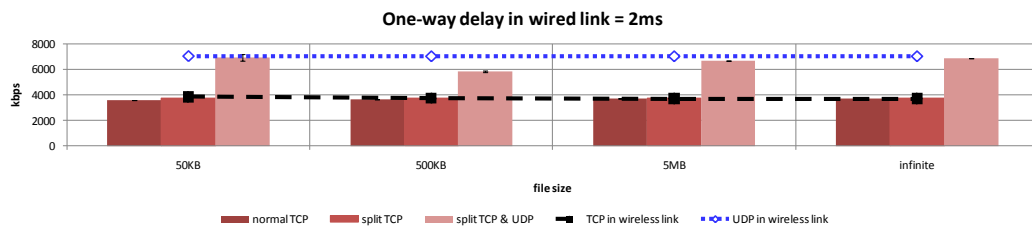


Figure A.17. Results for 1 FTP download and one-way delay of 2 ms (wired link).

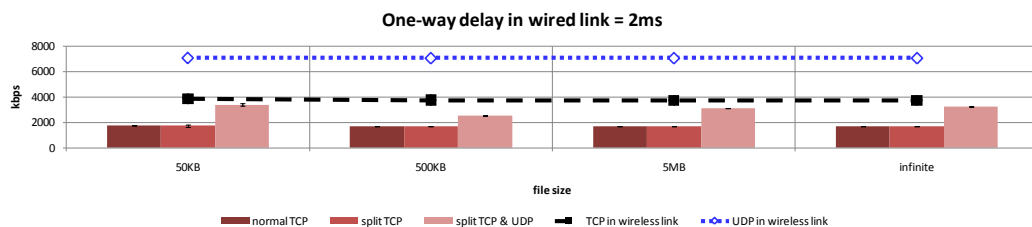


Figure A.18. Results for 1 FTP download with 3 Mbps of UDP traffic and one-way delay of 2 ms (wired link).

wired part is not restricting the TCP throughput (for that reason, the improvement is lower for very high RTT, e.g. 2 x 500 ms). In that case, it can be observed that higher error probabilities imply higher gains –increasing from approximately 65% to 80%–, thanks to the independence between wireless and wired links when using split TCP connections.

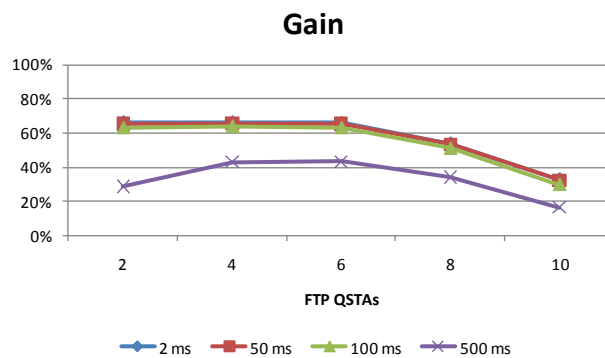


Figure A.19. Gain for several FTP users and 0% frame error rate.

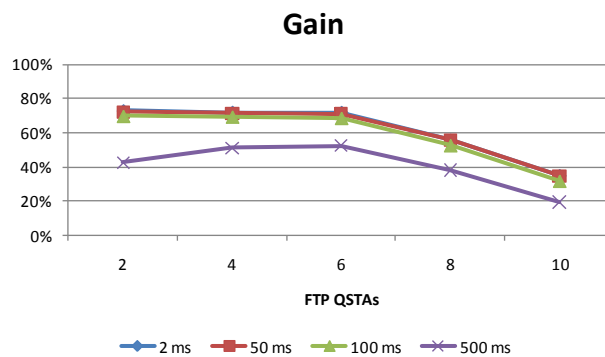


Figure A.20. Gain for several FTP users and 10% frame error rate.

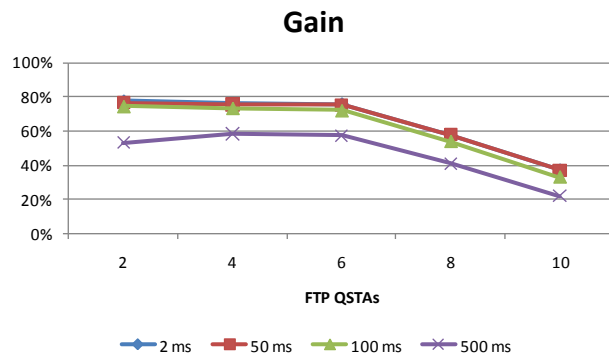


Figure A.21. Gain for several FTP users and 20% frame error rate.

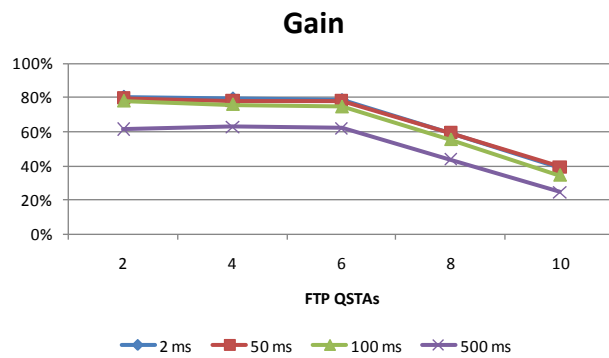


Figure A.22. Gain for several FTP users and 30% frame error rate.

A.6 Conclusions

Please refer to Chapter 9 for reading the main conclusions extracted from the work of this thesis.

Apéndice B

Retardo extra introducido por la capa MAC de IEEE 802.11e

En el caso de que el mecanismo de acceso HCCA esté activo, para el envío de datos en el enlace descendente es necesario que el coordinador híbrido primero compruebe que el medio está inactivo durante un período de tiempo definido (SIFS). La trama enviada contiene las cabeceras de las capas física y MAC además de los datos. Tras la recepción, la QSTA responde con una trama de confirmación (nuevamente tras un período SIFS). Por tanto, el tiempo total que transcurre durante este procedimiento se puede expresar según la Ecuación B.24.

$$t_{DL\ transfer} = SIFS + t_{DATA} + SIFS + t_{ACK} = SIFS + \left(\frac{h_{PHY} + h_{MAC} + N_i}{\Gamma_i} \right) + SIFS + \left(\frac{h_{PHY} + h_{ACK}}{\Gamma_{BASIC}} \right) \quad (\text{B.24})$$

donde Γ_i es la tasa binaria de datos de la capa física (11 Mbps en el caso de redes IEEE 802.11b) y Γ_{BASIC} es la tasa básica utilizada principalmente para señalización (1 Mbps para redes 802.11b).

214 | Retardo extra introducido por la capa MAC de IEEE 802.11e

A partir de esta expresión es fácil obtener el retardo extra debido a los procedimientos de la capa MAC y a las cabeceras, simplemente restando el tiempo requerido para el envío de los datos $\left(\frac{N_i}{\Gamma_i}\right)$. Suponiendo los parámetros especificados en la Tabla 4.1, el valor de este retardo viene dado por la Ecuación B.25.

$$\begin{aligned} t_{\text{retardo extra para transmisiones descendentes}} &= SIFS + \left(\frac{h_{PHY} + h_{MAC}}{\Gamma_i}\right) + SIFS + \\ \left(\frac{h_{PHY} + h_{ACK}}{\Gamma_{BASIC}}\right) &= 10^{-5} + \left(\frac{192+288}{11 \times 10^6}\right) + 10^{-5} + \left(\frac{192+112}{10^6}\right) = 367.6 \mu s \end{aligned} \quad (\text{B.25})$$

Si los datos se envían en el sentido ascendente, el QAP primero ha de enviar una trama *CF-Poll* a la QSTA origen, tras esperar un período PIFS con el medio inalámbrico inactivo. Si sólo hay una única trama de datos a enviar en la TXOP asignada (caso peor en términos de sobrecarga), el retardo introducido por la capa MAC de 802.11e se calcula mediante la Ecuación B.26.

$$\begin{aligned} t_{\text{retardo extra para transmisiones ascendentes}} &= PIFS + \left(\frac{h_{PHY} + h_{POLL}}{\Gamma_{BASIC}}\right) + SIFS + \\ \left(\frac{h_{PHY} + h_{MAC}}{\Gamma_i}\right) + SIFS + \left(\frac{h_{PHY} + h_{ACK}}{\Gamma_{BASIC}}\right) &= 3 \times 10^{-5} + \left(\frac{192+288}{10^6}\right) + 10^{-5} + \left(\frac{192+288}{11 \times 10^6}\right) + \\ 10^{-5} + \left(\frac{192+112}{10^6}\right) &= 877.6 \mu s \end{aligned} \quad (\text{B.26})$$

De esta manera, independientemente del sentido de la transmisión, el retardo extra debido a la capa MAC de las redes 802.11e siempre es inferior a un milisegundo, siendo por tanto despreciable frente al retardo introducido por el procedimiento de planificación de paquetes.

Apéndice C

Comprobación de la implementación de la variante de MSS para HCCA

Los resultados incluidos en este apéndice tienen como objetivo comprobar el correcto funcionamiento de la implementación de la variante del planificador MSS diseñada para HCCA en el simulador NS-2. Si bien no es posible verificar su comportamiento mediante simulación en todos los posibles casos, sí se puede evidenciar el cumplimiento de sus características teóricas (según el modelo analítico presentado en la Sección 4.1) en algunos casos relevantes. Las condiciones experimentales son las descritas en la Sección 4.3.1.

En concreto, se han realizado pruebas considerando sólo tráfico VoIP y una mezcla de tráfico VoIP con una sesión FTP (nótese que con un control de admisión como el usado por el planificador de referencia no se permitirían dos flujos FTP debido a sus requisitos de velocidad de transmisión). Para esta comprobación, las fuentes VoIP han generado tráfico CBR (sin detección de silencios) lo que corresponde a suponer las condiciones de peor caso.

216 | Comprobación de la implementación de la variante de MSS para HCCA

Con el tráfico generado únicamente por fuentes VoIP, la cota del retardo se puede calcular utilizando la Ecuación 4.6 y considerando $N_1 = N_{m\acute{a}x} = 280$ bytes, $R_1 = 74.7$ Kbps y $\Gamma_1 = 1$ Mbps. Con estos valores, el retardo máximo está limitado a 60 ms.

En el escenario con mezcla de tráfico, los valores utilizados han sido $N_1 = 280$ bytes, $N_2 = N_{m\acute{a}x} = 1500$ bytes, $R_1 = 74.7$ Kbps, $R_2 = 660$ Kbps y $\Gamma_1 = \Gamma_2 = 1$ Mbps, lo que implica unos retardos máximos $D_1 = 79.5$ ms para las conversaciones y $D_2 = 36.4$ ms para las descargas de ficheros.

La Figura C.1 presenta el retardo máximo (calculado como el 99 percentil del retardo experimentado, en el interfaz inalámbrico simulado, por los paquetes en el caso de que sólo haya tráfico VoIP, incluyendo el tráfico en ambos sentidos. Como se comentó, según el modelo teórico este retardo debe ser menor o igual que 60 ms. El modelo indica que estos requisitos se cumplirán siempre que el tráfico admitido en el sistema esté por debajo de la tasa mínima de transmisión en el nivel físico, i.e.

$$\sum R_i < \Gamma = 1 \text{ Mbps}$$

Para el códec utilizado, una conversación de voz genera

$$74.7 \frac{\text{Kbps}}{\text{flujo}} \times 2 \frac{\text{flujos}}{\text{conversación}} = 149.4 \frac{\text{Kbps}}{\text{conversación}}$$

luego el máximo número de conversaciones para el que se debería cumplir esta garantía será

$$\left\lfloor \frac{1 \text{ Mbps}}{149.4 \frac{\text{Kbps}}{\text{conversación}}} \right\rfloor = 6 \text{ conversaciones}$$

De esta forma se comprueba que, para este caso concreto, el planificador desarrollado verifica el modelo analítico ya que los resultados muestran un valor inferior hasta con 12 ~ 13 conversaciones, i.e. el doble del tráfico impuesto por este modelo. Esta diferencia en el número de conversaciones soportadas puede deberse a que la tasa mínima

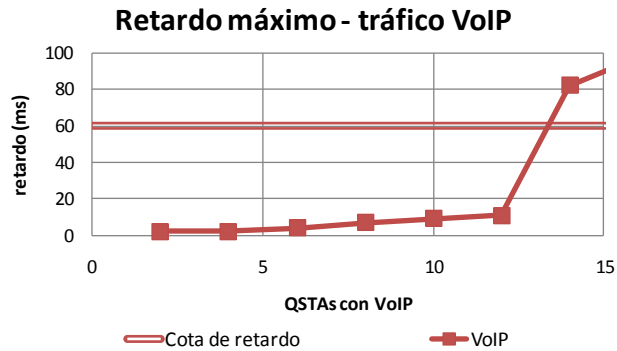


Figura C.1. Retardo máximo del planificador MSS con tráfico VoIP.

Γ utilizada (1 Mbps) es muy conservadora para intentar paliar la variabilidad del interfaz físico (tanto por las condiciones radio como por efectos del protocolo MAC, e.g. el proceso de contienda). Sin embargo, la velocidad de transferencia real en una red IEEE 802.11b normalmente es superior ya que puede variar, típicamente, entre 2 y 5 Mbps [100] [101] dependiendo de estos factores.

En el escenario con tráfico mixto (VoIP y FTP), el máximo número de conversaciones para el que se debe cumplir el requisito (Ecuación 4.6) será

$$\left\lfloor \frac{1 \text{ Mbps} - 660 \frac{\text{Kbps}}{\text{FTP}}}{149.4 \frac{\text{Kbps}}{\text{conversación}}} \right\rfloor = 2 \text{ conversaciones}$$

La Figura C.2 muestra el retardo máximo (99 percentil) tanto para VoIP como para FTP, cumpliendo el requisito de retardo para ambos servicios hasta con 6 conversaciones de voz, número superior a las 2 conversaciones exigidas.

También se ha comprobado que, para los escenarios simulados, se cumplen las exigencias impuestas para los planificadores utilizados en HCCA (Ecuación 2.5). Para ello, se ha calculado el valor mínimo de la diferencia entre el servicio real prestado por el planificador MSS (en bytes transmitidos) y el tráfico cursado exigido a un planificador HCCA según la Ecuación 2.5. Este cálculo ha sido realizado para cada flujo considerando

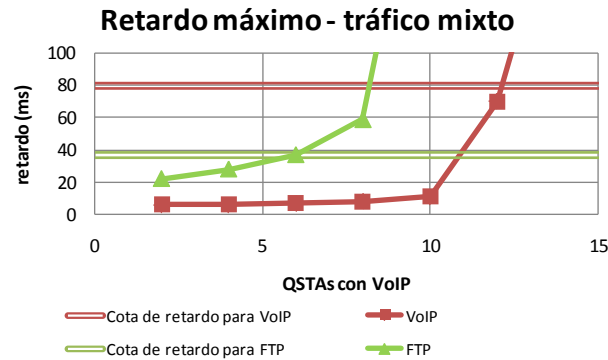


Figura C.2. Retardo máximo del planificador MSS con tráfico mixto (VoIP y FTP).

un Intervalo Especificado (véase la Sección 2.3.3) igual a un segundo (este intervalo puede variar aproximadamente entre 0 y 67 segundos [26]), eligiéndose el menor valor entre todos los flujos (caso peor) para las distintas simulaciones realizadas en cada punto de carga (las necesarias para un nivel de confianza fuese del 95 por ciento con un máximo de 10 réplicas).

La Figura C.3 muestra esta diferencia en el escenario que utiliza sólo tráfico generado por el servicio VoIP. Como se observa en la figura, este valor es siempre mayor o igual a 0 en el rango donde se verificaba la cota de retardo de paquetes (hasta 12 usuarios VoIP) por lo que, en dicho rango y en las simulaciones realizadas, el tráfico servido por MSS ha sido siempre superior al requerido.

De forma similar, las Figuras C.4 y C.5 muestran el mínimo exceso de servicio respecto al requerido en el escenario con tráfico mixto (VoIP y FTP) para cada uno de los servicios.

En esta comprobación se ha deshabilitado el regulador *dual-leaky-bucket* para la fuente FTP, ya que ésta limitaría la generación de tráfico exactamente al valor especificado en el TSPEC. Debido al control de congestión de TCP la tasa sería ligeramente inferior a ese límite, por lo que no se podría servir el tráfico requerido por el estándar al no haber sido generado. Nótese que esta dificultad se debe a la generación de tráfico y no al funcionamiento del planificador.

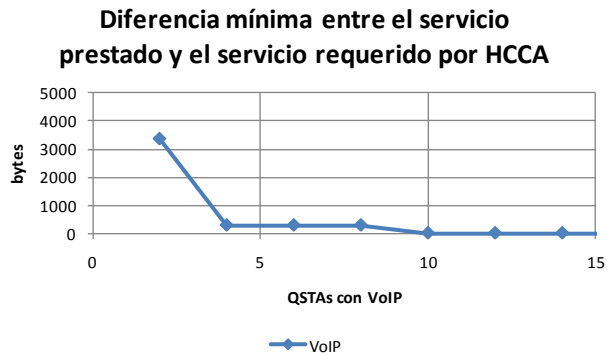


Figura C.3. Diferencia mínima entre el servicio prestado y el servicio requerido por HCCA del planificador MSS con tráfico VoIP.

Nuevamente, este valor es mayor o igual a 0 para el rango requerido (teóricamente hasta 2 QSTAs con tráfico VoIP; usando el rango con retardo acotado hasta 6 usuarios VoIP) por lo que se cumple la restricción impuesta a los planificadores HCCA.

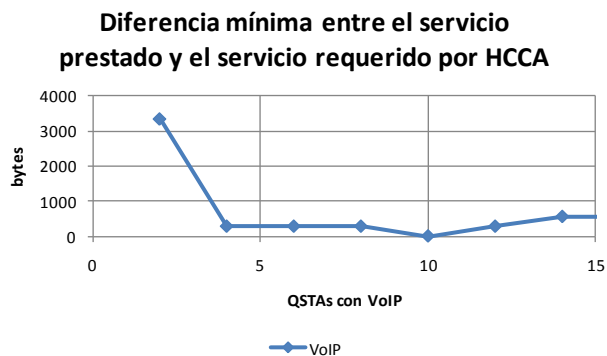


Figura C.4. Diferencia mínima entre el servicio prestado y el servicio requerido por HCCA para tráfico VoIP del planificador MSS con tráfico mixto (VoIP y FTP).

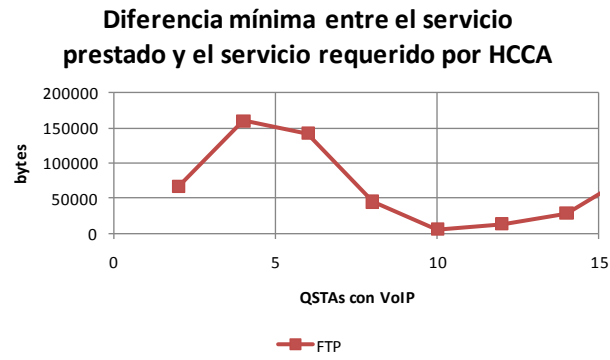


Figura C.5. Diferencia mínima entre el servicio prestado y el servicio requerido por HCCA para tráfico FTP del planificador MSS con tráfico mixto (VoIP y FTP).

Apéndice D

Selección de parámetros para el Método de los Territorios

La selección de los parámetros PFM_i , PAT_i y α_i afectan al comportamiento dinámico del Método del Territorio expuesto en el Capítulo 5. Para estimar adecuadamente estos parámetros, que dependen del servicio y la clase de tráfico concretos, habría que realizar un estudio pormenorizado –mediante el análisis de resultados obtenidos por simulación– para cada escenario y mezcla de tráficos dados.

Por este motivo se han realizado simulaciones usando tres valores que se han considerado representativos (*corto*, *intermedio* y *largo*) para los Períodos de Filtrado de MOS –0.2, 1 y 5 segundos– siendo los Períodos de Actualización de Territorio cinco veces su valor –1, 5 y 25 segundos–, de forma que se utilizan cinco muestras para tomar decisiones respecto a la asignación de recursos. El valor del peso en el promediado (véase la Sección 5.1), α_i , para estas simulaciones es 0.5 (salvo en el servicio de vídeo sobre TCP, que considera el tiempo de recuperación explicado en la Sección 3.2.3).

El tráfico generado ha seguido el mismo patrón que se utilizó en las simulaciones dinámicas para evaluar el rendimiento del Método del Territorio, explicado en la Sección 5.2.1. Éste se basaba en la mezcla de tráfico con *carga alta* (un usuario de *streaming* de

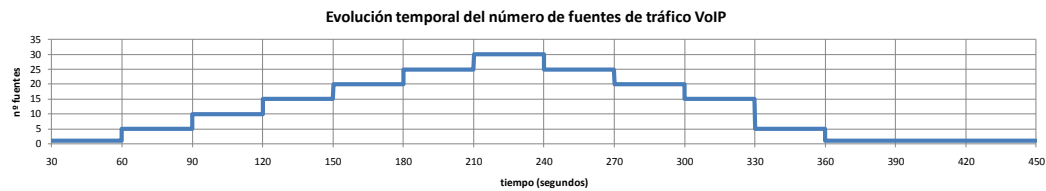


Figura D.1. Evolución temporal del tráfico VoIP durante el experimento con carga dinámica.

vídeo, dos descargas progresivas de vídeo, tres sesiones web y dos descargas de ficheros), modificando el número de fuentes VoIP durante la simulación entre 1 y 30. La evolución temporal de la carga correspondiente a las fuentes de voz se incluyó en la Figura 5.13, y se repite aquí para facilitar la lectura de este apéndice.

A continuación se muestran los resultados de las tres campañas de simulación realizadas, considerando la política de *reparto con derecho preferente*.

Resultados para $PFM = 0.2$ s, $PAT = 1$ s.

Las figuras siguientes (Figuras D.2 a D.6) muestran cómo el territorio conversacional fluctúa rápidamente dentro de los períodos con 5 o más conversaciones. Estas fluctuaciones provocan variaciones en el territorio de *streaming*, que a su vez generan picos negativos en la calidad percibida por los usuarios. Algo similar se observa también para los tráficos de vídeo sobre TCP y FTP.

Sin embargo, esta configuración parece ser adecuada para el tráfico web, adaptándose el MOS promediado al MOS instantáneo y reflejando el comportamiento a ráfagas de este tipo de tráfico.

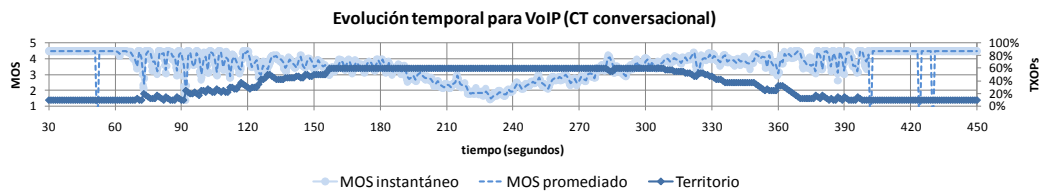


Figura D.2. Evolución temporal de MOS instantáneo, MOS promediado y territorio para VoIP (*CT conversacional*) utilizando PFM=0.2 s y PAT=1 s.

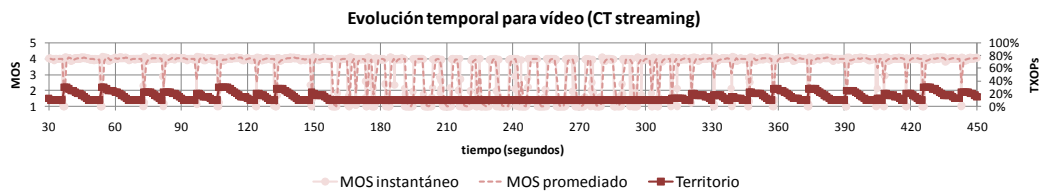


Figura D.3. Evolución temporal de MOS instantáneo, MOS promediado y territorio para *streaming* de vídeo (*CT streaming*) PFM=0.2 s y PAT=1 s.

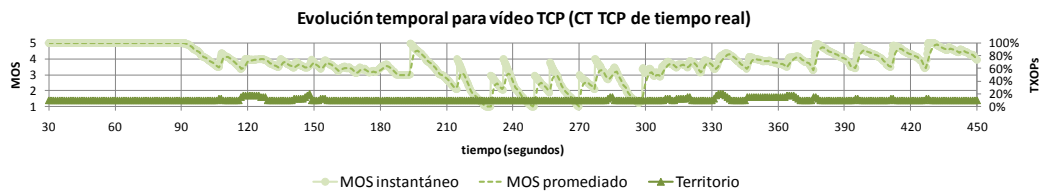


Figura D.4. Evolución temporal de MOS instantáneo, MOS promediado y territorio para *streaming* de vídeo sobre TCP (*CT TCP de tiempo rea*) utilizando PFM=0.2 s y PAT=1 s.

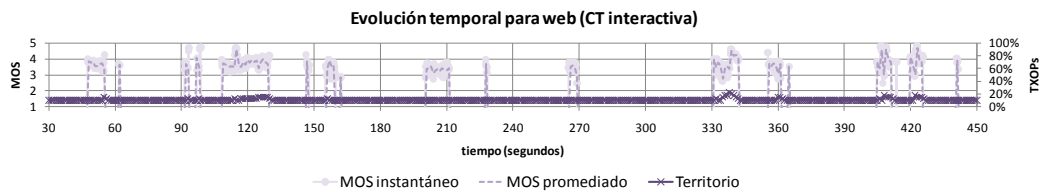


Figura D.5. Evolución temporal de MOS instantáneo, MOS promediado y territorio para navegación web (*CT interactiva*) utilizando PFM=0.2 s y PAT=1 s.

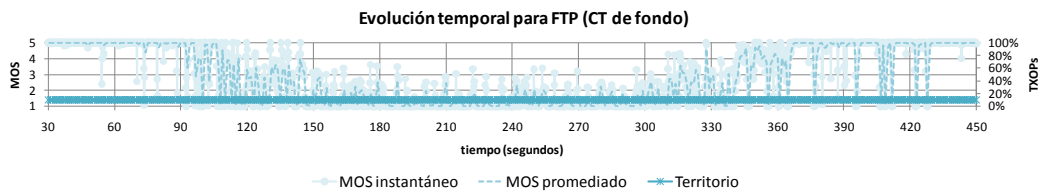


Figura D.6. Evolución temporal de MOS instantáneo, MOS promediado y territorio para FTP (*CT de fondo*) utilizando PFM=0.2 s y PAT=1 s.

Resultados para PFM = 1 s, PAT = 5 s.

Los resultados para esta nueva configuración parecen, en general, más suaves que los obtenidos anteriormente. La QoE del servicio VoIP varía más gradualmente que en el caso anterior, y los recursos de su territorio no presentan oscilaciones. Algo similar parece ocurrir para los tráficos de *streaming* de vídeo, vídeo sobre TCP y FTP.

Sin embargo, en el tráfico web se observa cómo han desaparecido algunos picos que antes sí se apreciaban. Esto se debe al reducido tiempo de descarga de una página web y al largo período de lectura, de forma que, si se distancia demasiado los instantes de medida de MOS instantáneo, se pierde información sobre el tráfico generado, su calidad y se imposibilita la toma de decisiones en estas ocasiones.

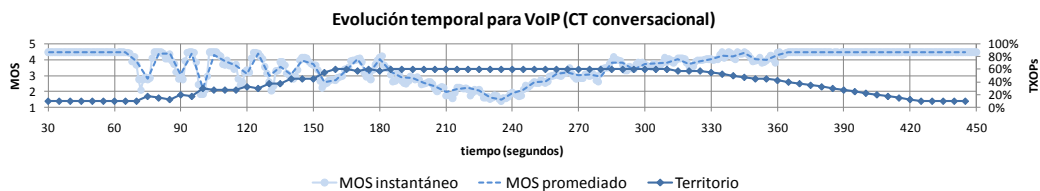


Figura D.7. Evolución temporal de MOS instantáneo, MOS promediado y territorio para VoIP (*CT conversacional*) utilizando PFM=1 s y PAT=5 s.

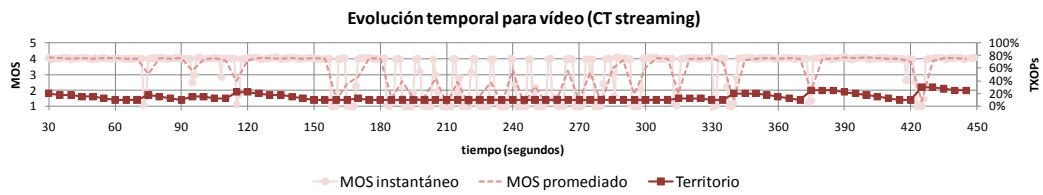


Figura D.8. Evolución temporal de MOS instantáneo, MOS promediado y territorio para *streaming* de vídeo (CT streaming) PFM=1 s y PAT=5 s.

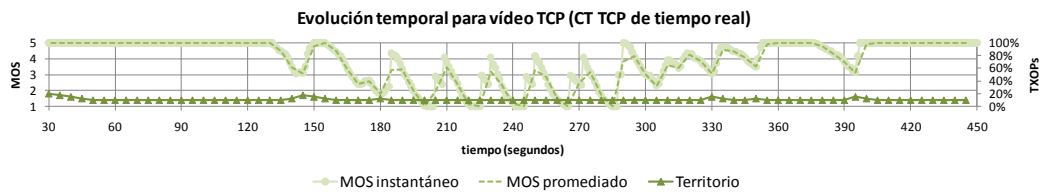


Figura D.9. Evolución temporal de MOS instantáneo, MOS promediado y territorio para *streaming* de vídeo sobre TCP (CT TCP de tiempo rea) utilizando PFM=1 s y PAT=5 s.

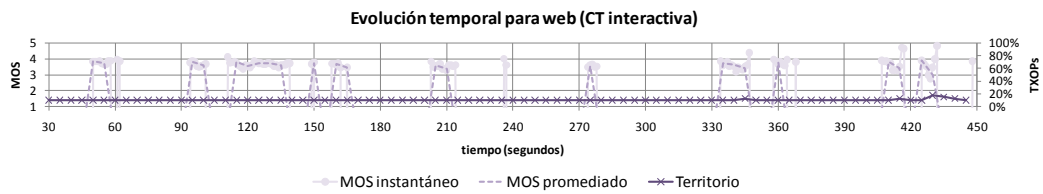


Figura D.10. Evolución temporal de MOS instantáneo, MOS promediado y territorio para navegación web (CT interactiva) utilizando PFM=1 s y PAT=5 s.

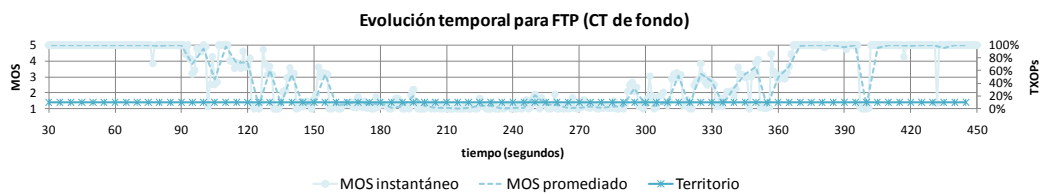


Figura D.11. Evolución temporal de MOS instantáneo, MOS promediado y territorio para FTP (CT de fondo) utilizando PFM=1 s y PAT=5 s.

Resultados para $PFM = 5$ s, $PAT = 25$ s.

Las figuras de estos últimos resultados (Figuras D.12 a D.16) muestran cómo, aumentando en exceso el valor de los parámetros PFM y PAT, se pierden detalles y se produce una falta de respuesta en la asignación de recursos. Así, los valores de MOS instantáneo y promediado no siguen las variaciones en la carga de tráfico, y la evolución de los territorios se vuelve muy lenta y no acorde a la dinámica del escenario.

Gracias a estas observaciones para las tres configuraciones analizadas, parece razonable utilizar $PFM_i = 1$ s y $PAT_i = 5$ s para los servicios VoIP, *streaming* de vídeo, vídeo sobre TCP y FTP; para el servicio de navegación web se utilizará $PFM = 0.2$ s y $PAT = 1$ s, de forma que las medidas y las decisiones se adapten al carácter fluctuante en su generación de tráfico.

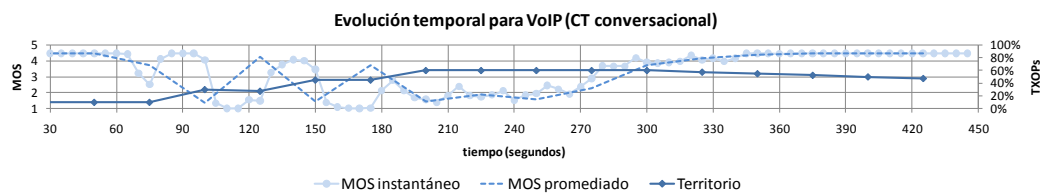


Figura D.12. Evolución temporal de MOS instantáneo, MOS promediado y territorio para VoIP (*CT conversacional*) utilizando $PFM=5$ s y $PAT=25$ s.

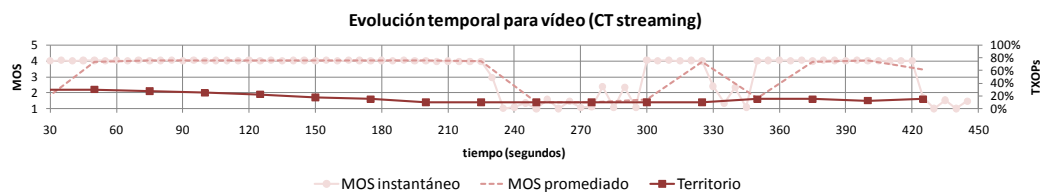


Figura D.13. Evolución temporal de MOS instantáneo, MOS promediado y territorio para *streaming* de vídeo (*CT streaming*) $PFM=5$ s y $PAT=25$ s.

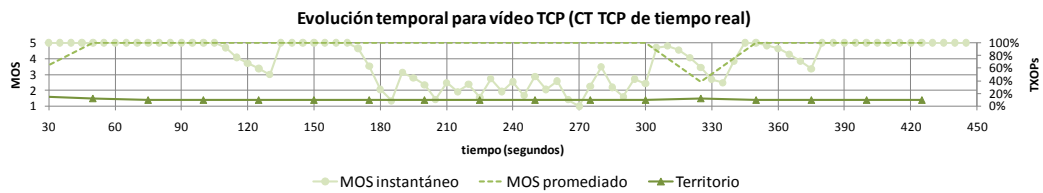


Figura D.14. Evolución temporal de MOS instantáneo, MOS promediado y territorio para *streaming* de vídeo sobre TCP (*CT TCP de tiempo rea*) utilizando PFM=5 s y PAT=25 s.

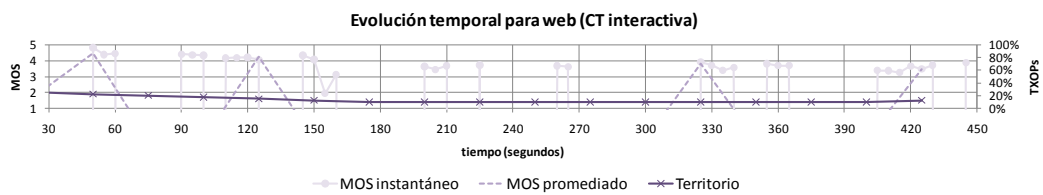


Figura D.15. Evolución temporal de MOS instantáneo, MOS promediado y territorio para navegación web (*CT interactiva*) utilizando PFM=5 s y PAT=25 s.

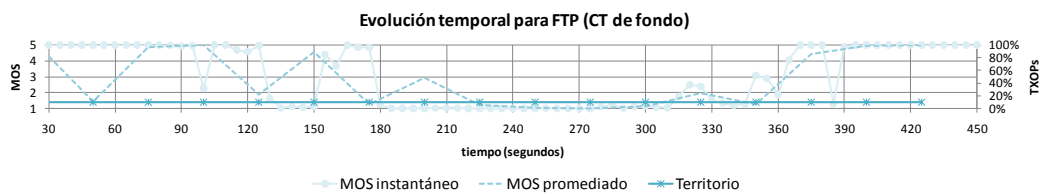


Figura D.16. Evolución temporal de MOS instantáneo, MOS promediado y territorio para FTP (*CT de fondo*) utilizando PFM=5 s y PAT=25 s.

Tras este estudio sobre los parámetros PFM y PAT, se incluye un análisis del impacto del valor del parámetro α_i . Para ello, se han utilizado los valores de PFM y PAT razonados en el estudio anterior. De forma similar a ese estudio, se han usado tres valores representativos: 0.1, 0.5 y 0.9. Un valor bajo indica que el MOS promediado es muy parecido a la muestra actual de MOS instantáneo (Ecuación 5.1), mientras que un valor alto indica que el MOS promediado considera sobre todo las muestras pasadas. Este estudio debe considerarse como un análisis preliminar y no como una optimización, ya

que ésta dependería del escenario concreto.

Nótese que el valor de este parámetro para el servicio de vídeo sobre TCP depende del tiempo de recuperación ($TR = 1$ s, por lo que α_i vale 0.48 para este servicio) considerado, por lo que su valor no se modificará en las simulaciones realizadas.

A continuación se muestran los resultados para las configuraciones del parámetro α_i comentadas. El escenario simulado utiliza el mismo tráfico *dinámico* utilizado para el estudio de los parámetros PFM y PAT.

Resultados para $\alpha_i = 0.1$.

Como se comentó, un valor bajo de α_i hace que el MOS promediado sólo tenga en cuenta prácticamente la última muestra de MOS instantáneo. Así, cualquier variación brusca en el MOS instantáneo podría aparecer en el MOS promediado, haciendo que surgieran oscilaciones en el tamaño del territorio.

Sin embargo, el período de filtrado de MOS (PFM) tiene un efecto similar de promediado y suavizado, que parece reducir el impacto de este parámetro ya que se obtienen resultados muy similares a los conseguidos para $\alpha_i = 0.5$ (Figuras D.22 a D.26). Valores reducidos del parámetro PFM podrían hacer más notable las diferencias entre distintos valores de α_i , pero no es el objetivo de este estudio analizar todas las posibilidades sino justificar unos valores razonables para las simulaciones de evaluación del rendimiento.

Resultados para $\alpha_i = 0.5$.

Este valor de α_i podría ser un buen compromiso entre suavidad y velocidad de respuesta, reduciéndose las oscilaciones en el MOS promediado al considerar más muestras de MOS instantáneo.

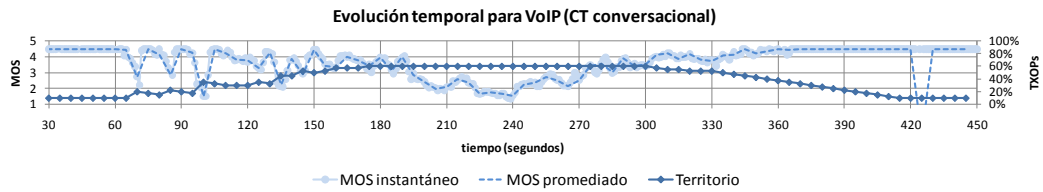


Figura D.17. Evolución temporal de MOS instantáneo, MOS promediado y territorio para VoIP (*CT conversacional*) utilizando $\alpha_i = 0.1$.

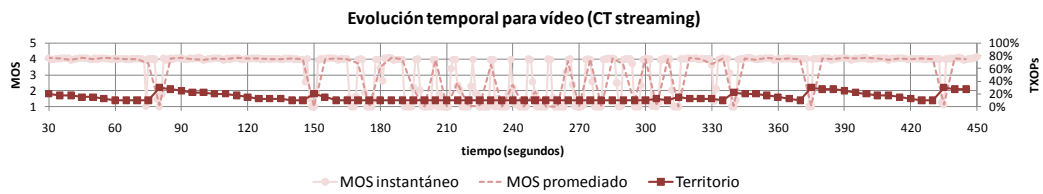


Figura D.18. Evolución temporal de MOS instantáneo, MOS promediado y territorio para *streaming* de vídeo (*CT streaming*) utilizando $\alpha_i = 0.1$.

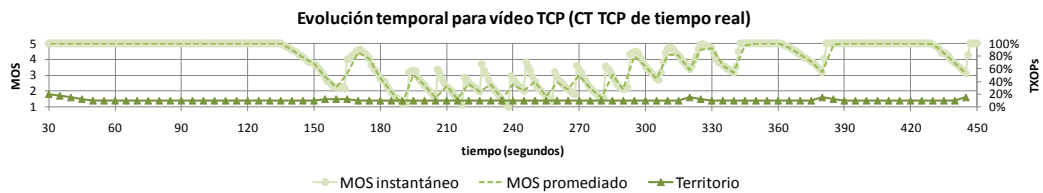


Figura D.19. Evolución temporal de MOS instantáneo, MOS promediado y territorio para *streaming* de vídeo sobre TCP (*CT TCP de tiempo rea*) utilizando $\alpha_i = 0.1$.

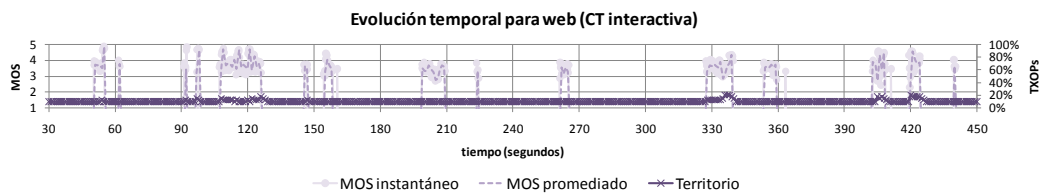


Figura D.20. Evolución temporal de MOS instantáneo, MOS promediado y territorio para navegación web (*CT interactiva*) utilizando $\alpha_i = 0.1$.

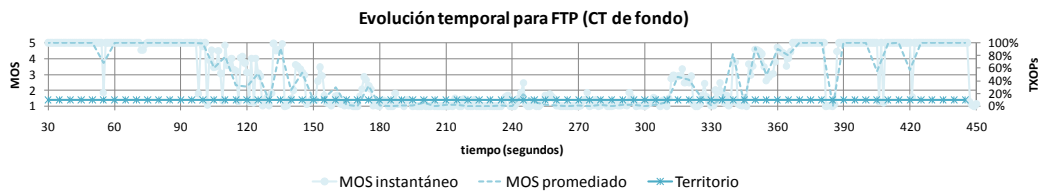


Figura D.21. Evolución temporal de MOS instantáneo, MOS promediado y territorio para FTP (CT de fondo) utilizando $\alpha_i = 0.1$.

No obstante, como se comentó anteriormente, las diferencias entre un valor α_i pequeño (0.1) y uno intermedio (0.5) parecen no ser demasiado importantes debido al efecto de suavizado del período de filtrado de MOS (PFM). En ambos casos, las curvas obtenidas presentan el mismo tipo de picos (tanto en profundidad como duración) y el MOS promediado sigue bastante fielmente al MOS instantáneo.

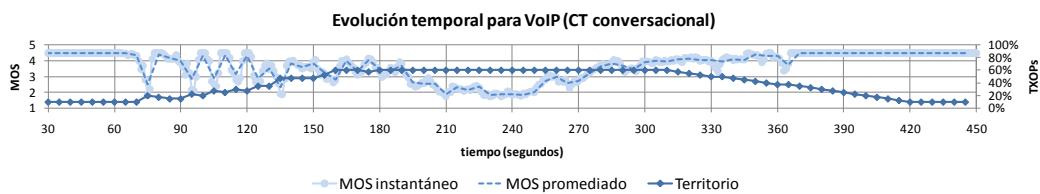


Figura D.22. Evolución temporal de MOS instantáneo, MOS promediado y territorio para VoIP (CT conversacional) utilizando $\alpha_i = 0.5$.

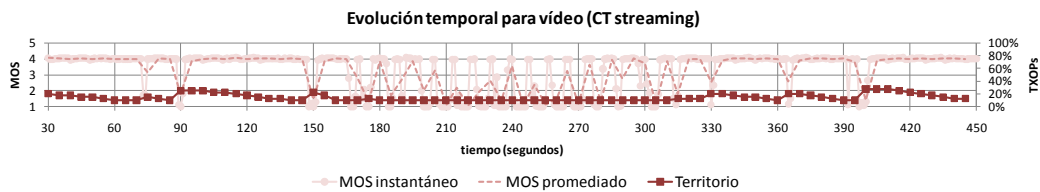


Figura D.23. Evolución temporal de MOS instantáneo, MOS promediado y territorio para *streaming* de vídeo (CT streaming) utilizando $\alpha = 0.5$.

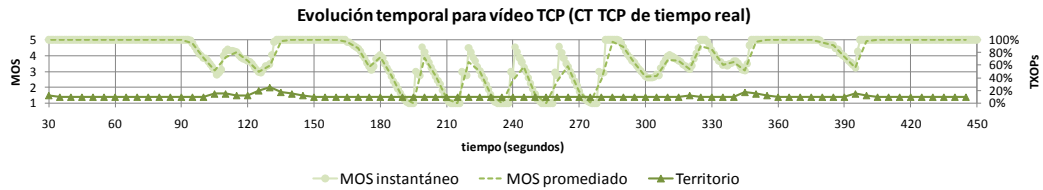


Figura D.24. Evolución temporal de MOS instantáneo, MOS promediado y territorio para *streaming* de vídeo sobre TCP (*CT TCP de tiempo rea*) utilizando $\alpha_i = 0.5$.

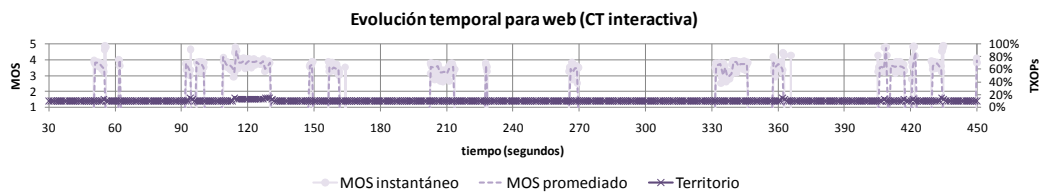


Figura D.25. Evolución temporal de MOS instantáneo, MOS promediado y territorio para navegación web (*CT interactiva*) utilizando $\alpha_i = 0.5$.

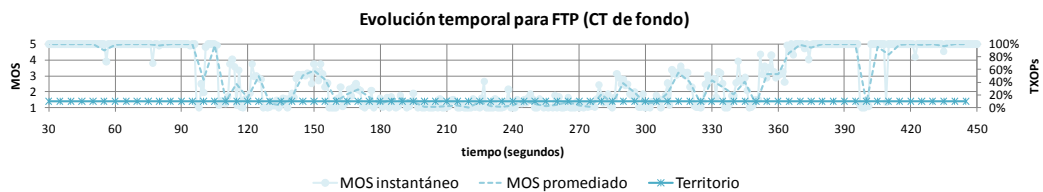


Figura D.26. Evolución temporal de MOS instantáneo, MOS promediado y territorio para FTP (*CT de fondo*) utilizando $\alpha_i = 0.5$.

Resultados para $\alpha_i = 0.9$.

En el caso de un valor de α_i muy alto el MOS promediado tarda excesivamente en seguir las variaciones del MOS instantáneo, provocando que los territorios también se actualicen con retardo. Esto se observa claramente en el caso del tráfico VoIP (Figura D.27), cuya calidad instantánea experimenta fluctuaciones importantes provocando que la percepción del usuario se degrade. También se observa para el resto de servicios (salvo

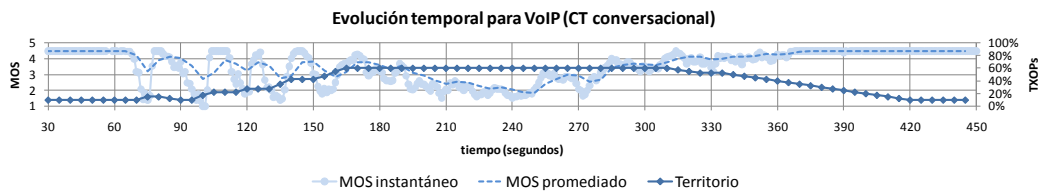


Figura D.27. Evolución temporal de MOS instantáneo, MOS promediado y territorio para VoIP (*CT conversacional*) utilizando $\alpha_i = 0.9$.

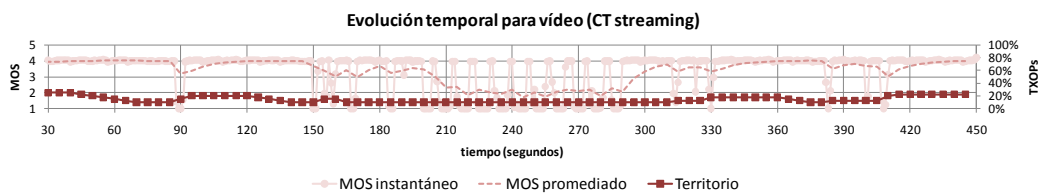


Figura D.28. Evolución temporal de MOS instantáneo, MOS promediado y territorio para *streaming* de vídeo (*CT streaming*) utilizando $\alpha_i = 0.9$.

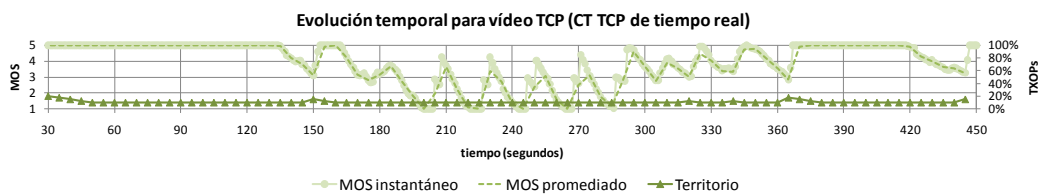


Figura D.29. Evolución temporal de MOS instantáneo, MOS promediado y territorio para *streaming* de vídeo sobre TCP (*CT TCP de tiempo rea*) utilizando $\alpha_i = 0.9$.

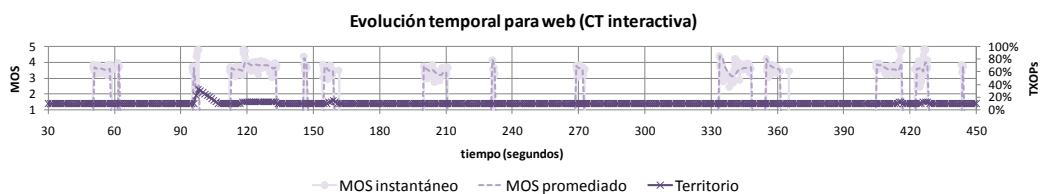


Figura D.30. Evolución temporal de MOS instantáneo, MOS promediado y territorio para navegación web (*CT interactiva*) utilizando $\alpha_i = 0.9$.

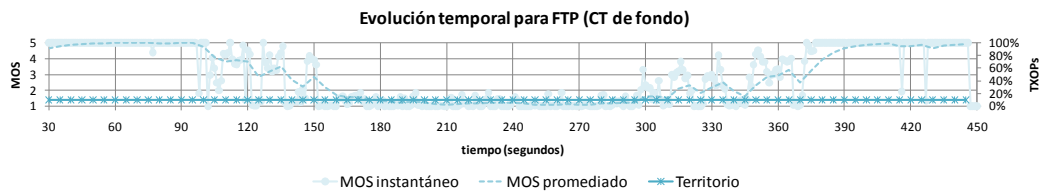


Figura D.31. Evolución temporal de MOS instantáneo, MOS promediado y territorio para FTP (CT de fondo) utilizando $\alpha_i = 0.9$.

vídeo sobre TCP, cuyo parámetro α_i depende del tiempo de recuperación), lo que indica que un valor excesivamente alto degradará el rendimiento de la red.

Si bien parece que el valor del parámetro α_i elegido no afectará dramáticamente al rendimiento del sistema, las conclusiones extraídas de este estudio y los razonamientos efectuados han motivado que se haya optado por utilizar un valor de 0.5 en las campañas de simulación realizadas. Este valor parece un compromiso razonable entre velocidad de respuesta y suavidad (falta de oscilaciones).

Bibliografía

- [1] IEEE std. 802.11-1997, part 11: wireless LAN medium access control (MAC) and physical layer (PHY) specifications, June 1997.
- [2] IEEE std. 802.11e-2005, part 11: wireless LAN medium access control (MAC) and physical layer (PHY) specifications, amendment 8: medium access control (MAC) quality of service enhancements, November 2005.
- [3] A. Grilo, M. Macedo, and M. Nunes. A scheduling algorithm for QoS support in IEEE 802.11 networks. *IEEE Wireless Communications*, 10(3):36–43, February 2005.
- [4] Claudio Cicconetti, Luciano Lenzini, Enzo Mingozzi, and Giovanni Stea. Design and performance analysis of the Real-Time HCCA scheduler for IEEE 802.11e WLANs. *Computer Networks*, 51(9):2311–2325, 2007.
- [5] Augusto Foronda, Yuhi Higuchi, Chikara Ohta, Masahiko Yoshimoto, and Yoji Okada. Service interval optimization with delay bound guarantee for HCCA in IEEE 802.11e WLANs. *IEICE Transactions*, 90-B(11):3158–3169, 2007.
- [6] Thanasis Korakis and Leandros Tassiulas. Providing quality of service guarantees in wireless LANs compliant with 802.11e. *Computer Networks*, 47(2):239–255, 2005.
- [7] Inanc Inan, Feyza Keceli, and Ender Ayanoglu. An adaptive multimedia QoS scheduler for 802.11e wireless LANs. In *Proceedings of the 2006 IEEE International Conference on Communications*, pages 5263–5270, December 2006.
- [8] Pierre Ansel, Qiang Ni, and Thierry Turletti. FHCF: A simple and efficient scheduling scheme for IEEE 802.11e wireless LAN. *Mobile Networks and Applications*, 11(3):391–403, June 2006.
- [9] Gennaro Boggia, Pietro Camarda, Luigi A. Grieco, and Saverio Mascolo.

Feedback-based control for providing real-time services with the 802.11e MAC. *IEEE/ACM Transactions on Networking*, 15(2):323–333, 2007.

[10] L. Yang. Enhanced HCCA for real-time traffic with QoS in IEEE 802.11e based networks. In *Proceedings of the 2005 IEE International Workshop on Intelligent Environments (IE '05)*, volume 2005, pages 203–209, 2005.

[11] Zul Bin Muhamad Noh, Takahiro Suzuki, and Shuji Tasaka. A packet scheduling scheme for audio-video transmission with IEEE 802.11e HCCA and its application-level QoS assessment. In *Proceedings of the 2006 Asia-Pacific Conference on Communications*, pages 1–5, December 2006.

[12] Dimitris Skyrianoglou, Nikos Passas, and Apostolis Salkintzis. ARROW: An efficient traffic scheduling algorithm for IEEE 802.11e HCCA. *IEEE Transactions on Wireless Communications*, 5(12):3558–3567, December 2006.

[13] Claudio Cicconetti, Luciano Lenzini, Enzo Mingozzi, and Giovanni Stea. An efficient cross layer scheduler for multimedia traffic in wireless local area networks with IEEE 802.11e HCCA. *SIGMOBILE Mobile Computing and Communications Review*, 11(3):31–46, 2007.

[14] Guocong Song and Ye Li. Cross-layer optimization for OFDM wireless networks - part I: theoretical framework. *IEEE Transactions on Wireless Communications*, 4(2):614–624, 2005.

[15] Guocong Song and Ye Li. Cross-layer optimization for OFDM wireless networks - part II: algorithm development. *IEEE Transactions on Wireless Communications*, 4(2):625–634, March 2005.

[16] Pablo Ameigeiras, Juan J. Ramos-Munoz, Jorge Navarro-Ortiz, Preben Mogensen, and Juan M. Lopez-Soler. QoE oriented cross-layer design of a resource allocation algorithm in beyond 3G systems. *Computer Communications*, 33(5):571–582, October 2009.

[17] S. Khan, J. Brehmer, W. Kellerer, W. Utschick, and E. Steinbach. Application-

driven cross-layer optimization for video streaming over wireless networks. *IEEE Communications Magazine*, 44:122–130, 2006.

[18] A. Saul. Simple optimization algorithm for MOS-based resource assignment. In *Proceedings of the 2008 Vehicular Technology Conference (VTC Spring 2008)*, pages 1766–1770, May 2008.

[19] Andreas Saul and Gunther Auer. Multiuser resource allocation maximizing the perceived quality. *EURASIP Journal on Wireless Communications and Networking*, 2009:1–15, January 2009.

[20] Jorge Navarro-Ortiz, Pablo Ameigeiras, Juan J. Ramos-Munoz, and Juan M. Lopez-Soler. An 802.11e HCCA scheduler with an end-to-end quality aware territory method. *Computer Communications*, 32(11):1281–1297, 2009.

[21] Jorge Navarro-Ortiz, Juan M. Lopez-Soler, and Giovanni Stea. Quality of experience based resource sharing in ieee 802.11e hcca. In *Proceedings of the 16th European Wireless conference (EW 2010)*, pages 1–8, April 2010.

[22] ITU-T recommendation G.107. International telephone connection and circuits. General definitions. The E model, a computational model for use in transmission planning, March 2005.

[23] Lingfen S. Asiya Khan. Content-based video quality prediction for MPEG4 video streaming over wireless networks. *Journal of Multimedia*, 4(4):228–239, August 2009.

[24] Yasuhiro Sato, Shin’Ichiro Kitano, Shingo Ata, and Ikuo Oka. The measurement of user’s subjective assessment on throughput degradation in bulk data transmission. In *Proceedings of the 11th Asia-Pacific Symposium on Network Operations and Management*, pages 236–245, 2008.

[25] Pablo Ameigeiras, Juan J. Ramos-Munoz, Jorge Navarro-Ortiz, Juan Ramiro-Moreno, and Juan M. Lopez-Soler. QoE evaluation of scheduling algorithms for NRT services in LTE. In *Proceedings of the ICT Mobile and Wireless Communications Summit (ICT-MobileSummit 2009)*, June 2009.

238 | Bibliografia

- [26] IEEE std. 802.11-2007, part 11: wireless LAN medium access control (MAC) and physical layer (PHY) specifications, June 2007.
- [27] IEEE 802.11 official timelines, http://grouper.ieee.org/groups/802/11/reports/802.11_timelines.htm.
- [28] Gianni Costanzi. A Linux kernel based implementation of 802.11e HCCA, <http://utenli.lycos.it/giannicostanzi/my-projects.html>, Master Thesis.
- [29] ITU-T recommendation P.10/G.100. Vocabulary and effects of transmission parameters on customer opinion of transmission quality, July 2008.
- [30] ITU-T recommendation E.800. Quality of telecommunication services: concepts, models, objectives and dependability planning. Terms and definitions related to the quality of telecommunication services, September 2008.
- [31] ITU-T P.800. Methods for subjective determination of transmission quality - Series P: telephone transmission quality; methods for objective and subjective assessment of quality, August 1996.
- [32] ITU-T recommendation G.1030. Estimating end-to-end performance in IP networks for data applications, November 2005.
- [33] ITU-T recommendation G.1070. Opinion model for video-telephony applications, April 2007.
- [34] ITU-T recommendation G.1080. Quality of experience requirements for IPTV services, December 2008.
- [35] 3GPP TS 23.107. Technical specification group services and system aspects; quality of service (QoS) concept and architecture (release 9), v9.0.0, December 2009.
- [36] 3GPP TS 23.203. Policy and charging control architecture (release 9), v9.2.0, December 2009.

- [37] W3C recommendation for SOAP version 1.2 specification, <http://www.w3.org/tr/soap12/>, April 2007.
- [38] Roy T. Fielding. *PhD Dissertation "Architectural Styles and the Design of Network-based Software Architectures"*, chapter 5 "Representational State Transfer (REST)", http://www.ics.uci.edu/fielding/pubs/dissertation/rest_arch_style.htm. PhD thesis, University of California, Irvine, 2000.
- [39] YouTube - Broadcast Yourself, <http://www.youtube.com>.
- [40] ITU-T recommendation G.711. Pulse code modulation (PCM) of voice frequencies, November 1988.
- [41] ITU-T recommendation P.862. Perceptual evaluation of speech quality (PESQ): an objective method for end-to-end speech quality assessment of narrow-band telephone networks and speech codecs, February 2001.
- [42] A. Takahashi, H. Yoshino, and N. Kitawaki. Perceptual QoS assessment technologies for VoIP. *IEEE Communications Magazine*, 42(7):28–34, July 2004.
- [43] Floriano De Rango, Mauro Tropea, Peppino Fazio, and Salvatore Marano. Overview on VoIP: Subjective and objective measurement methods. *International Journal of Computer Science and Network Security*, 6(1):140–153, January 2006.
- [44] J. Klaue, B. Rathke, and A. Wolisz. EvalVid - a framework for video transmission and quality evaluation. In *Proceedings of the 13th International Conference on Modelling Techniques and Tools for Computer Performance Evaluation (TOOLS '03)*, pages 255–272. 2003.
- [45] Fabio Sonnati. Implementing a dual-threshold buffering strategy in Flash Media Server, http://www.adobe.com/devnet/flashmediaserver/articles/fms_dual_buffering.html, July 2006.
- [46] Saâd Biaz and Nitin H. Vaidya. "De-randomizing" congestion losses to improve TCP performance over wired-wireless networks. *IEEE/ACM Transactions on*

240 | Bibliografía

Networking, 13(3):596–608, 2005.

[47] Vijender B. Reddy and Anil K. Sarje. Differentiation of wireless and congestion losses in TCP. In *Proceedings of the 2nd International Conference on Communication Systems Software and Middleware (COMMSWARE '07)*, pages 1–5, January 2007.

[48] V. Arya and T. Turetli. Accurate and explicit differentiation of wireless and congestion losses. In *Proceedings of the 23rd International Conference on Distributed Computing Systems (ICDCS '03)*, pages 877–882, 2003.

[49] Mozilla Firefox, <http://www.mozilla-europe.org/es/firefox/>.

[50] A. King. The average web page, <http://www.optimizationweek.com/reviews/average-web-page>, October 2006.

[51] Vector, <http://www.vector.co.jp/>.

[52] Broadband networking report, <http://www.musen-lan.com/speed/htmldata/>.

[53] Rajendra K. Jain, Dah-Ming W. Chiu, and William R. Hawe. Dec-tr-301. A quantitative measure of fairness and discrimination for resource allocation in shared computer systems. Technical report, Digital Equipment Corporation, September 1984.

[54] A. Francini and F. M. Chiussi. Minimum-latency dual-leaky-bucket shapers for packet multiplexers: theory and implementation. In *Proceedings of the 8th International Workshop on Quality of Service (IWQOS 2000)*, pages 19–28, June 2000.

[55] H. Jonathan Chao, Yau-Ren Jenq, Xiaolei Guo, and Cheuk H. Lam. Design of packet-fair queuing schedulers using a RAM-based searching engine. *IEEE Journal on Selected Areas in Communications*, 17(6):1105–1126, 1999.

[56] A. K. Parekh and R. G. Gallager. A generalized processor sharing approach to flow control in integrated services networks: the multiple node case. *IEEE/ACM Transactions on Networking*, 2(2):137–150, April 1994.

- [57] Abhay K. Parekh and Robert G. Gallager. A generalized processor sharing approach to flow control in integrated services networks: the single-node case. *IEEE/ACM Transactions on Networking*, 1(3):344–357, August 1993.
- [58] H. Jonathan Chao and Jun S. Hong. Design of an ATM shaping multiplexer with guaranteed output burstiness. *Computer Systems Science and Engineering*, 12:131–141, 1996.
- [59] Dimitrios Stiliadis and Anujan Varma. Latency-rate servers: a general model for analysis of traffic scheduling algorithms. *IEEE/ACM Transactions on Networking*, 6(5):611–624, 1998.
- [60] Lixia Zhang. VirtualClock: a new traffic control algorithm for packet-switched networks. *ACM Transactions on Computer Systems*, 9(2):101–124, 1991.
- [61] Qian Wu, Mingwei Gong, and Carey Williamson. TCP fairness issues in IEEE 802.11 wireless LANs. *Computer Communications*, 31(10):2150–2161, 2008.
- [62] Network Simulator 2, <http://www.isi.edu/nsnam/ns/>.
- [63] Matteo M. Andreozzi, Giovanni Stea, and Carlo Vallati. A framework for large-scale simulations and output result analysis with NS-2. In *Proceedings of the 2nd International Conference on Simulation Tools and Techniques (Simutools '09)*, pages 1–7, 2009.
- [64] Cicconetti. A software architecture for simulating IEEE 802.11e HCCA. In *Proceedings of the 3rd International Workshop on Internet Performance, Simulation, Monitoring and Measurement (IPS-MoMe 2005)*, pages 97–104, March 2005.
- [65] ITU-T recommendation P.59. Telephone transmission quality objective measuring apparatus. Artificial conversational speech, March 1993.
- [66] A. Bacioccola, C. Cicconetti, and G. Stea. User-level performance evaluation of VoIP using NS-2. In *Proceedings of the 2nd International Conference on Performance Evaluation Methodologies and Tools (ValueTools '07)*, pages 1–10, 2007.

242 | Bibliografia

- [67] Ce-Kuen S. Chih-Heng Ke. An evaluation framework for more realistic simulations of MPEG video transmission. *Journal of Information Science and Engineering*, 24(2):425–440, 2008.
- [68] Microsoft® Windows Media™ Player, <http://www.microsoft.com/windows/windowsmedia/default.msp>.
- [69] Xu Cheng, C. Dale, and Jiangchuan Liu. Statistics and social network of YouTube videos. In *Proceedings of the 16th International Workshop on Quality of Service (IWQoS 2008)*, pages 229–238, 2008.
- [70] M. Molina, P. Castelli, and G. Foddis. Web traffic modeling exploiting TCP connections' temporal clustering through HTML-REDUCE. *IEEE Network*, 14(3):46–55, August 2002.
- [71] Sam Jansen and Anthony McGregor. Simulation with real world network stacks. In *Proceedings of the 37th Conference on Winter Simulation (WSC '05)*, pages 2454–2463, 2005.
- [72] David X. Wei and Pei Cao. NS-2 TCP-Linux: an NS-2 TCP implementation with congestion control algorithms from Linux. In *Proceeding of the 2006 Workshop on NS-2: the IP network simulator (WNS2 '06)*, pages 9+, 2006.
- [73] Floyd. RFC 2582, the New-Reno modification to TCP's fast recovery algorithm, April 1999.
- [74] Alberto Medina, Mark Allman, and Sally Floyd. Measuring the evolution of transport protocols in the internet. *SIGCOMM Computer Communication Review*, 35(2):37–52, April 2005.
- [75] Flavius C. Bogdan Moraru. Practical analysis of TCP implementations: Tahoe, Reno, NewReno. In *Proceedings of the 2nd RoEduNet International Conference*, June 2003.
- [76] Toni Janevski and Ivan Petrov. Optimization of TCP/IP over 802.11 wireless

networks in home environment. In *Communications Infrastructure. Systems and Applications in Europe*, volume 16, chapter 25, pages 240–251. Springer Berlin Heidelberg, 2009.

[77] Rekha Jain, Apurva Gaiwak, and P. D. Vyavahare. Performance comparison of TCP NewReno and Vegas for IEEE 802.11e in multihop wireless networks. In *Proceedings of the 4th International Conference on Wireless Communication and Sensor Networks (WCSN '08)*, pages 212–217, December 2008.

[78] Jiwei Chen, Mario Gerla, Yeng Lee, and Medy Sanadidi. TCP with variance control for multihop IEEE 802.11 wireless networks. In *Proceedings of the 2006 Military Communications Conference (MILCOM 2006)*, volume 0, pages 1–7, 2006.

[79] Claudio Cicconetti, Enzo Mingozzi, and Giovanni Stea. An integrated framework for enabling effective data collection and statistical analysis with NS-2. In *Proceeding from the 2006 workshop on NS-2: the IP network simulator (WNS2 '06)*, 2006.

[80] Qiang N. Pierre Ansel. FHCF implementation in NS2, <http://planete.inria.fr/qni/fhcf/>, March 2004.

[81] S. Khan, S. Duhovnikov, E. Steinbach, M. Sgroi, and W. Kellerer. Application-driven cross-layer optimization for mobile multimedia communication using a common application layer quality metric. In *Proceedings of the 2006 International Conference on Wireless Communications and Mobile Computing (IWCMC '06)*, pages 213–218, 2006.

[82] Shoaib Khan, Svetoslav Duhovnikov, Eckehard Steinbach, and Wolfgang Kellerer. MOS-based multiuser multiapplication cross-layer optimization for mobile multimedia communication. *Advances in MultiMedia*, 2007(1):6, 2007.

[83] Andreas Saul. Simple optimization algorithm for MOS-based resource assignment. In *Proceedings of the 67th IEEE Vehicular Technology Conference (VTC Spring 2008)*, pages 1766–1770, May 2008.

[84] Andreas Saul and Gunther Auer. Multiuser resource allocation maximizing the perceived quality. *EURASIP Journal on Wireless Communications and Networking*,

2009:1–15, 2009.

[85] Antonio Grilo and Mario Nunes. Performance evaluation of IEEE 802.11e. In *Proceeding of the 13th IEEE International Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC 2002)*, 2002.

[86] F. Keceli, I. Inan, and E. Ayanoglu. Weighted fair uplink/downlink access provisioning in IEEE 802.11e WLANs. In *Proceedings of the 2008 IEEE International Conference on Communications (ICC '08)*, pages 2473–2479, May 2008.

[87] Ajay Bakre and B. R. Badrinath. I-TCP: Indirect TCP for mobile hosts. In *Proceedings of the 15th International Conference on Distributed Computing Systems*, pages 136–143, 1995.

[88] Hakyung Jung, Nakjung Choi, Yongho Seok, Taekyoung Kwon, and Yanghee Choi. Augmented split-TCP over wireless LANs. In *Proceedings of the 2006 IEEE International Conference on Communication (ICC '06)*.

[89] Dimitris Velenis, Dimitris Kalogeras, and Basil S. Maglaris. SaTPEP (Satellite TCP Performance Enhancing Proxy) model for NS-2 version 2.27, <http://www.netmode.ntua.gr/dbelen/satpep/>.

[90] Dimitris Velenis, Dimitris Kalogeras, and Basil S. Maglaris. SaTPEP: a TCP Performance Enhancing Proxy for satellite links. In *Proceedings of the 2nd International IFIP-TC6 Networking Conference on Networking Technologies, Services, and Protocols; Performance of Computer and Communication Networks; and Mobile and Wireless Communications (NETWORKING '02)*, pages 1233–1238, 2002.

[91] Rahul Jain and Teunis J. Ott. Design and implementation of split TCP in the Linux kernel. In *Proceedings of the 49th IEEE Global Telecommunications Conference (GLOBECOM 2006)*, pages 1–6, November 2006.

[92] Jitendra Padhye, Victor Firoiu, Donald F. Towsley, and James F. Kurose. Modeling TCP Reno performance: a simple model and its empirical validation. *IEEE/ACM Transactions on Networking*, 8(2):133–145, April 2000.

- [93] A. Wierman and T. Osogami. A unified framework for modeling TCP-Vegas, TCP-SACK, and TCP-Reno. In *Proceedings of the 11th IEEE/ACM International Symposium on Modeling, Analysis and Simulation of Computer Telecommunications Systems (MASCOTS 2003)*, pages 269–278, 2003.
- [94] Nadim Parvez, Anirban Mahanti, and Carey Williamson. An analytic throughput model for TCP NewReno. Technical report, Department of Computer Science, University of Calgary, January.
- [95] N. Katsuhiko, H. Okada, T. Yamazato, M. Katayama, and A. Ogawa. New analytical model for the TCP throughput in wireless environment. In *Proceedings of the 53rd Vehicular Technology Conference (VTC 2001 Spring)*, volume 3, pages 2128–2132, May 2001.
- [96] Carsten Burmeister, Ulrich Killat, and Jens Bachmann. TCP over rate-adaptive WLAN - an analytical model and its simulative verification. In *Proceedings of the 2006 International Symposium on on World of Wireless, Mobile and Multimedia Networks (WOWMOM '06)*, pages 339–348, 2006.
- [97] Lili Qiu, Yin Zhang, and Srinivasan Keshav. Understanding the performance of many TCP flows. *Computer Networks*, 37(3-4):277–306, 2001.
- [98] Ramón Agüero, Marta Garcá-Arranz, and Luis Muñoz. Accurate simulation of 802.11 indoor links: A bursty channel model based on real measurements. *EURASIP Journal on Wireless Communications and Networking*, 2010:1–18, 2010.
- [99] Qiang Ni, Tianji Li, Thierry Turletti, and Yang Xiao. Saturation throughput analysis of error-prone 802.11 wireless networks. *Wireless Communications and Mobile Computing*, 5(8):945–956, 2005.
- [100] M. Heusse, F. Rousseau, G. Berger-Sabbatel, and A. Duda. Performance anomaly of 802.11b. In *Proceedings of the 22th IEEE Conference of the IEEE Computer and Communications Societies (INFOCOM 2003)*, volume 2, pages 836–843 vol.2, 2003.
- [101] M. K. Lee, H. A. Latchman, R. E. Newman, S. Katar, and L. Yonge. Field

246 | Bibliografia

performance comparison of IEEE 802.11b and HomePlug 1.0. In *Proceedings of the 27th IEEE Conference on Local Computer Networks (LCN '02)*, volume 0, pages 598–599, 2002.