



Universidad de Zaragoza

Centro Politécnico Superior



Scheduling y provisión de QoS en redes móviles 4G OFDMA

Proyecto Fin de Carrera
Ingeniería de Telecomunicación

Autor: Luis Salvo Barco

Director: Dra. Ángela Hernández Solana



Departamento de
Ingeniería Electrónica
y Comunicaciones

Universidad Zaragoza

Zaragoza, Julio 2011

SCHEDULING Y PROVISIÓN DE QoS EN REDES MÓVILES 4G OFDMA

Resumen

En el desarrollo de las redes de nueva generación, destaca la tecnología LTE (Long Term Evolution), basada en la técnica de acceso OFDMA (Orthogonal Frequency Division Multiple Access). Una tarea importante en estos sistemas es el aprovechamiento de los recursos espectrales y de potencia disponibles garantizando unos determinados valores para los parámetros de calidad de servicio o QoS (Quality of Service) establecidos. Para conseguir este aprovechamiento existen diversas técnicas como la gestión del control de admisión y congestión, el control de las interferencias intercelulares, control de la movilidad, balanceo de carga, control de potencia, “scheduling” de paquetes, etc.

En este proyecto fin de carrera se evaluarán procedimientos de scheduling flexibles y eficientes capaces de explotar la diversidad frecuencial y temporal del canal, satisfaciendo las necesidades de los usuarios en la célula y proporcionando QoS diferenciada. Las propuestas analizadas y evaluadas incluyen criterios de ordenación dependientes de tasa, retardo y condiciones de canal. Se considerarán de forma explícita las restricciones impuestas por los mecanismos de mitigación de interferencias y control de potencia, suposiciones realistas con respecto a la capacidad de estimación del canal y los niveles de interferencia, la granularidad en la asignación de subportadoras y el conjunto de esquemas de modulación y codificación a utilizar. Se considerarán estructuras MAC compatibles con este modo de funcionamiento, y sobre ellas se procederá al estudio, adaptación y evaluación mediante simulación de los esquemas de gestión de recursos radio.

El objetivo es maximizar el número de usuarios soportables. Para ello se evaluarán los algoritmos de scheduling empleados en diversas situaciones, cambiando los tipos de tráfico y los periodos de actividad y silencio empleados para modelar el sistema. También se probará el efecto de la aplicación de un bloqueo inicial sobre los usuarios con peores condiciones de canal para los distintos algoritmos, comparando los resultados obtenidos con los que se sacaron sin este bloqueo y determinando en cada escenario la conveniencia o no de su utilización.

ÍNDICE DE CONTENIDOS

Resumen.....	3
Abreviaturas y acrónimos.....	11
Capítulo 1: INTRODUCCIÓN.....	13
1.1: Definición del proyecto.....	13
1.2: Objetivos y alcance del proyecto.....	14
Capítulo 2: DESCRIPCIÓN DEL SISTEMA.....	15
2.1: Consideraciones generales.....	15
2.2: Reúso fraccional de frecuencias (FFR).....	18
2.3: FFR ordenado con adaptación de potencia (FFRopa).....	19
2.4: Asignación de recursos y scheduling.....	21
Capítulo 3: ALGORITMOS PREVIOS AL PROYECTO.....	25
3.1: Retardo: Largest Delay First (LDF).....	25
3.2: Proportional Fair (PF).....	26
3.3: Modified Largest Weighted Delay First (MLWDF).....	27
3.4: Exponential Rule (EXP).....	28
Capítulo 4: ALGORITMOS DE SCHEDULING.....	29
4.1: Urgency and Efficiency-based Packet Scheduling: UEPS.....	29
4.2: Variante de MLWDF y EXP.....	32
4.3: Clasificación en grupos de urgencia.....	33
4.3.1: División en subgrupos según tiempo de vida del paquete.....	34
4.3.2: División en subgrupos según el estado de los buffers.....	34
4.4: Combinación de Proportional Fair y Retardo.....	35
Capítulo 5: RESULTADOS OBTENIDOS.....	37
5.1: Descripción del modelo de sistema.....	37
5.1.1: Parámetros generales.....	37
5.1.2: Selección de MCSs.....	39
5.1.3: Modelo de generación de CQIs.....	39
5.1.4: Modelo de tráfico.....	41
5.1.5: Modelo de retransmisión de TUs erróneos.....	42
5.2: Ráfagas A y B. Sin bloqueo.....	43
5.2.1: Ráfagas A. Tráfico uniforme.....	43
5.2.2: Ráfagas A. Escenario multiservicio.....	47
5.2.3: Ráfagas B. Tráfico uniforme.....	52
5.2.4: Ráfagas B. Escenario multiservicio.....	54
5.3: Ráfagas A y B. Bloqueo inicial.....	57
5.3.1: Bloqueo por retardo. Ráfagas A.....	58
5.3.2: Bloqueo por retardo. Ráfagas B.....	62
5.3.3: Bloqueo por estado de los buffers. Ráfagas A.....	66

Scheduling y provisión de QoS en redes móviles 4G OFDMA

5.3.4: Bloqueo por estado de los buffers. Ráfagas B.....	69
5.4: Conclusiones sobre los resultados.....	72
Capítulo 6: CONCLUSIONES Y LÍNEAS FUTURAS.....	75
Referencias.....	77
Anexo A: DESCRIPCIÓN DEL SIMULADOR.....	79
A.1: Estructura del simulador.....	79
Anexo B: RESULTADOS COMPLETOS AL APLICAR BLOQUEO INICIAL.....	85
B.1: Bloqueo por retardo.....	85
B.1.1: Ráfagas A. Tráfico uniforme.....	85
B.1.2: Ráfagas A. Escenario multiservicio.....	89
B.1.3: Ráfagas B. Tráfico uniforme.....	95
B.1.4: Ráfagas B. Escenario multiservicio.....	98
B.2: Bloqueo por estado de los buffers.....	104
B.2.1: Ráfagas A. Tráfico uniforme.....	104
B.2.2: Ráfagas A. Escenario multiservicio.....	107
B.2.3: Ráfagas B. Tráfico uniforme.....	113
B.2.4: Ráfagas B. Escenario multiservicio.....	116

ÍNDICE DE TABLAS

Tabla 5.1: Parámetros del sistema.....	38
Tabla 5.2: Modelo de propagación multicamino.....	38
Tabla 5.3: SINR umbral y tasa de datos para cada MCS.....	39
Tabla 5.4: Parámetros del modelo CQIs.....	41
Tabla 5.5: Perfiles de tráfico.....	42

ÍNDICE DE FIGURAS

Figura 2.1: RBs y scheduling tiempo-frecuencia.....	16
Figura 2.2: FFR con sectorización en el exterior.....	18
Figura 2.3: Asignación ordenada en la subbanda interior (FFRopa).....	19
Figura 2.4: Ejemplos de adaptación de la potencia en FFRopa.....	19
Figura 2.5: Esquema general del scheduling.....	22
Figura 4.1: Función utilizada para el algoritmo UEPS (servicios en tiempo real).....	30
Figura 4.2: Función $ U(t) $ para distintos valores de los parámetros a y c	31
Figura 4.3: Función $ U'(t) $ para distintos valores de los parámetros a y c	31
Figura 4.4: Clasificación en subgrupos de urgencia y reorganización.....	33
Figura 5.1: Modelo temporal de generación de CQIs.....	40
Figura 5.2: Resultados para los distintos métodos con tráfico uniforme (rA).....	44
Figura 5.3: Comparación de los distintos métodos con tráfico uniforme (rA).....	46
Figura 5.4.1: Resultados para los distintos métodos en un escenario multiservicio (50% $d_{max}=50ms$ +50% $d_{max}=250ms$) (rA) (I).....	49
Figura 5.4.2: Resultados para los distintos métodos en un escenario multiservicio (50% $d_{max}=50ms$ +50% $d_{max}=250ms$) (rA) (II).....	50
Figura 5.5: Comparación de los distintos métodos en un escenario multiservicio (50% $d_{max}=50ms$ +50% $d_{max}=250ms$) (rA).....	51
Figura 5.6: Resultados para los distintos métodos con tráfico uniforme (rB).....	52
Figura 5.7: Comparación de los distintos métodos con tráfico uniforme (rB).....	53
Figura 5.8: Resultados para los distintos métodos en un escenario multiservicio (50% $d_{max}=50ms$ +50% $d_{max}=250ms$) (rB).....	55
Figura 5.9: Comparación de los distintos métodos en un escenario multiservicio (50% $d_{max}=50ms$ +50% $d_{max}=250ms$) (rB).....	56
Figura 5.10: Algoritmos PF y Retardo con/sin bloqueo con tráfico uniforme (rA).....	59
Figura 5.11: Comparación con/sin bloqueo para los distintos algoritmos (rA). Tráfico uniforme.....	59
Figura 5.12: Algoritmos Retardo, 1-UEPS y PF con/sin bloqueo en un escenario multiservicio (50% $d_{max}=50ms$ +50% $d_{max}=250ms$) (rA).....	61
Figura 5.13: Comparación con/sin bloqueo para los distintos algoritmos (rA). Escenario multiservicio (50% $d_{max}=50ms$ +50% $d_{max}=250ms$)	62
Figura 5.14: Algoritmos Retardo, EXP, PF y 1-UEPS con/sin bloqueo con tráfico uniforme (rB).....	63
Figura 5.15: Resultados con/sin bloqueo para los distintos algoritmos (rB). Tráfico uniforme.....	64
Figura 5.16: Algoritmos Retardo, MLWDFvar y PF con/sin bloqueo en un escenario multiservicio (50% $d_{max}=50ms$ +50% $d_{max}=250ms$) (rB).....	65
Figura 5.17: Comparación con/sin bloqueo para los distintos algoritmos (rB). Escenario multiservicio (50% $d_{max}=50ms$ +50% $d_{max}=250ms$)	66
Figura 5.18: Algoritmos PF y Retardo con/sin bloqueo con tráfico uniforme (rA).....	67

Scheduling y provisión de QoS en redes móviles 4G OFDMA

Figura 5.19: Comparación con/sin bloqueo para los distintos algoritmos (rA). Tráfico uniforme.....	67
Figura 5.20: Algoritmo Retardo con/sin bloqueo en un escenario multiservicio (50% $d_{max}=50ms$ +50% $d_{max}=250ms$) (rA).....	68
Figura 5.21: Comparación con/sin bloqueo para los distintos algoritmos (rA). Escenario multiservicio (50% $d_{max}=50ms$ +50% $d_{max}=250ms$)	69
Figura 5.22: Algoritmos Retardo, MLWDF, EXP y 1-UEPS con/sin bloqueo con tráfico uniforme (rB).....	70
Figura 5.23: Resultados con/sin bloqueo para los distintos algoritmos (rB). Tráfico uniforme.....	70
Figura 5.24: Algoritmos Retardo y MLWDFvar con/sin bloqueo en un escenario multiservicio (50% $d_{max}=50ms$ +50% $d_{max}=250ms$) (rB).....	71
Figura 5.25: Comparación con/sin bloqueo para los distintos algoritmos (rB). Escenario multiservicio (50% $d_{max}=50ms$ +50% $d_{max}=250ms$)	72
Figura A.1: Estructura de clases del simulador.....	80
Figura A.2: Ejemplo de despliegue con wrap-around.....	81
Figura A.3: Representación de los 7 canales de cada UE.....	82
Figura B.1: Comparación entre algoritmos con/sin bloqueo con tráfico uniforme(rA). 86	
Figura B.2: Algoritmo MLWDFvar con tráfico uniforme (rA).....	87
Figura B.3: Algoritmo 1-UEPS con tráfico uniforme (rA).....	88
Figura B.4: Comparación bloqueos para los distintos algoritmos (rA). Tráfico uniforme..	
.....	89
Figura B.5: Comparación entre algoritmos con/sin bloqueo en un escenario multiservicio (50% $d_{max}=50ms$ +50% $d_{max}=250ms$) (rA). Tráfico de cota 50ms.....	90
Figura B.6: Comparación entre algoritmos con/sin bloqueo en un escenario multiservicio (50% $d_{max}=50ms$ +50% $d_{max}=250ms$) (rA). Tráfico de cota 250ms.....	91
Figura B.7: Algoritmo MLWDFvar en el escenario multiservicio (rA).....	92
Figura B.8: Algoritmo 1-UEPS en el escenario multiservicio (rA).....	93
Figura B.9: Comparación bloqueos para los distintos algoritmos (rA). Escenario multiservicio.....	94
Figura B.10: Comparación entre algoritmos con/sin bloqueo con tráfico uniforme (rB)...	
.....	95
Figura B.11: Algoritmo MLWDFvar con tráfico uniforme (rB).....	96
Figura B.12: Algoritmo 1-UEPS con tráfico uniforme (rB).....	97
Figura B.13: Comparación bloqueos para los distintos algoritmos (rB). Tráfico uniforme.	
.....	98
Figura B.14: Comparación entre algoritmos con/sin bloqueo en un escenario multiservicio (50% $d_{max}=50ms$ +50% $d_{max}=250ms$) (rB). Tráfico de cota 50ms.....	99
Figura B.15: Comparación entre algoritmos con/sin bloqueo en un escenario multiservicio (50% $d_{max}=50ms$ +50% $d_{max}=250ms$) (rB). Tráfico de cota 250ms.....	100
Figura B.16: Algoritmo MLWDFvar en el escenario multiservicio (rB).....	101
Figura B.17: Algoritmo 1-UEPS en el escenario multiservicio (rB).....	102

Figura B.18: Comparación bloqueos para los distintos algoritmos (rB). Escenario multiservicio.....	103
Figura B.19: Comparación entre algoritmos con/sin bloqueo con tráfico uniforme (rA)..	104
Figura B.20: Algoritmo MLWDFvar con tráfico uniforme (rA).....	105
Figura B.21: Algoritmo 1-UEPS con tráfico uniforme (rA).....	106
Figura B.22: Comparación bloqueos para los distintos algoritmos (rA). Tráfico uniforme.	107
Figura B.23: Comparación entre algoritmos con/sin bloqueo en un escenario multiservicio (50% $d_{max}=50ms$ +50% $d_{max}=250ms$) (rA). Tráfico de cota 50ms.....	108
Figura B.24: Comparación entre algoritmos con/sin bloqueo en un escenario multiservicio (50% $d_{max}=50ms$ +50% $d_{max}=250ms$) (rA). Tráfico de cota 250ms.....	109
Figura B.25: Algoritmo MLWDFvar en el escenario multiservicio (rA).....	110
Figura B.26: Algoritmo 1-UEPS en el escenario multiservicio (rA).....	111
Figura B.27: Comparación bloqueos para los distintos algoritmos (rA). Escenario multiservicio.....	112
Figura B.28: Comparación entre algoritmos con/sin bloqueo con tráfico uniforme (rB)...	113
Figura B.29: Algoritmo MLWDFvar con tráfico uniforme (rB).....	114
Figura B.30: Algoritmo 1-UEPS con tráfico uniforme (rB).....	115
Figura B.31: Comparación bloqueos para los distintos algoritmos (rB). Tráfico uniforme.	116
Figura B.32: Comparación entre algoritmos con/sin bloqueo en un escenario multiservicio (50% $d_{max}=50ms$ +50% $d_{max}=250ms$) (rB). Tráfico de cota 50ms.....	117
Figura B.33: Comparación entre algoritmos con/sin bloqueo en un escenario multiservicio (50% $d_{max}=50ms$ +50% $d_{max}=250ms$) (rB). Tráfico de cota 250ms.....	118
Figura B.34: Algoritmo MLWDFvar en el escenario multiservicio (rB).....	119
Figura B.35: Algoritmo 1-UEPS en el escenario multiservicio (rB).....	120
Figura B.36: Comparación bloqueos para los distintos algoritmos (rB). Escenario multiservicio.....	121

ABREVIATURAS Y ACRÓNIMOS

1-UEPS	↔	UEPS modificado
3GPP	↔	3rd Generation Partnership Project
ARQ	↔	Automatic Repeat request
AWGN	↔	Additive White Gaussian Noise
BLER	↔	Block Error Rate
BS	↔	Base Station
BWA	↔	Broadband Wireless Access
CDMA	↔	<i>Code Division Multiple Access</i>
CINR	↔	Channel to Interference and Noise Ratio
CQI	↔	Channel Quality Indicator
DL	↔	DownLink
eNB	↔	evolved Node B
E-UTRA	↔	Evolved UMTS Terrestrial Radio Access network
EXP	↔	EXponential rule
FD	↔	Frequency Domain
FIFO	↔	First Input First Output
FFR	↔	Fractional Frequency Reuse
FFRopa	↔	FFR with ordering and power adaptation
FFT	↔	Fast Fourier Transform
HARQ	↔	Hybrid ARQ
HoL	↔	Head of Line
ICI	↔	InterCarrier Interference
ICIC	↔	InterCarrier Interference Coordination
LDF	↔	Largest Delay First
LTE	↔	Long Term Evolution
MAC	↔	Medium Access Control
MCS	↔	Modulation and Coding Scheme
MLWDF	↔	Modified Largest Weighted Delay First
MLWDFvar	↔	Variante de MLWDF y EXP
ms	↔	Milisegundos
npaq3reorg	↔	Algoritmo de división en subgrupos según el estado de los buffers
OFDMA	↔	Orthogonal Frequency-Division Multiple Access
PF	↔	Proportional Fair
PFret	↔	Algoritmo que usa como criterio PF*retardo
PSA	↔	Power and Subcarrier Allocation

Scheduling y provisión de QoS en redes móviles 4G OFDMA

QAM	↔	<i>Quadrature amplitude modulation</i>
QoS	↔	Quality of Service
QPSK	↔	Quadrature Phase Shift Keying
rA	↔	Ráfagas tipo A
rB	↔	Ráfagas tipo B
RB	↔	Resource Block
Retardo	↔	Algoritmo de scheduling por ratio entre el tiempo de espera en cola y la cota máxima del retardo
RRM	↔	Radio Resource Management
SC-FDMA	↔	Single-Carrier Frequency-Division Multiple Access
SINR	↔	Signal to Interference and Noise Ratio
SPUC	↔	Signal Processing Unit C++
TD	↔	Time Domain
tem3reorg	↔	Algoritmo de división en subgrupos según tiempo de vida del paquete
TTI	↔	Time Transmission Interval
TU	↔	Transport Unit
UE	↔	User Equipment
UEPS	↔	Urgency and Efficiency-based Packet Scheduling
UEPSder	↔	UEPS aplicado con la fórmula de la literatura (usando $ U'(t) $)
UL	↔	UpLink
WiMAX	↔	Worldwide Interoperability for Microwave Access

1.INTRODUCCIÓN

1.1: Definición del proyecto

En la actualidad, en el mundo de la telefonía móvil, que continúa su constante evolución, se está desarrollando la cuarta generación de tecnologías (4G). Siguiendo la norma del 3rd Generation Partnership Project (3GPP), Long Term Evolution (LTE) será el estándar para el acceso radio. El incesante incremento en el uso y en la demanda de servicios multimedia en movilidad ha hecho que el desarrollo de tecnologías móviles celulares de banda ancha (BWA) sea extremadamente importante para los operadores y para los organismos de estandarización.

LTE será la clave para el despegue del internet móvil o servicios como la transmisión de datos a más de 300 metros y videos de alta definición. Lo novedoso de LTE es la interfaz radioeléctrica basada en Orthogonal Frequency Division Multiple Access (OFDMA) para el enlace descendente (DL) y Single-Carrier FDMA (SC-FDMA) para el enlace ascendente (UL). El uso de estas técnicas con capas físicas potentes y flexibles hace que los parámetros de las diferentes conexiones puedan adaptarse a las condiciones del canal, niveles de interferencia, carga del sistema o variaciones de las fuentes de tráfico. Pero para tener eficiencia en la provisión de calidad de servicio (QoS), es básico un diseño eficaz de la gestión y asignación de recursos radio (RRM). La implementación de estos mecanismos no está en las especificaciones, dejando libre su desarrollo.

El entorno inalámbrico es muy hostil para la provisión de calidad de servicio debido a su variabilidad con el tiempo. Esto representa un serio desafío para el soporte de servicios con restricciones de retardo máximo en sistemas inalámbricos. Los sistemas de comunicaciones ya estandarizados con restricciones QoS en entornos inalámbricos sólo pueden garantizar los requisitos para un porcentaje de los casos. Esto implica una caída del servicio, generando los cortes de llamadas o los mensajes de “red ocupada”. A la hora de diseñar un sistema inalámbrico, el objetivo es poder garantizar la QoS al mayor número posible de usuarios del sistema, utilizando para ello diversas técnicas. En este contexto, el diseño eficiente de estrategias de scheduling dependientes de las condiciones de canal y del tráfico (incluida la selección adaptativa de tasa y/o potencia de transmisión) es un aspecto determinante, al igual que lo es el diseño de esquemas de gestión de interferencias. De hecho, la mejora de la capacidad del sistema pasa por el diseño y aplicación conjunta y coordinada de todos estos mecanismos.

Mediante las técnicas de scheduling se busca realizar el mejor reparto posible de los recursos del sistema, buscando en todo momento el satisfacer los requerimientos de los usuarios en términos de tasa, cotas de retardo o tasa de descarte de paquetes. Con dos componentes (temporal y frecuencial), los algoritmos de scheduling temporal sirven para determinar en qué orden se sirve a los usuarios, mientras que la componente frecuencial busca obtener el máximo partido de la diversidad multiusuario y multicanal en sistemas OFDMA.

1.2: Objetivos y alcance del proyecto

El objetivo de este proyecto, centrado en el enlace descendente de un sistema basado en LTE, es la evaluación de procedimientos de scheduling flexibles y eficientes capaces de explotar la diversidad frecuencial y temporal del canal, satisfaciendo las necesidades de los usuarios en la célula y proporcionando calidad de servicio diferenciada. Se quieren proponer estrategias más óptimas desde el punto de vista de utilización de los recursos del sistema y de la provisión de QoS en función del tipo de servicio.

Las propuestas que se han analizado incluyen criterios de asignación dependientes de tasa, retardo y condiciones del canal. Los algoritmos de scheduling empleados deben ser capaces de efectuar una selección adaptativa del esquema de codificación más adecuado en función de los recursos físicos asignados a cada usuario y de las condiciones de canal e interferencia observados por éste en ellos.

Para evaluar estas propuestas se ha utilizado un simulador implementado en C++, en el que se incorporaron las distintas propuestas de scheduling, permitiendo el análisis pormenorizado y detallado de las mismas. En las simulaciones se contemplaron evaluaciones realistas, modelando de forma precisa la capa física, aplicando los modelos de propagación formulados en las recomendaciones del 3GPP y teniendo en cuenta las limitaciones de potencia de transmisión de las estaciones base, los distintos grados de movilidad de los usuarios y los diversos tipos de tráfico.

El simulador tenía ya programados varios algoritmos de scheduling, por lo que el trabajo consistió en buscar nuevos procedimientos que mejoraran los resultados de los anteriores, programarlos convenientemente en el simulador y, tras comprobar su correcto funcionamiento, aguardar a los resultados arrojados por el mismo.

A lo largo de esta memoria se procurará introducir el sistema sobre el que se aplicarán estos algoritmos, y mostrar y comentar los resultados que con ellos se obtienen, haciendo las pertinentes comparaciones con los que previamente se hallaban en el simulador, para usar éstos como referencia.

2. DESCRIPCIÓN DEL SISTEMA

En esta sección se intenta introducir al lector a las condiciones del medio inalámbrico. Para ello, en el apartado 2.1 se explican de un modo general los parámetros tenidos en cuenta y el modo en que éstos se han tratado por medio de diversas estrategias: tecnología de acceso, control de interferencias, gestión de potencia, control de errores, scheduling, etc.

A partir del apartado 2.2 se centra en los algoritmos de gestión de interferencias y asignación de recursos. En la elaboración de este proyecto finalmente se utilizó solamente el algoritmo FFR_{opa} [1], variante del conocido FFR (Fractional Frequency Reuse) [2] combinado con los diferentes algoritmos de scheduling que más adelante se desarrollarán, pero se considera necesaria una explicación del FFR para su comprensión.

2.1: Consideraciones generales

OFDMA ha sido seleccionada para ser la tecnología de acceso múltiple para los sistemas inalámbricos de banda ancha como LTE o WiMAX por su gran flexibilidad y escalabilidad para dar cabida a un gran número de usuarios con diferentes demandas de tráfico y condiciones de canal. El sistema planteado tendrá reuso frecuencial unidad, es decir, todas las células usarán el mismo ancho de banda y a la misma frecuencia, de forma que se consigue una alta eficiencia espectral.

Esto hace que el sistema requiera el uso de esquemas de coordinación de ICI (ICIC) para mejorar el rendimiento en la parte exterior de la célula. Los esquemas de ICIC definen una serie de restricciones en cuanto a la asignación de recursos, generalmente frecuencia y potencia, que han de ser tenidos en cuenta. El objetivo de usar al mismo tiempo ICIC y algoritmos de asignación dinámica de potencia y subportadoras (PSA) es conseguir un mejor rendimiento en cuanto a la gestión de las interferencias y uso de recursos.

El reuso fraccional de frecuencias (FFR) es un esquema de ICIC ampliamente aceptado en redes OFDMA con reuso frecuencial unidad. Haciendo uso del mismo, el ancho de banda del sistema se divide en dos subbandas, una interior y una exterior. En la subbanda interior se aplica reuso unidad, es decir, los bloques de recursos (RBs) de dicha subbanda pueden ser ocupados por usuarios de cualquier sector de la celda. En la subbanda exterior, por el contrario, se aplica sectorización de tal forma que sólo los

Scheduling y provisión de QoS en redes móviles 4G OFDMA

usuarios (UEs: User Equipment) del sector pueden ocupar los RBs asignados en la subbanda exterior a dicho sector. Además, la potencia de transmisión en la subbanda exterior es más alta que en la subbanda interior, con el fin de alcanzar a los usuarios situados en el límite de la celda.

La flexibilidad de asignación que proporciona en la zona central de la célula incrementa las fluctuaciones de ICI, ya que la mayor ICI puede venir de cualquier sector de las células vecinas. Además, cambios rápidos en la ocupación de las frecuencias puede causar un efecto ON-OFF en la potencia transmitida. Esto conduce a una alta inexactitud en la estimación de la ICI, que eleva el Block Error Rate (BLER) y, en consecuencia, el ratio de descarte de paquetes y los retardos de transmisión.

Para superar estos inconvenientes, se propone un algoritmo PSA denominado FFRopa, que mejora la estabilidad de la ICI y reduce el descarte de paquetes y el retardo mediante la aplicación de un enfoque coordinado para ordenar la asignación de recursos de la frecuencia.

Los recursos radio son generalmente asignados a usuarios activos de acuerdo a alguna política de scheduling de paquetes. Se requiere una adecuada combinación del algoritmo PSA con el scheduling para lograr eficientemente los requerimientos de QoS y obtener cierta ganancia en diversidad que incremente el throughput global del sistema.

El scheduling temporal-frecuencial de paquetes [3], con la asignación de potencia y la selección adaptativa del esquema de modulación y codificación (MCS) están localizados en la estación base (BS). Cada célula realiza independientemente sus decisiones de scheduling teniendo en cuenta las limitaciones del esquema de ICIC. Los recursos frecuenciales del sistema se dividen en *resource blocks* (RBs), siendo éstos la menor cantidad posible de recursos asignables (Figura 2.1). Un RB contiene *NSC* subportadoras OFDM, y se usa la misma potencia de transmisión y MCS para todas las subportadoras de un mismo RB. En nuestro caso se utilizaron RBs de 12 subportadoras OFDM, separadas entre sí $\Delta f=15$ kHz.

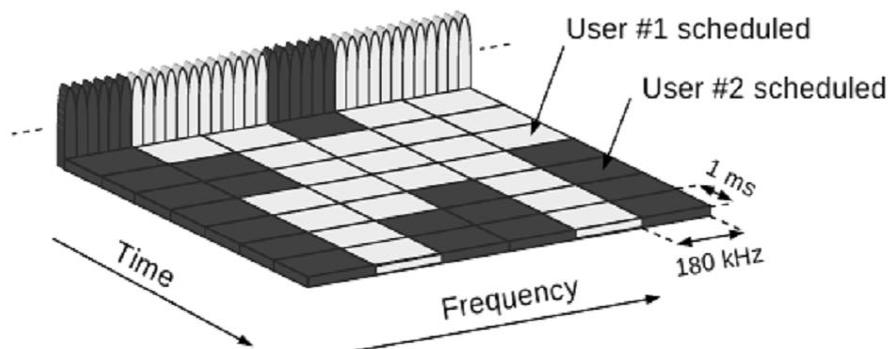


Figura 2.1: RBs y scheduling tiempo-frecuencia

Scheduling y provisión de QoS en redes móviles 4G OFDMA

Existen limitaciones para el máximo de potencia de transmisión para cada RB (P_{RB}^{max}) y la máxima potencia de salida en el conjunto de la célula (P_T^{max}). Así, siendo P_i la potencia transmitida en el RB i y P_{out} la potencia total en la célula, estas limitaciones imponen que:

$$P_{out} = \sum_{i=1}^{N_{RB}} P_i \leq P_T^{max} \quad \text{con} \quad P_i \leq P_{RB}^{max} \quad (11.1)$$

Cumpliendo esta ecuación, el valor de la potencia transmitida en cada sector nunca es superior a P_T^{max} , aun cuando todos los RBs sean asignados. Aunque a priori sería posible una adaptación conjunta de potencia y MCS según las condiciones del canal para mejorar la eficiencia global del sistema. Para limitar las fluctuaciones de ICI, se recomienda que la potencia transmitida esté entre unos valores dados por una máscara.

La información sobre las condiciones del canal la envía cada usuario usando el indicador de la calidad del canal (CQI), incluyendo el ratio canal a ruido e interferencias (CINR) en cada RB. Multiplicando el valor de la CINR por la potencia de transmisión se obtiene la relación señal a ruido e interferencia (SINR) esperada para cada RB. De acuerdo a las condiciones de canal de cada usuario, la MCS objetivo se define como la MCS más baja que permite transmitir los datos requeridos por el usuario en el mínimo número posible de RBs, satisfaciendo siempre los requisitos de SINR. Una vez asignados los recursos de frecuencia y potencia de transmisión necesarios para servir las tasas de datos requeridas por cada usuario, se asignan más recursos a aquellos usuarios que aún tengan datos para mejorar su QoS y la eficiencia espectral o bien para recuperar retrasos en el tráfico.

Los mecanismos de ARQ son protocolos utilizados para el control de errores en la transmisión de datos, garantizando la integridad de los mismos. El mecanismo de ARQ aplicado consiste básicamente en la retransmisión de los bloques de datos que han sido recibidos erróneamente. Cada conexión lógica de datos entre un móvil y su estación base tiene dos buffers distintos, ambos con política FIFO: el buffer de salida, en el que se almacenan las nuevas unidades de transporte (TUs) de datos pendientes de ser transmitidos; y el buffer de ARQ, que contiene los TUs a retransmitir porque fueron recibidos erróneamente. La prioridad de un paquete de ARQ es siempre superior a la de los bloques de datos, con lo que si un móvil recibe una asignación de recursos, los bloques de ARQ son los que preferentemente serán transmitidos.

2.2: Reúso fraccional de frecuencias (FFR)

El algoritmo de asignación de recursos radio fraccional *FFR* es la aplicación directa del método de coordinación basado en reúso fraccional con sectorización en la subbanda exterior (Figura 2.2). Los usuarios con peores condiciones de canal recibirán servicio preferiblemente en la subbanda exterior, ya que al transmitir con mayor potencia y hacer uso de sectorización, presenta unas condiciones más favorables para ellos. En la zona interior, por el contrario, no existen restricciones en la asignación de RBs, lo que permite ajustar los RBs asignados a cada sector en función de su carga. De todos modos, la potencia de transmisión en cada RB sólo se transmitirá en el sector al que pertenece el UE receptor del RB. Por tanto, la interferencia en un RB determinado no proviene siempre del mismo sector, lo que introduce un nuevo efecto de variabilidad en la interferencia que provocará un mayor número de errores a la hora de estimar las condiciones en las que se encuentra el usuario.

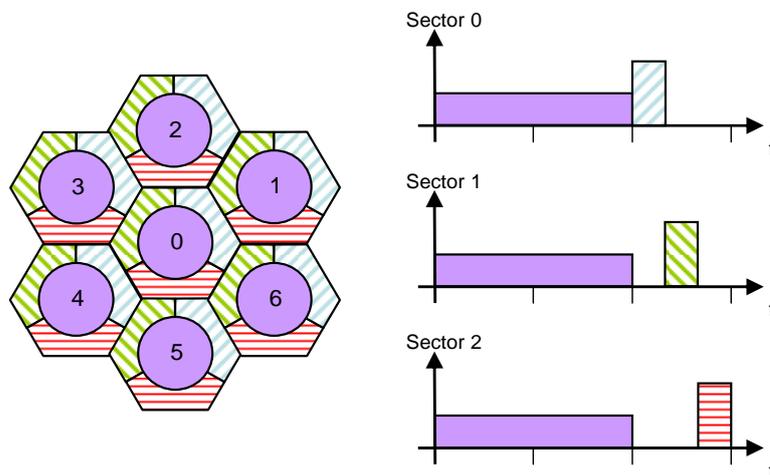


Figura 2.2: FFR con sectorización en el exterior

Uno de los puntos a favor del sistema fraccional con sectorización en el exterior es que simplifica la coordinación entre celdas. El esquema resulta más flexible para adaptarse a distribuciones heterogéneas de UEs entre sectores dentro de la celda. Esto no sucede en los sistemas de coordinación con ordenación de las asignaciones en el interior, puesto que se ha de decidir a priori qué parámetros concretos debe aplicar cada celda, siendo de este modo el cluster la estructura fundamental a replicar en el despliegue. Por tanto, el sistema de reúso fraccional aporta un despliegue y una gestión de red más sencilla, pero a costa de un empeoramiento de las prestaciones ofrecidas.

2.3: FFR ordenado con adaptación de potencia (FFRopa)

En el algoritmo FFR antes referido los RBs se van asignando ordenadamente a partir de un RB inicial a los UEs del sector, y además hace uso de una asignación de potencia constante, es decir, la misma potencia de transmisión es asignada a todos los RBs de la misma subbanda (M_{med}). Al evitar las variaciones en la potencia transmitida se reduce naturalmente la variabilidad de la interferencia. Sin embargo, esta restricción en potencia supone que determinados usuarios no reciban una asignación porque su potencia requerida exceda, aunque sea en una pequeña cantidad, el valor de la máscara M_{med} . Es por esto que se define el algoritmo *FFRopa*, que se basa en una asignación ordenada de los RBs de la subbanda interior, emulando un reuso 1/3 flexible, al tiempo que se aplica un ajuste acotado de la potencia (Figuras 2.3 y 2.4). De este modo, cuando un usuario requiera una potencia ($P_{k,i}$) inferior a M_{med} (UE0), la diferencia $M_{med} - P_{k,i}$ contribuye a crear un margen de potencia para asignar una potencia de transmisión superior a M_{med} a usuarios en peores condiciones (UE1). La adaptación de potencia se circunscribe únicamente a la subbanda interior, puesto que en la exterior es preferible estabilizar la interferencia lo máximo posible, con lo que se mantiene la asignación con potencia constante.

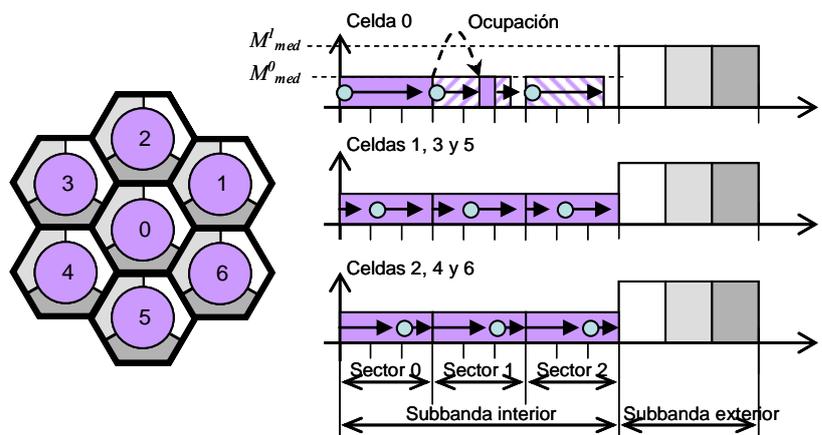


Figura 2.3: Asignación ordenada en la subbanda interior (FFRopa)

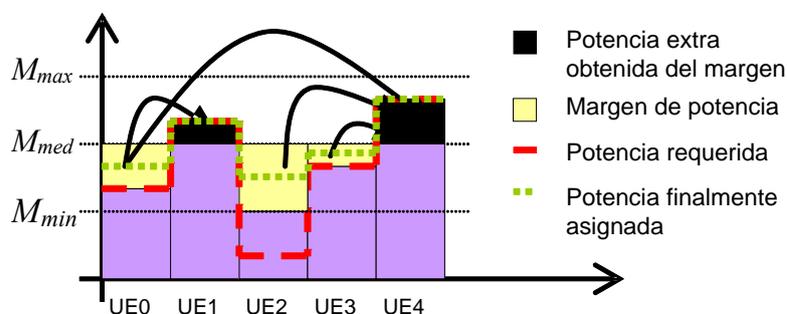


Figura 2.4: Ejemplos de adaptación de la potencia en FFRopa

Scheduling y provisión de QoS en redes móviles 4G OFDMA

Con el fin de mantener una baja variabilidad de la interferencia en la subbanda interior, la potencia finalmente asignada sólo puede variar entre dos valores prefijados (M_{min} y M_{max}). De este modo, si la potencia requerida es inferior a M_{min} , el algoritmo asume que la potencia requerida es M_{min} , como ocurre con el UE2. Por otro lado, cuando la potencia requerida excede M_{max} , el usuario no podrá ser asignado y deberá intentar ser ubicado en la subbanda de mayor potencia.

Se ha juzgado conveniente realizar la asignación de potencia definitiva al final, de modo que se puede ajustar a las asignaciones realizadas. En este sentido, mientras los usuarios van siendo asignados, la potencia preasignada a ellos es M_{med} si su potencia requerida es inferior a M_{med} , o directamente la potencia requerida $P_{k,i}$ si la excede, mientras que el margen de potencia logrado se va actualizado de acuerdo a la potencia realmente requerida por los usuarios.

Una vez se han realizado todas las asignaciones de RBs, se procede a reajustar la potencia preasignada de modo que no se exceda el límite máximo de potencia transmitida en la celda. Para ello, comenzamos computando la potencia total asignada hasta el momento. Si esta potencia no excede el límite máximo, porque ningún usuario ha excedido M_{med} o hay RBs sin asignar que compensan el exceso de potencia, se valida la preasignación de potencia realizada. Así, los usuarios con buenas condiciones de canal experimentarán un gran margen frente a interferencias, lo que permitirá incrementar su MCS para transmitir de forma más eficiente. Cuando la potencia total preasignada exceda el límite máximo permitido en la celda, es necesario extraer potencia de los usuarios con mejores condiciones de canal. En nuestro caso, hemos optado por extraer potencia de todos los UEs con una potencia requerida inferior a M_{med} de forma proporcional al margen de interferencia que experimentan. Así, más potencia es tomada de los UEs con mayor margen y viceversa, lo que supone un tratamiento más igualitario de los usuarios.

En cuanto a la asignación ordenada a partir de un RB inicial, sucede que cuando un UE no puede recibir servicio por condiciones de canal en el RB que le corresponde por la ordenación, no recibe asignación y debe esperar a que lleguen a habilitarse otros RBs para recibirla. Este mecanismo es ineficaz puesto que se impide innecesariamente la asignación de UEs prioritarios por el simple hecho de que no pueden recibir servicio en el RB que les corresponde.

Para evitar este bloqueo, la versión de FFRopa aquí utilizada parte de un grupo inicial de RBs que pueden ser asignados en cada subbanda sin seguir un ordenamiento predefinido: al inicio del algoritmo se realiza una estimación de los RBs que van a ser necesarios para la asignación de los UEs que han solicitado acceso. Dado que es improbable un cambio brusco en el número de UEs activos entre dos intervalos de transmisión (TTIs) consecutivos, el número de RBs ocupados en cada subbanda en el TTI anterior es una buena estimación de los RBs necesarios en el presente TTI. Así, si

en la subbanda preasignada al sector 0 se ocuparon en el TTI anterior un total de N RBs, en el TTI actual se asume que los primeros N RBs de la subbanda van a ser asignados. Por tanto, los primeros UEs elegirán sin ningún tipo de orden los RBs que necesiten de entre este grupo inicial. Una vez ocupados todos ellos, se procede a la asignación del resto de RBs de la subbanda del sector siguiendo su ordenamiento.

De forma resumida, el algoritmo se ejecuta de forma similar a FFR pero con un grupo inicial de RBs que pueden ser asignados en cada subbanda sin seguir un ordenamiento predefinido, y con la gestión de potencia descrita anteriormente: un usuario recibirá una asignación aunque su potencia requerida supere la máscara media, siempre que exista un margen de potencia anterior suficiente. Finalmente se aplica el reajuste de potencia con el fin de que la potencia total asignada no supere el límite máximo de la celda.

2.4: Asignación de recursos y scheduling

Como ya se ha especificado, el proceso de scheduling está centralizado en la BS, y se lleva a cabo con una periodicidad de 1ms. La asignación de recursos (frecuencia, potencia y tasa de transmisión), se realiza teniendo en cuenta información de la calidad del canal CQI y otros parámetros dependientes del tráfico, como el estado de los buffers, y la prioridad de los usuarios, que dependerá, entre otros factores, de los parámetros de calidad de servicio del tipo de servicio.

Con dos dominios claros de priorización, el dominio temporal (TD: Time Domain) y el frecuencial (FD: Frequency Domain), la función del scheduling incluye la selección de los bloques físicos asignables, junto con la selección del esquema de modulación y codificación más apropiado y la potencia de transmisión. El dominio temporal trata de determinar los UEs que deben recibir servicio en un determinado periodo de scheduling (atendiendo a criterios de tráfico o de calidad de canal), mientras que el dominio frecuencial establece la asignación específica de recursos frecuenciales a los UEs previamente seleccionados o priorizados.

Por lo general, el objetivo del scheduling temporal es conseguir un equilibrio entre el throughput global y un reparto equitativo de recursos, mientras que el scheduling frecuencial trata de maximizar el throughput. Aunque las decisiones de scheduling TD y FD pueden ser secuenciales, con frecuencia ambos criterios aparecen de forma conjunta en la definición de las estrategias de scheduling. La interacción entre los distintos factores involucrados se ilustra en la Figura 2.5.

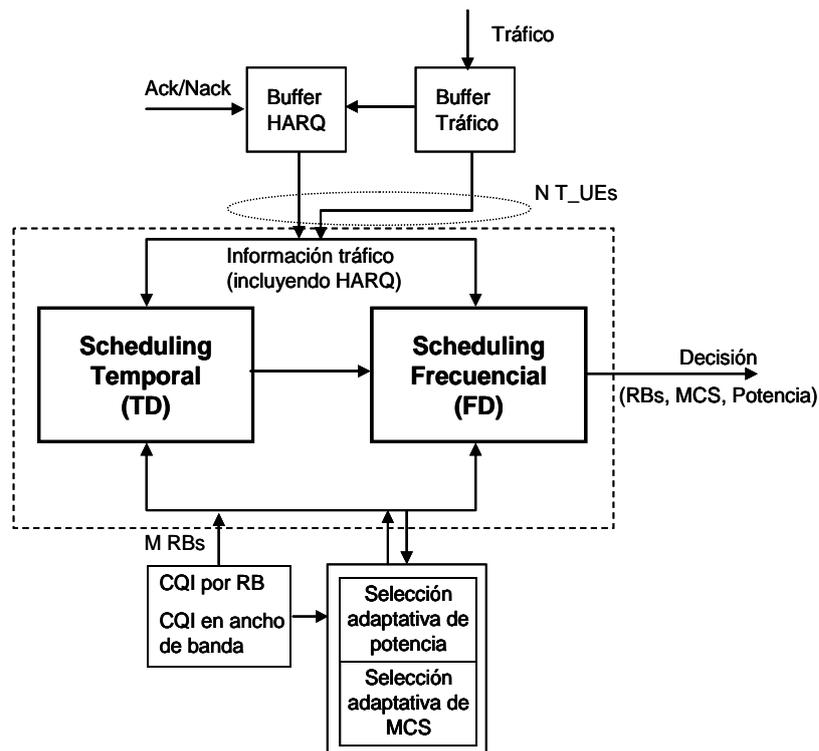


Figura 2.5: Esquema general del scheduling

Considerando que ha de adaptarse a las condiciones de canal, de tráfico y requisitos de QoS, son muchos los factores que condicionan el funcionamiento del scheduling, como por ejemplo la capacidad de disponer de información fiable sobre las condiciones de canal (y del estado de los buffers) y la capacidad de señalización. En relación al primer punto, hay que tener en cuenta que el uso de la información del CQI para la adaptación de MCS y scheduling dinámico está limitado por la calidad de la información proporcionada. Ésta depende de los errores en la medida, condicionada por el número de pilotos de referencia utilizados en la ventana de medida; de la pérdida de información debida al formato de envío de CQI elegido, del retardo en la transmisión, recepción y procesado de los datos, y de los errores en la recepción del CQI, entre otros factores.

Por lo que respecta a la señalización, el número de usuarios que puede recibir recursos para transmitir en cada TTI está limitado por el número de canales que transportan la información de scheduling para ambas direcciones de enlace (UL, DL).

Interesa la implementación de estrategias de asignación de recursos que contemplen de forma conjunta el dominio frecuencial y temporal, de acuerdo con el diagrama de la Figura 2.5.

El enfoque más sencillo sería establecer en primer lugar el conjunto de UEs que deberían recibir servicio en un periodo de scheduling (atendiendo, por ejemplo, a la

Scheduling y provisión de QoS en redes móviles 4G OFDMA

cantidad de servicio recibido y los parámetros de QoS negociados), para posteriormente asignar los recursos frecuenciales a dichos usuarios en función de las condiciones de canal. El scheduling temporal, en primera instancia, selecciona los N usuarios que van a pasar a una segunda fase de scheduling frecuencial, que se encargará de asignar los recursos disponibles, de la forma más conveniente, entre esos UEs, dando prioridad a las retransmisiones sobre las primeras transmisiones. El principal inconveniente de este planteamiento radica en que en realidad ambas fases no deberían ser secuenciales, ya que la solución óptima implica la consideración conjunta de los criterios que aplican el scheduling temporal y frecuencial.

En las siguientes secciones se explicarán diversos algoritmos de scheduling y se compararán los resultados obtenidos en cuanto a descarte de paquetes respecto al número de usuarios en cada célula. Algunos algoritmos se encontraban en el sistema anteriormente a la realización de este proyecto, por lo que se explicarán por separado, pero se usarán sus resultados para compararlos con los que se han desarrollado durante el mismo.

A la hora de valorar el comportamiento de las propuestas específicas de gestión de recursos a nivel de enlace que se proponen es importante resaltar que éstas consideran de forma explícita las restricciones impuestas por los mecanismos de mitigación de interferencias y control de potencia, suposiciones realistas con respecto a la capacidad de estimación del canal y de los niveles de interferencia, la granularidad en la asignación de subportadoras y en el conjunto de esquemas de modulación y codificación a utilizar.

3. ALGORITMOS PREVIOS AL PROYECTO

Esta sección está dedicada a recopilar y explicar los algoritmos que ya se hallaban implementados en el sistema antes de comenzar con la elaboración de este proyecto. El sistema comentado en la sección anterior se probó sobre un simulador cuyo esquema puede encontrarse en el Anexo A. Todos los algoritmos aquí comentados y los realizados con motivo de este proyecto también se programaron sobre este mismo simulador.

Para simplificar el análisis, se establecen algunos criterios, que serán aplicados a todos ellos. En primer lugar, siendo $\hat{r}_{j,k}[n]$ el throughput, o tasa instantánea estimada para el usuario j -ésimo en el RB k -ésimo, dicha estimación se asumirá que se obtiene en base a los reportes de CQI enviados por el UE a la estación base. La estimación de la tasa, se efectuará suponiendo una distribución idéntica de potencia entre todos los RBs del ancho de banda del canal. Con este planteamiento, se focaliza el scheduling en la asignación dinámica de tasa. En todo caso, si se contemplase la posibilidad de asignar distintos niveles de potencia, debería garantizarse en todo momento que la potencia total de transmisión en la estación base no supere el límite máximo establecido en las especificaciones del sistema. Se considera en todos los casos que el número de usuarios que esperan recibir servicio es N y el número de RBs disponibles es M .

3.1: Retardo: Largest Delay First (LDF)

La asignación de RBs se adapta en función del retardo que experimentan los paquetes de cada UE. Siendo W_j el retardo acumulado del paquete situado en la cabecera de la cola del usuario UE j , para cada RB se escoge el UE que presenta el retardo más alto (ecuación III.1). Una vez asignado el RB, el valor de W_j se actualiza con el tiempo de espera del siguiente paquete en cola. Con este planteamiento es posible asignar varios RBs al mismo UE siempre que sea éste el UE con el retardo más alto. El scheduler solo necesita manejar marcas temporales relativas al instante de generación de nuevos paquetes en la cola. En cualquier caso, la asignación del RB en un sistema real sólo se produciría en el caso de que la relación señal a ruido estimada en recepción sea capaz de satisfacer alguno de los requisitos de SINR establecidos para el conjunto acotado de MCS elegibles en el sistema. En caso contrario, la asignación no sería posible y se continuaría con el siguiente UE más prioritario.

$$\hat{j} = \arg \max_j \{W_j[n]\} \quad (III.1)$$

Cálculo de W_j para todos los UEs

Mientras haya RBs

{

$$\hat{j} = \arg \max_j \{W_j[n]\}$$

Si no es posible asignar un RB continúa en el siguiente UE más prioritario

}

Este criterio se modificó un poco para no dar una mayor ventaja a aquellos paquetes con una mayor cota máxima de retardo, ya que pueden acumular más retardo y en ese caso tendrían siempre prioridad sobre aquellos con una cota menor. Para hacerlo más justo y adaptable a distintos tipos de tráfico, se decidió ponderar el tiempo acumulado de espera por la cota máxima de retardo que el paquete admite (ecuación III.2).

$$\hat{j} = \arg \max_j \left\{ \frac{W_j[n]}{d_{\max}} \right\} \quad (III.2)$$

3.2: Proportional Fair (PF)

Proportional Fair (“PF”): esta estrategia de selección prioriza al usuario con mayor ratio entre la tasa promedio alcanzable por el usuario j , $\bar{r}_j[n]$, en los RBs que le pueden ser asignados en el TTI n -ésimo (calculada con estimación ideal del canal) frente a la tasa promedio servida $T_j[n]$. Este parámetro se calcula de forma recursiva considerando un intervalo de filtrado de la estimación de longitud t_f intervalos de scheduling tal y como se refleja en la ecuación III.4 [4]. La tasa promedio se calcula de acuerdo con la expresión III.3. [5]

$$\hat{j} = \arg \max_j \left\{ \frac{\hat{r}_j[n]}{T_j[n]} \right\} \quad (III.3)$$

$$T_j[n] = \begin{cases} \left(1 - \frac{1}{t_f}\right) T_j[n-1] + \left(\frac{1}{t_f}\right) \hat{r}_j[n] & \text{si el UE } j \text{ recibe servicio en el instante } n \\ \left(1 - \frac{1}{t_f}\right) T_j[n-1] & \text{en otro caso} \end{cases} \quad (III.4)$$

Scheduling y provisión de QoS en redes móviles 4G OFDMA

Con este planteamiento se reduce la probabilidad de que un usuario con malas condiciones de canal no reciba servicio. Es posible contemplar dos posibles alternativas de implementación. En la primera, el valor de $T_j[n]$ se actualiza únicamente al inicio de cada nuevo periodo de scheduling. En la segunda, el valor de $T_j[n]$ se actualiza cada vez que se produce la asignación de un RB.

Cálculo de $T_j[n]$ para todos los UEs

Mientras haya RBs

{

$$\hat{j} = \arg \max_j \left\{ \frac{\hat{r}_{j,k}[n]}{T_j[n]} \right\}$$

Actualizar $T_j[n]$ (versión2)

Si no es posible asignar un RB continúa en el siguiente UE más prioritario

}

3.3: Modified Largest Weighted Delay First (MLWDF)

Es uno de los algoritmos de scheduling más populares para servicios en tiempo real. Tiene en cuenta además de los parámetros de calidad de servicio (retardo y tasa), las condiciones de canal. Elige la mejor combinación de retraso-tasa sirviendo al UE que maximiza la ecuación (III.5), donde j es el índice del usuario seleccionado y k el del RB, aplicando el siguiente algoritmo. [6]

Cálculo de $W_j[n]$ para todos los UEs

Mientras haya RBs disponibles

{

$$\{\hat{j}, \hat{k}\} = \arg \max_{j,k} \{ F(j,k) \}$$

Si no hay ningún RB disponible, continúa con el siguiente par

Actualizar $W_j[n]$

}

$$F(j,k) = a_j W_j[n] \frac{\hat{r}_{j,k}[n]}{\bar{r}_j[n]} \quad \text{con} \quad a_j = -\frac{\log_{10}(\delta_j)}{T_j} \quad (\text{III.5})$$

En la ecuación III.5, W_j es el retardo acumulado del paquete situado en la cabecera de la cola del usuario j , δ_j es la probabilidad con la que se desea alcanzar los requerimientos de retardo T_j del usuario j , $\hat{r}_{j,k}[n]$ es el throughput instantáneo en el RB k y $\bar{r}_j[n]$ es el throughput instantáneo promedio que puede observar el UE en todo el ancho de banda del sistema. El algoritmo presenta menores grados de libertad que el algoritmo LDF.

En las simulaciones se decidió usar los valores $T_j=0.8*d_{max}$ y $\delta_j=0.05$.

3.4: Exponential rule (EXP)

Algoritmo similar al MLWDF pero sustituyendo la función (III.5) por (III.6) [7], donde N es el número de usuarios que se reparte los recursos:

$$F(j,k) = a_j W_j [n] \frac{\hat{r}_{j,k}[n]}{\bar{r}_j[n]} \exp\left(\frac{a_j W_j [n] - \overline{aW}}{1 + \sqrt{\overline{aW}}}\right) \quad \text{con} \quad \overline{aW} = \frac{1}{N} \sum_{j=1}^N a_j W_j [n] \quad (\text{III.6})$$

En este caso, el término $a_j W_j \frac{\hat{r}_{j,k}[n]}{\bar{r}_j[n]}$ es más o menos dominante a medida que el retardo del UE j se aproxima a la media de los retardos de todos los UEs que esperan recibir servicio. En otro caso, el factor exponencial proporciona mayor prioridad al usuario con mayor retardo.

4. ALGORITMOS DE SCHEDULING

Este es el apartado en el que se introducen los algoritmos de scheduling analizados, programados y simulados durante la ejecución de este proyecto. Éstos fueron probados, al igual que los de la sección anterior, sobre el simulador explicado en el anexo A.

Al igual que en la sección anterior, se asumen algunas denominaciones comunes a todos los algoritmos comentados, siendo $\hat{r}_{j,k}[n]$ el throughput, o tasa instantánea estimada para el usuario j -ésimo en el RB k -ésimo, N es el número de usuarios que esperan recibir servicio y el número de RBs disponibles es M .

4.1: Urgency and Efficiency-based Packet Scheduling: UEPS

El primer algoritmo con el que probamos las posibles mejoras en el scheduling de paquetes es el algoritmo UEPS [8] [9].

Este algoritmo tiene como objetivo maximizar el throughput de los terminales de usuario con tráfico sin restricciones temporales, al tiempo que se cumple con los requisitos de calidad QoS de los que requieran servicios en tiempo real. Se le denomina UEPS porque los dos factores principales a tener en cuenta en la asignación de recursos son la urgencia de los paquetes, representada por una función de utilidad dependiente del tiempo (time-utility), y la eficiencia del uso de los RBs, indicada por el estado del canal.

Para contabilizar la eficiencia en la utilización de los recursos radio se puede recurrir al algoritmo de scheduling del Proportional Fair, utilizando la relación entre el throughput instantáneo ($\hat{r}_j[n]$) y el promediado ($T_j[n]$) de cada UE j . Sin embargo, un aspecto importante es determinar si el valor $\hat{r}_j[n]$ se va a referir a la tasa instantánea que se obtiene utilizando todo el ancho de banda o si se calcula para cada RB por separado.

$$ratio = \frac{\hat{r}_j[n]}{T_j[n]} \quad (IV.1)$$

En el caso del factor dependiente del retardo, se establecen dos funciones diferentes, una para los servicios en tiempo real y otros para los servicios en tiempo no real. De

forma genérica las funciones pueden ser hard o soft dependiendo de si la respuesta al límite de la cota de retardo es el descarte del paquete o no (soft). La función de utilidad utilizada se ilustra en la Figura 4.1. Los servicios en tiempo real, que requieren no superar un plazo temporal, deberían regirse por una función de utilidad hard cuyo valor sea sólo distinto de 0 en un intervalo anterior al límite de descarte del paquete. En determinados servicios como mail u otros tráficos FTP (File Transfer Protocol) puede que no tengan ninguna restricción de retraso o que el límite sea muy elevado.

El criterio de ordenación temporal de las transmisiones es la derivada de la función de utilidad temporal, $|U'(t)|$. La motivación es detectar cambios en la tolerancia al retardo. En el caso particular de los servicios en tiempo real, para evitar la discontinuidad que aparece en el límite del retardo, se define una función de suavizado, de tal forma que finalmente los valores absolutos de la derivada de la función de utilidad son:

$$U(t) = \begin{cases} e^{-a(t-c)} / (1 + e^{-a(t-c)}) & \text{para servicios en tiempo real} \\ 1 - e^{at} / e^D & \text{para servicios en tiempo no real} \end{cases} \quad (\text{IV.2})$$

Donde a y c son parámetros que definen la pendiente y el punto de inflexión de la función de utilidad y D es la cota de retardo del servicio en tiempo no real.

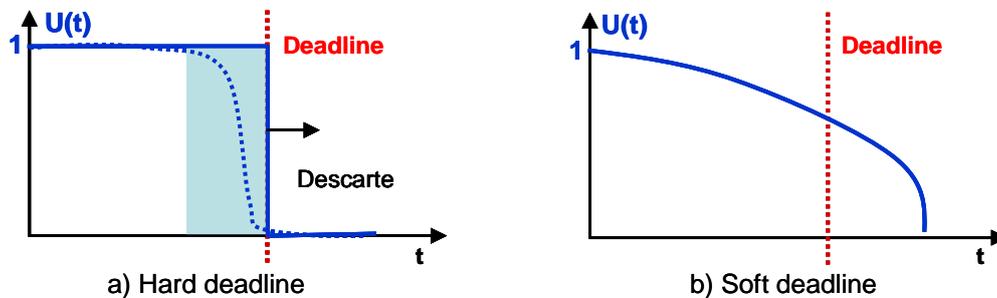


Figura 4.1: Función utilizada para el algoritmo UEPS (servicios en tiempo real)

Finalmente, el criterio global de scheduling consiste en seleccionar el UE j que recibirá servicio de acuerdo con la ecuación (IV.3).

$$\hat{j} = \arg \max_j \left\{ |U'(t)| \frac{\hat{r}_j[t]}{T_j[t]} \right\} \quad (\text{IV.3})$$

Scheduling y provisión de QoS en redes móviles 4G OFDMA

Para los parámetros a y c , se escogieron varios valores atendiendo a la forma de sus gráficas y al servicio con los que iban a ser probados, quedando finalmente los representados en la Figura 4.2.

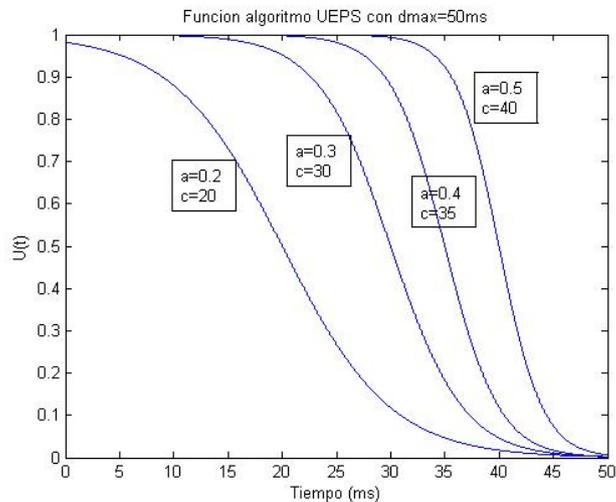


Figura 4.2: Función $|U(t)|$ para distintos valores de los parámetros a y c

Al ver las gráficas que se obtenían al dibujar la función $|U'(t)|$ (Figura 4.3), y tras ver que los resultados no eran nada buenos aplicando el algoritmo tal y como viene expresado en [8] (ver apartado resultados, Figura 5.2), se decidió probar otras posibilidades, con el convencimiento de que un algoritmo que diera un buen resultado debía tener en cuenta al mismo tiempo el retardo acumulado de los paquetes y la tasa que requieren en sus transmisiones.

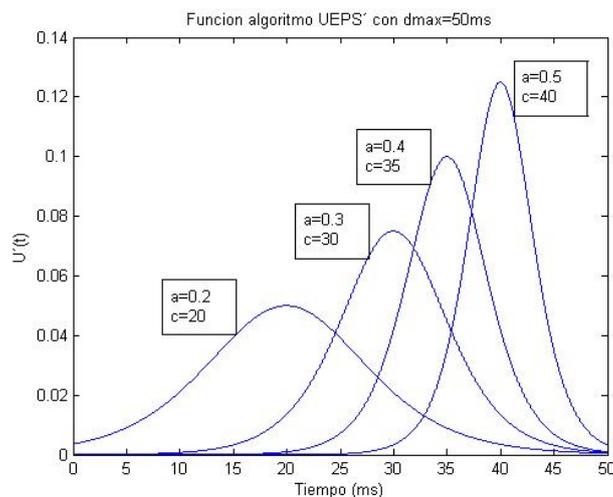


Figura 4.3: Función $|U'(t)|$ para distintos valores de los parámetros a y c

Así, se decidió hacer algunas modificaciones al algoritmo para probar otras posibilidades, siendo la que finalmente se empleó la siguiente:

- Hacer uso de la función $|1 - U(t)|$ en lugar de la derivada, para dar mayor prioridad a los paquetes con mayor retardo acumulado:

$$\hat{j} = \arg \max_j \left\{ |1 - U(t)| \frac{\hat{r}_j[t]}{T_j[t]} \right\} \quad (\text{IV.4})$$

Con este nuevo algoritmo sí se obtuvieron mejores resultados, por lo que fue el que se probó con más posibilidades de tráfico, bloqueo a usuarios con malas condiciones de canal, etc.

4.2: Variante de MLWDF y EXP

En [10] se propone un algoritmo de scheduling alineado con las propuestas MLWDF y EXP. En este caso, la función de prioridad se calcula de acuerdo con la expresión (IV.5):

$$P_{j,k}[n] = \begin{cases} \frac{\hat{r}_{j,k}[n]}{T_j[n]} \cdot \frac{1}{(d_{\max} - d_j[n])^\phi} \cdot \frac{1}{(D_{\max} - D_j[n])^\theta} \cdot W & \text{servicios en tiempo real} \\ \frac{\hat{r}_{j,k}[n]}{T_j[n]} & \text{servicios en tiempo no real} \end{cases} \quad (\text{IV.5})$$

Donde:

$\hat{r}_{j,k}[n]$ es el throughput instantáneo promedio del UE j en el RB k , en el instante de scheduling n .

$T_j[n]$ es la tasa media del UE j -ésimo en un tiempo t_f .

d_{\max} es el máximo retardo permitido.

D_{\max} es el máximo retardo de jitter permitido.

$d_j[n]$ es el retardo estimado del primer paquete en la cola del usuario j -ésimo.

$D_j[n]$ es el retardo estimado para llenar el buffer receptor al final del intervalo de scheduling.

W es un parámetro peso, para garantizar que el servicio en tiempo real es prioritario.

ϕ , θ son pesos de penalización que se ajustan en función de los requisitos de los servicios.

$$\hat{j} = \arg \max_j \{P_{j,k}[n]\} \quad (IV.6)$$

Para la programación y simulación de este algoritmo se decidió prescindir de los parámetros de jitter de la ecuación (IV.5) (D_{max} , $D_j[n]$ y θ), quedando una ecuación de nuevo dependiente del retardo acumulado y de la tasa (IV.7), que fue la que finalmente se utilizó.

$$P_{j,k}[n] = \begin{cases} \frac{\hat{r}_{j,k}[n]}{T_j[n]} \cdot \frac{1}{(d_{max} - d_j[n])^\phi} \cdot W & \text{servicios en tiempo real} \\ \frac{\hat{r}_{j,k}[n]}{T_j[n]} & \text{servicios en tiempo no real} \end{cases} \quad (IV.7)$$

4.3: Clasificación en grupos de urgencia

Otra idea que se planteó para desarrollar algún nuevo algoritmo fue la de clasificar los paquetes que requerían servicio en distintos grupos de urgencia, según cuánta tuvieran en función de diversos criterios. Se dispuso que un buen criterio sería clasificar los paquetes en 3 subgrupos: muy urgentes, urgentes y no urgentes, dándose siempre una mayor prioridad a los primeros y menor a los últimos.

Finalmente se tomó la decisión de realizar una doble organización: una inicial para decidir los tres subgrupos en que clasificamos cada paquete y una segunda para organizar los paquetes dentro de cada uno de los subgrupos. Se intenta ilustrar esta idea con la Figura 4.4.

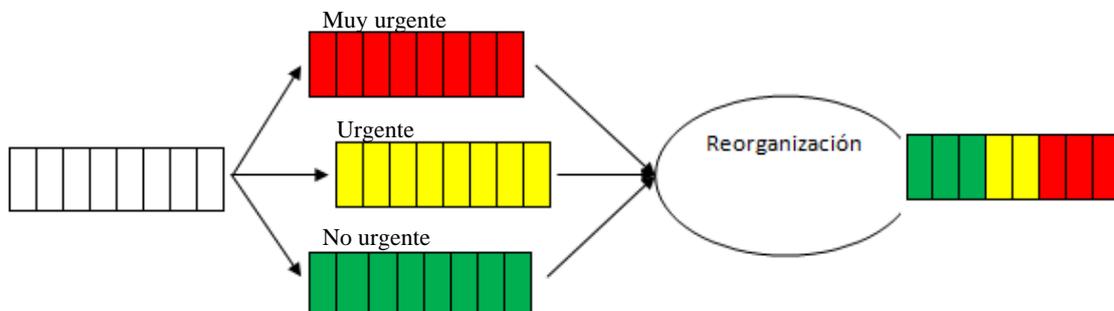


Figura 4.4: Clasificación en subgrupos de urgencia y reorganización

Tras pensar en varios algoritmos posibles para la decisión de los grupos y para la posterior reorganización y probar algunos de los mismos para ver los resultados obtenidos, se decidió quedarse con las dos posibilidades que se presentan en los siguientes subapartados.

4.3.1: División en subgrupos según tiempo de vida del paquete

En este caso, se decidió realizar la clasificación en subgrupos de urgencia teniendo en cuenta el retardo acumulado del paquete en la cabecera de la cola de cada UE, es decir, en función del mismo colocábamos a los paquetes en uno de los tres subgrupos (mayor urgencia cuanto mayor retardo acumulado tiene el paquete), para después ordenar los paquetes dentro de cada subgrupo en función de otro criterio cualquiera.

Los criterios seleccionados finalmente para ordenar los paquetes de cada subgrupo fueron los siguientes: ordenar los paquetes de los grupos muy urgente y no urgente en función de la tasa que requieran por medio del criterio *Proportional Fair* (III.3), y los del grupo urgente por el del *retardo* (III.2).

4.3.2: División en subgrupos según el estado de los buffers

Para esta segunda opción decidimos no tener en cuenta únicamente el paquete en la cabecera de la cola, sino el estado del buffer de cada UE al completo [11]. Se propone un criterio que toma en consideración la cantidad de datos que el usuario presenta en sus buffers, estimando la cantidad de TTIs que requeriría para dar salida a todos los TUs almacenados, junto con el tiempo de vida de dichos paquetes. Esto da una idea de si el usuario va a ser capaz idealmente de transmitir los paquetes que tiene almacenados en cola dentro del margen dado por su tiempo de vida. En concreto, la fórmula aplicada se muestra en la ecuación IV.8, donde n_TTIs_req es el número estimado de TTIs requerido para transmitir todos los datos existentes en los buffers, $(d_{max}-W_i)$ representa el tiempo de vida del paquete en la cabecera del buffer (HoL) del usuario (al asumir que los buffers tienen política FIFO, los paquetes sucesivos tendrán igual o mayor tiempo de vida) y α es un parámetro para tener en cuenta las retransmisiones.

$$F_i = \frac{n_TTIs_req}{(n_TTIs_req + (d_{max} - W_i))(1 - \alpha)} \quad (IV.8)$$

El valor de n_TTIs_req (ecuación IV.9) se estima asumiendo que en adelante el usuario va a recibir un RB por TTI con una determinada MCS. Así, si la tasa servida con esta MCS es r_x , tenemos:

$$n_TTIs_req = \frac{bits_en_buffers}{r_x} \quad (IV.9)$$

Tendrá una mayor prioridad el usuario que tenga una mayor cantidad de paquetes en su buffer: $\hat{j} = \arg \max_i \{F_i\}$

A la hora de parametrizar la fórmula del criterio de decisión se han tomado los siguientes valores: $\alpha=0.1$ (lo que supone dimensionar para un 10% de TUs retransmitidos). Aunque el valor de α se ha tomado constante en estas simulaciones, una posible mejora del criterio pasaría por tomar en consideración la tasa de BLER que el usuario ha sufrido realmente en las últimas transmisiones para obtener una estimación más precisa del margen de tiempo disponible para transmitir los paquetes en cola.

4.4: Combinación de Proportional Fair y Retardo

Finalmente, se decidió probar una combinación entre estos dos criterios anteriormente comentados, intentando conseguir un algoritmo que tuviera en cuenta tanto los valores del retardo acumulado como los de la tasa requerida a la hora de decidir la prioridad de un paquete.

$$\hat{j} = \arg \max_j \left\{ \frac{W_j[n]}{d_{\max}} \times \frac{\hat{r}_{j,k}[n]}{T_j[n]} \right\} \quad (\text{IV.10})$$

Como puede apreciarse en la ecuación (IV.10), se probó un algoritmo que usa una multiplicación de ambos criterios para escoger las prioridades de cada TU.

5.RESULTADOS OBTENIDOS

El objetivo de este apartado es evaluar las prestaciones de distintas propuestas de scheduling temporal en combinación con las estrategias de scheduling frecuencial y coordinación de interferencias propuestas y evaluadas a lo largo de esta memoria. Se consideran los algoritmos introducidos y comentados en los apartados 3 y 4: Proportional Fair (PF), Modified Largest Weighted Delay First (MLWDF), ratio entre el tiempo de espera en cola y la cota máxima del retardo (Retardo), Exponential Rule (EXP), Urgency and Efficiency-based Packet Scheduling (UEPSder) y la modificación del mismo (1-UEPS), variante de MLWDF y EXP (MLWDFvar), división en subgrupos según retardo acumulado del paquete (tem3reorg), división en subgrupos según el estado de los buffers (npaq3reorg) y la combinación del PF y el Retardo (PFret). Dichos criterios de prioridad se combinan con criterios dependientes del canal. Como paso previo, antes de detallar las propuestas evaluadas, se presenta la descripción del modelo de sistema considerado.

5.1: Descripción del modelo de sistema

Aquí se ofrece una descripción del modelo de sistema utilizado en las simulaciones de sistema que permitirán comparar las prestaciones ofrecidas por las distintas propuestas de scheduling.

5.1.1: Parámetros generales

Las simulaciones para la comparación de las distintas propuestas de scheduling se han llevado a cabo por medio de la plataforma de evaluación programada en C++, descrita en el anexo A. Los parámetros generales del sistema utilizados en la simulación se recogen en la Tabla 5.1.

El despliegue celular considerado consta de 19 celdas trisectorizadas que son simuladas con la técnica *wrap-around* para evitar efectos de borde en las celdas del extremo del despliegue. Para los desvanecimientos Rayleigh se han aplicado a los modelos extendidos propuestos en E-UTRA [12] que pueden verse en la Tabla 5.2. En la realidad, la respuesta frecuencial del canal varía con el tiempo como resultado principalmente del desplazamiento del terminal móvil.

Scheduling y provisión de QoS en redes móviles 4G OFDMA

Parámetros	Valores	
Despliegue celular	19 celdas trisectorizadas. Técnica wrap-around	
Distancia entre BSs	1732 m	
Modelo de propagación	$L = 128.1 + 37,6 \log_{10} d (km)$	
Desviación de los desvanecimientos lentos	8dB	
Distancia de coherencia de los desvanecimientos lentos	50m	
Correlación de los desvanecimientos	0,5	
Ruido térmico	-174dBm/Hz	
Factor de ruido del receptor	5dB	
Frecuencia central	2,5GHz	
Ancho de banda del sistema	5MHz	
Periodo de scheduling (TTI)	1ms	
Número de RBs	27	
Tamaño del RB	12 subportadoras OFDM durante 1 TTI	
Modelo de desvanecimientos Rayleigh	Velocidad del T_UE (km/h)	Modelo de desvanecimientos Rayleigh
	3	Pedestrian-A extendido
Modelo del espectro Doppler	Jakes	
Potencia máxima transmitida en la BS	43dBm	
Potencia máxima transmitida por RB	32dBm	
Ganancia de antena en BS	15dB	
Patrón de radiación horizontal de antena en el T_BS	70°(-3dB) con 20 dB de front-to-back ratio	
Ganancia de antena en el UE	0dB	
Ganancia de diversidad en transmisión	3dB	
Ganancia de diversidad en recepción	3dB	

Tabla 5.1: Parámetros del sistema

Pedestrian A extendido (3km/h)	
Retardo del camino (ns)	Potencia relativa (dB)
0	0,0
30	-1,0
70	-2,0
90	-3,0
110	-8,0
190	-17,2
410	-20,8

Tabla 5.2: Modelo de propagación multicamino

La distribución de los usuarios dentro de la celda es homogénea tanto entre sectores como zonas, es decir, existe el mismo número de usuarios en cada sector de la celda y a su vez, éstos se distribuyen homogéneamente entre las 3 zonas de que consta un

sector. La movilidad de los usuarios está restringida también a su zona, con lo que la carga va a permanecer siempre con dicha distribución homogénea. En estas condiciones se simula el algoritmo FFRopa descrito anteriormente para diferentes condiciones de carga de la celda.

5.1.2: Selección de MCSs

El ajuste de la MCS permite variar la tasa servida a cada usuario en los RBs asignados de acuerdo con las condiciones particulares de canal y con el estado de sus buffers. Asumiendo que la SINR promedio en el RB es conocida por medio del mecanismo de indicación de calidad del canal (CQI), la BS sólo podrá asignar MCSs cuya SINR umbral sea inferior a la SINR promedio observada en el RB. Se asume que la SINR umbral con la MCS m es aquella SINR para la que se obtiene una tasa de error en el bloque del 1%.

Como se apunta en la Tabla 5.1, el tamaño elegido para el RB es de 12 subportadoras OFDM durante una subtrama (TTI) de 1 ms. Esta configuración es equivalente a 12 subportadoras durante 10 símbolos netos de datos, es decir, 120 subportadoras de datos por TTI. En la Tabla 5.3 se recoge la SINR umbral y la tasa de datos neta por RB para cada una de las 6 MCSs consideradas en las simulaciones.

MCS	Configuración	SINR umbral (dB)	Tasa de datos (bits/RB)
0	QPSK 1/2	5,2	120
1	QPSK 3/4	9	180
2	16QAM 1/2	12,5	240
3	16QAM 3/4	16,4	360
4	64QAM 2/3	22,4	480
5	64QAM 3/4	24,5	540

Tabla 5.3: SINR umbral y tasa de datos para cada MCS

5.1.3: Modelo de generación de CQIs

Como se ha descrito anteriormente, a la hora de asignar RBs y elegir MCSs es necesario disponer de medidas de la calidad del canal. Un modelo genérico de CQIs se muestra en la Figura 5.1. En nuestras simulaciones, la medida de calidad del canal es la relación canal a interferencia y ruido (CINR), calculada como se muestra a continuación:

$$CINR_k = \frac{|h_k|^2}{I_k + N} \quad (V.1)$$

Scheduling y provisión de QoS en redes móviles 4G OFDMA

donde h_k es la respuesta impulsional del canal en la portadora k , I_k es la potencia interferente en la portadora k y N es la potencia promedio del ruido AWGN (Additive White Gaussian Noise). Multiplicando el valor de CINR por la potencia de transmisión se obtiene una estimación de la SINR que observará el UE en recepción, con lo que se puede realizar la asignación de RBs y MCSs. Un CQI consta por tanto de un valor de CINR observado por RB, promediando el CINR de las 12 subportadoras del mismo durante las W_{CQI} subtramas anteriores. El valor de esta ventana de promediado afecta tanto a la precisión de las medidas de calidad del enlace como a la variabilidad de las mismas. Valores de W_{CQI} altos suponen arrojan un CQI más estable, que son adecuados para casos de movilidad alta pues es prácticamente imposible seguir con precisión la evolución del canal. Por el contrario, a velocidades bajas, un valor de W_{CQI} bajo permite seguir con mayor precisión la evolución del canal puesto que éste varía muy lentamente.

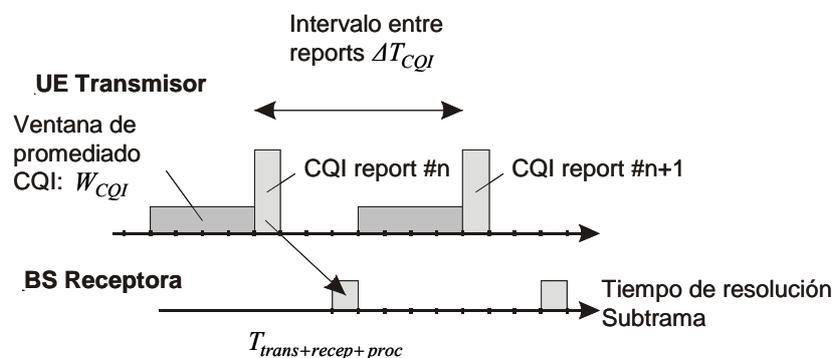


Figura 5.1: Modelo temporal de generación de CQIs

En un sistema real, las medidas de calidad del enlace obtenidas por el UE en una subtrama no pueden estar disponibles en la estación base para su uso en el scheduling en la siguiente subtrama. Existe un retardo t_{CQI} entre la generación de un CQI y su uso en el scheduler (figura 5.1), inherente a la propia comunicación, que aglutina el tiempo necesario para reservar recursos para la transmisión del CQI en el uplink así como la propia transmisión y el tiempo que requiere la BS para procesarlo y ponerlo a disposición del scheduler. De forma genérica, entre dos generaciones consecutivas de un CQI por parte de un usuario pasan un número determinado de subtramas (ΔT_{CQI}) cuyo valor se puede ajustar generalmente dependiendo de la velocidad del UE. Cuanto menor sea la velocidad de desplazamiento del móvil, menor periodicidad en la generación de CQIs se requiere, pues los cambios que experimenta el canal son más lentos.

En las simulaciones presentadas en este apartado se han considerado los valores presentados en la Tabla 5.4 para los parámetros relacionados con el modelo de generación de CQIs. Los valores de t_{CQI} y W_{CQI} son los usualmente considerados en la

bibliografía, mientras que hemos asumido que se genera un CQI cada subtrama con el fin de evaluar las prestaciones de los mecanismos de scheduling sin el error debido a no disponer de medidas actualizadas cada TTI, aunque el CQI usado en el scheduling de la trama t se corresponda con las medidas realizadas en la trama $t-t_{CQI}$.

Parámetros	Valores
W_{CQI}	4ms
Δ_{CQI}	1ms
t_{CQI}	2ms

Tabla 5.4: Parámetros del modelo de CQIs

5.1.4: Modelo de tráfico

En las simulaciones presentadas en este apartado se asume que cada usuario mantiene una única conexión con la BS, transmitiéndose a través de ella un tráfico a ráfagas tipo ON-OFF. El tiempo de permanencia en cada estado se calcula mediante una variable aleatoria con distribución exponencial de medias T_{ON} en el estado activo y T_{OFF} en el estado de silencio. El factor de actividad del tráfico (α) se define por tanto como T_{ON}/T_{OFF} .

Durante el estado activo, la fuente genera cada TTI un paquete de datos de tamaño L . Para gestionar la transmisión del flujo de datos a través del interfaz radio, los paquetes se dividen en bloques más pequeños, denominados *Transport Unit* (TU), que es la unidad mínima de transmisión que se ha considerado en las simulaciones. Dado que la unidad mínima de asignación de recursos es un RB (no pueden asignarse fracciones de RBs), el tamaño del TU es igual a la cantidad neta de datos que pueden transmitirse en un RB cuando se hace uso de la MCS más baja del sistema, esto es, 120 bits. Cuando el tamaño del paquete L no es múltiplo del tamaño del TU, el último TU se rellena con ceros de padding hasta completar los 120 bits. Consideraremos por tanto que la tasa promedio requerida por la conexión es de L bits/TTI si el tamaño del paquete es múltiplo del tamaño del TU ó $120 \cdot \lceil L/120 \rceil$ bits/TTI en caso contrario.

Cada paquete de datos, y por tanto cada uno de los TUs en los que se divide, tiene una cota máxima de retardo de d_{max} subtramas. Si un TU no ha sido recibido correctamente al cabo de d_{max} subtramas es descartado.

Las tablas siguientes recogen los distintos perfiles de tráfico utilizados en las simulaciones de este apartado. Cabe destacar que el factor de actividad se mantiene siempre en $1/4$, variándose el tiempo medio de permanencia en cada estado, lo que da lugar a dos tipos de perfiles: Ráfagas A y Ráfagas B. Se pretende de este modo estudiar el impacto que la presencia de ráfagas más largas de datos tiene en las prestaciones. Se consideran también distintas cotas máximas de retardo (Tabla 5.5).

Parámetros	Valores			
	Perfil Ráfagas A		Perfil Ráfagas B	
α	1/4			
T_{ON}	30ms		300ms	
T_{OFF}	90ms		900ms	
L	240 bits			
d_{max}	Ráfagas A0	50ms	Ráfagas B0	50ms
	Ráfagas A1	250ms	Ráfagas B1	250ms

Tabla 5.5: Perfiles de tráfico.

Para modelar con precisión la realidad, los usuarios deberían de generar llamadas de duración aleatoria, durante las cuales se transmiten paquetes con alguno de los perfiles antes descritos. Cuando una llamada finaliza, el usuario deja de transmitir paquetes pero se mantiene registrado en la celda, desplazándose por el escenario a la espera de iniciar una nueva llamada. Por tanto, a la hora de simular este escenario real, se necesitarían crear muchos más móviles para un valor de carga dado frente al caso de generación continua de tráfico (con alguno de los perfiles anteriores). Este exceso de usuarios supone un consumo adicional de memoria por el simulador y un crecimiento importante del tiempo de simulación. En nuestras simulaciones se opta por una solución que permite modelar de forma similar la existencia de llamadas pero sin aumentar el tiempo de simulación: para cada conexión se genera un número aleatorio de transiciones ON-OFF (N_{ON-OFF}) con distribución uniforme entre 4 y 6, reduciéndose en una unidad cada vez que el tráfico pasa del estado ON al OFF. Cuando N_{ON-OFF} llega a cero, se chequean los buffers de la conexión y, si no queda ningún paquete pendiente de transmisión, el UE se reubica en una nueva posición, generando unos nuevos valores para cada uno de los desvanecimientos. De este modo se emula la existencia de más usuarios con llamadas activas e inactivas sin la necesidad de consumir más memoria ni tiempo de simulación.

5.1.5: Modelo de retransmisión de TUs erróneos

Para mantener la integridad de los datos transmitidos, los tráficos considerados en las simulaciones siguientes requieren de un mecanismo de recuperación de errores (ARQ: Automatic Repeat Request) para la retransmisión de TUs erróneos. En nuestro caso, hemos decidido tratar los TUs a retransmitir como si de un TU de nuevos datos se tratara, pero con mayor prioridad dentro de la conexión. Cada conexión lógica de datos entre un móvil y su estación base tiene dos buffers distintos, ambos con política FIFO (First Input First Output): el *buffer de salida*, en el que se almacenan los nuevos TUs de datos pendientes de ser transmitidos; y el *buffer de ARQ*, que contiene los TUs a retransmitir porque fueron recibidos erróneamente. La prioridad de un TU de ARQ es siempre superior a la de los bloques de datos, con lo que si un móvil recibe una asignación de recursos, los TUs de ARQ serán transmitidos preferentemente.

De forma genérica, cuando se transmite un bloque de datos en el DL transcurre un cierto tiempo hasta que la estación base recibe la indicación del móvil sobre si el bloque ha sido recibido correcta o erróneamente. Por tanto, los TUs pendientes de ser retransmitidos deberán permanecer un cierto tiempo T_{HARQ} en la cola de ARQ antes de estar verdaderamente disponibles para su retransmisión. En nuestro caso, este valor se ha fijado en 6ms. Del mismo modo, para evitar que el mecanismo de ARQ bloquee la comunicación con sucesivas retransmisiones del mismo bloque de datos, se ha fijado un límite máximo de 4 retransmisiones, con lo que un TU se descarta cuando su cuarta retransmisión es recibida erróneamente.

5.2: Ráfagas A y B. Sin bloqueo.

En primera instancia se muestran los resultados obtenidos para los distintos algoritmos propuestos y evaluados, no aplicando ningún tipo de bloqueo a los usuarios con peores condiciones de canal y usando primero un perfil de ráfagas A (rA), con tiempos T_{ON} y T_{OFF} de 30ms y 90ms respectivamente y posteriormente ráfagas B (rB), con tiempos de 300ms y 900ms. Al considerar estos servicios como emulación de servicios en tiempo real, imponiendo restricciones de retardo "hard", se asume que los paquetes que superan un retardo por encima de la cota máxima especificada, son descartados por el sistema. Asumiendo este criterio, las prestaciones, en todos los casos, se evalúan en términos de la evolución de los porcentajes de TUs descartados en función del número de usuarios en cada celda. Es preciso indicar que se establece como criterio para determinar el número máximo de usuarios soportables en el sistema, una cota para la probabilidad de descarte del 1%. Asimismo, se incluyen resultados relativos al retardo medio de los TUs en función de la carga de la celda.

5.2.1: Ráfagas A. Tráfico uniforme.

El primer escenario corresponde al caso en el que todos los usuarios de la red generan el mismo tipo de tráfico, con ráfagas del perfil A0 ($d_{max}=50ms$).

En primer lugar, en la figura 5.2 se procede a la evaluación de las prestaciones de las estrategias con distintos parámetros de funcionamiento, así como la comparación de las prestaciones arrojadas por las mejores configuraciones de todas ellas. Para interpretar los resultados mostrados en la figura, se introduce como paso previo el significado de la leyenda de las gráficas:

- MLWDFvar: cada línea representa el descarte para distintos valores de la variable ϕ (ecuación IV.5).
- UEPSder y 1-UEPS: los valores de la leyenda se refieren a los valores de los parámetros c y a respectivamente (ecuación IV.2).

Scheduling y provisión de QoS en redes móviles 4G OFDMA

- Tem3reorg: los valores de la leyenda representan los valores de retardo acumulado a partir de los cuales los paquetes cambian de grupo de urgencia.
- Npaq3reorg: la leyenda da los valores de la ecuación IV.8 a partir de los que los paquetes cambian de grupo de urgencia.

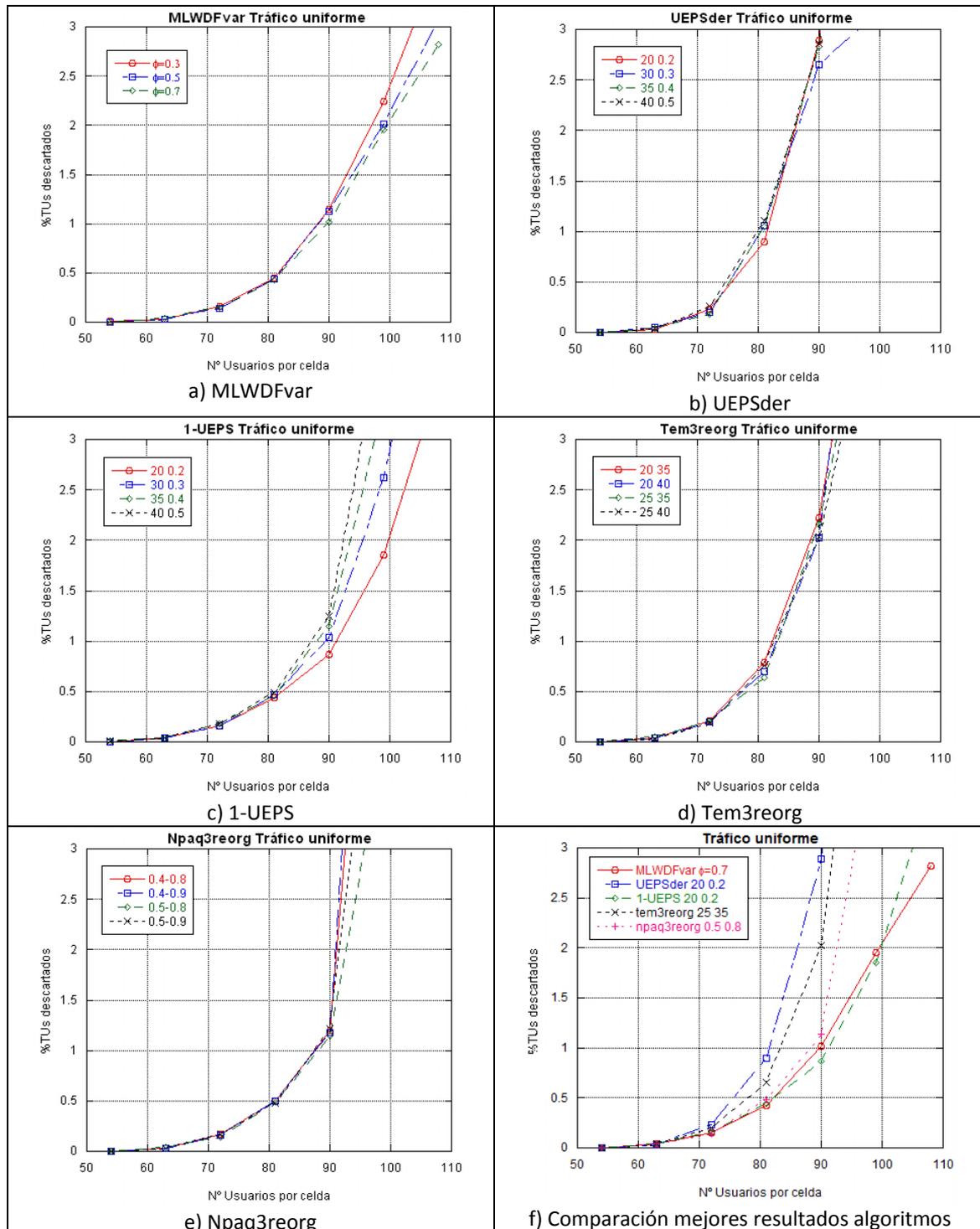


Figura 5.2: Resultados para los distintos métodos con tráfico uniforme (rA).

En cada uno de los esquemas, se probó con varios valores para los distintos parámetros en cada uno de los algoritmos, mostrándose aquí los que ofrecían un

mejor comportamiento. Se observa que para los algoritmos UEPSder (Fig. 5.2b), tem3reorg (Fig. 5.2d) y npaq3reorg (Fig. 5.2e), los valores elegidos para los parámetros afectan poco en el porcentaje de descarte, manteniéndose valores muy similares en todos ellos. En el caso del MLWDFvar (Fig. 5.2a) aunque no se aprecian diferencias significativas, particularmente en el rango de carga en el que se considera razonable fijar el funcionamiento del sistema (probabilidad de descarte por debajo del 1%), sí se puede observar como el mejor caso es aquel para el que $\phi=0.7$, valor para el que existe un mejor equilibrio entre la importancia dada por el algoritmo a sus factores dependientes de la tasa y del retardo acumulado.

El algoritmo que ve afectados sus resultados en mayor medida por los parámetros es el 1-UEPS. Para este tipo de tráfico, se observa que los valores que mejores resultados ofrecen son $c=20$, $\alpha=0.2$, ya que con estos valores se empieza antes a variar los valores de $|1-U(t)|$.

Al comparar los mejores resultados de los algoritmos, en la Figura 5.2f puede verse que los algoritmos con un mejor funcionamiento son el MLWDFvar y el 1-UEPS (Fig. 5.2c), mientras que el UEPSder y tem3reorg tienen un elevado número de descartes en cuanto crece el número de usuarios por celda. En lo que respecta a UEPSder, su peor comportamiento se debe a que, como puede verse en la Figura 4.3, la función $|U'(t)|$ proporciona mayor prioridad a los usuarios cuyo retardo acumulado está alrededor del parámetro c , causando que paquetes con mayor tiempo en el buffer queden con prioridades menores, incrementando el riesgo de descarte. Sin embargo, con el algoritmo 1-UEPS, que modifica la versión original UEPSder, siempre damos mayor prioridad a los paquetes con más retardo acumulado. Viendo que el algoritmo UEPSder es el que tiene un peor comportamiento en cuanto a número de descartes y dado que se cuenta con una variación (1-UEPS) que funciona significativamente mejor, se tomó la decisión de descartarlo para las sucesivas simulaciones.

El hecho de que los algoritmos que emplean la organización en dos fases diferenciadas (tem3reorg y npaq3reorg) tengan un mayor porcentaje de descartes que otros sucede porque al realizar la primera organización seleccionamos los paquetes con menor retardo acumulado en su paquete de cabecera (tem3reorg), o aquellos con el buffer más desocupado (npaq3reorg), y los ubicamos en el grupo de los no urgentes, que inevitablemente esperarán a que el resto de paquetes que se ubiquen en los otros grupos sean servidos. Estos paquetes, los no urgentes, serán difícilmente transmitidos o sólo una parte de ellos logrará recibir los recursos requeridos, por lo que irán acumulando retardo hasta pasar a formar parte de alguna de las otras dos categorías. En cuanto a la diferencia entre ambos, la causa de que el tem3reorg funcione peor que el npaq3reorg, reside en que el npaq3reorg crea los grupos en función del estado del buffer, estimando la cantidad de TTIs que requeriría para dar salida a todos los TUs

Scheduling y provisión de QoS en redes móviles 4G OFDMA

almacenados, mientras que el `tem3reorg` sólo tiene en cuenta el retardo acumulado del paquete en cabecera de la cola.

A continuación, en la figura 5.3 se comparan las prestaciones de los algoritmos referidos arriba con las prestaciones de los algoritmos PF, MLWDF, EXP, Retardo y la modificación PFret (Figura 5.3a). También se incluye una evaluación de estos algoritmos en términos del retardo acumulado (Figura 5.3b).

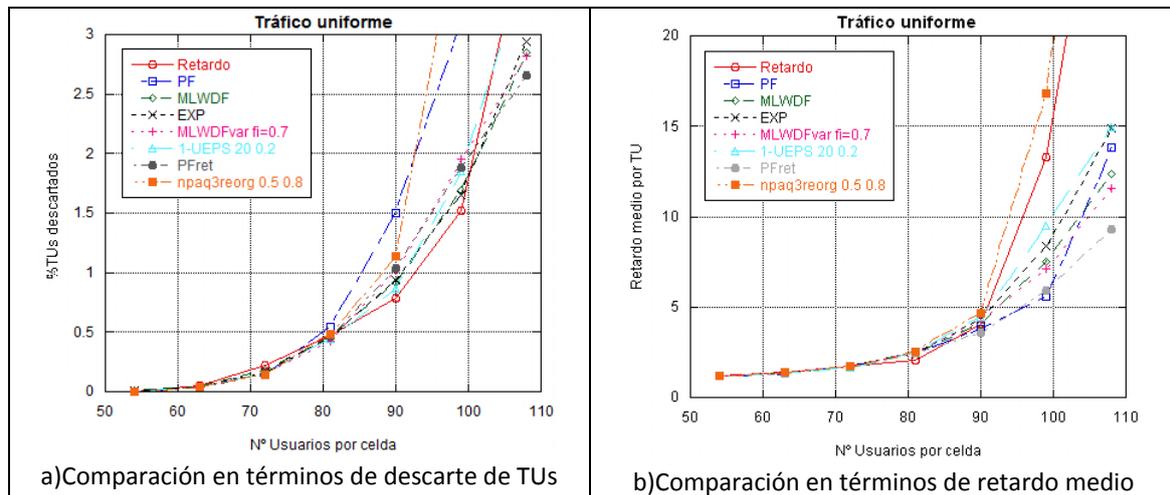


Figura 5.3: Comparación de los distintos métodos con tráfico uniforme (rA).

En lo que respecta a esta comparación en cuanto a porcentaje de descartes, los algoritmos que tienen peores resultados son el ya mencionado `npaq3reorg` y el PF. El caso del PF se debe a que la métrica aplicada en el scheduling sólo tiene en cuenta los valores de tasa promedio alcanzable y servida, en ningún caso atiende a factores dependientes del retardo. Esto implica que con frecuencia, usuarios que acumulan retardo debido a malas condiciones de canal, reciban pocos recursos en el sistema. En general, hay que tener en cuenta que para garantizar QoS en escenarios con servicios con restricciones de retardo, será necesario flexibilizar criterios como la maximización del throughput o reparto equitativo de recursos frecuenciales entre usuarios, para permitir que usuarios con peores condiciones accedan a los recursos del sistema dentro de las cotas de retardo especificadas. No obstante, como se verá más adelante, conviene evitar que estos usuarios monopolicen la capacidad del sistema. En contraposición al algoritmo PF, el algoritmo que mayor número de usuarios acepta manteniendo sus porcentajes de descarte por debajo del 1% es el del Retardo, ya que la prioridad de asignación se vincula de forma directa con el retardo acumulado en cola. No obstante, al no considerar el estado del canal, cuando la carga del sistema empieza a ser alta (el número de usuarios en el sistema implica una estimación alta de ocupación de los recursos frecuenciales del sistema con una MCS media), la probabilidad de descarte se dispara frente a otros esquemas. En este caso, el scheduling de paquetes con elevado retardo, que corresponden a usuarios con muy

malas condiciones de canal, obliga al uso para éstos de MCSs más bajas ocupando un número mayor de RBs. Esto resta recursos para el resto de paquetes/usuarios, que incluso observando buenas condiciones de canal, verán como aumenta su demora, llegando eventualmente a producirse el descarte de paquetes. En cualquier caso, el esquema Retardo sigue proporcionando mejores prestaciones que PF en todas las situaciones.

En un término intermedio se encuentran los algoritmos MLWDF, EXP, 1-UEPS, MLWDFvar, y el PFret, ya que dependen fuertemente del PF (todos ellos llevan ese factor en su ecuación), pero también tienen en cuenta de alguna forma el retardo, lo que hace que mejoren las características del PF sin llegar a las del Retardo, aunque aventajan a este en que con la acumulación de usuarios sus descartes no se disparan. No obstante, considerando como criterio la satisfacción de la cota del 1% en la probabilidad de descarte, la estrategia Retardo es la que admite un mayor número de usuarios en el sistema garantizando los parámetros de QoS. Las capacidades del sistema para estos algoritmos en cuanto a número de usuarios que podrían aceptarse satisfaciendo la cota del 1% en la probabilidad de descarte son similares entre sí, puesto que nos encontramos en un escenario de tráfico uniforme que hace que las diferencias entre los resultados de cada método sean mínimas.

La figura 5.3b muestra las prestaciones en términos de retardo promedio para los distintos algoritmos. En ella se observa que conforme aumenta la carga de la celda, los algoritmos que mayor retardo acumulan son el Retardo y el npaq3reorg. Éste último emplea en la segunda fase de su priorización el método del Retardo para ordenar a los paquetes urgentes. En vista de estos resultados, se concluye que al dar prioridad a los paquetes exclusivamente en función del retardo acumulado, los paquetes con menor retardo dan paso a las transmisiones de los paquetes más urgentes, viendo incrementado su tiempo de permanencia en cola. Esto es, en condiciones de carga media alta es posible garantizar que un mayor número de usuarios transmitan dentro de la cota de retardo, pero a costa de un incremento del retardo promedio de espera. No obstante, al aumentar mucho la carga, la demora en cola es muy difícil de compensar más adelante, y tanto el retardo como el descarte acaban creciendo rápidamente.

5.2.2: Ráfagas A. Escenario Multiservicio.

En este escenario se estudian las prestaciones ofrecidas por los algoritmos de scheduling analizados en el apartado anterior para el caso de considerar dos tipos de tráfico con distintos requerimientos de QoS. En concreto, el 50% de los usuarios de cada zona generan tráfico con el perfil de ráfagas A0 (con cota de retardo de 50ms) y el otro 50% con perfil A1 (con cota de retardo de 250ms), ambos con $T_{ON}=30ms$ y $T_{OFF}=90ms$. En este escenario podremos estudiar la capacidad de los algoritmos de scheduling para lidiar con distintos requisitos de QoS. Un algoritmo de scheduling que

Scheduling y provisión de QoS en redes móviles 4G OFDMA

se comporte adecuadamente deberá retrasar las transmisiones de los UEs con mayor cota de retardo en favor de los que tienen menor cota, con lo que debemos analizar el descarte y retardo sufrido por los usuarios de cada uno de los perfiles de tráfico de forma independiente, en lugar de analizar el global de la celda.

Dado que existen dos tipos de tráfico con distintos requerimientos de QoS, la prioridad de asignación de cada paquete deberá obtenerse de acuerdo con la cota de retardo de su flujo, es decir, el algoritmo de scheduling deberá parametrizarse adecuadamente de acuerdo con dicha cota. A modo de ejemplo, en el caso de 1-UEPS, la prioridad de los paquetes que pertenezcan a un flujo con cota 50ms se calculará con los valores a y c , que aparecen en la figura 4.2 mientras que para los paquetes pertenecientes a un flujo con cota 250ms, se utilizarán los siguientes valores: $a=0.1$ y $c=200$; $a=0.2$ y $c=220$.

Las prestaciones obtenidas por los distintos algoritmos de scheduling considerados se muestran en las figuras 5.4.1 y 5.4.2. Es preciso tener en cuenta que en este escenario, el número máximo de usuarios soportables, establecida una cota máxima para la probabilidad de descarte del 1%, vendrá dada por el servicio que experimente peores prestaciones.

La leyenda para los distintos algoritmos en el caso del escenario multiservicio (50% $d_{max}=50ms$ +50% $d_{max}=250ms$), que aparece en las figuras 5.4.1 y 5.4.2 está diseñada como se explica a continuación:

- MLWDFvar: cada línea representa el descarte para distintos valores del peso W , utilizado para priorizar el tráfico con menor cota máxima de retardo (ecuación IV.5).
- 1-UEPS: los valores de la leyenda se refieren a los valores de los parámetros c utilizados para los dos tipos de tráfico ($d_{max1}=50ms$ y $d_{max2}=250ms$ respectivamente) (ecuación IV.2).
- Tem3reorg: los valores de la leyenda representan los valores de retardo acumulado a partir de los cuales los paquetes cambian de grupo de urgencia para los dos tráficos.
- Npaq3reorg: la leyenda da los valores de la ecuación IV.8 a partir de los que los paquetes cambian de grupo de urgencia.

Scheduling y provisión de QoS en redes móviles 4G OFDMA

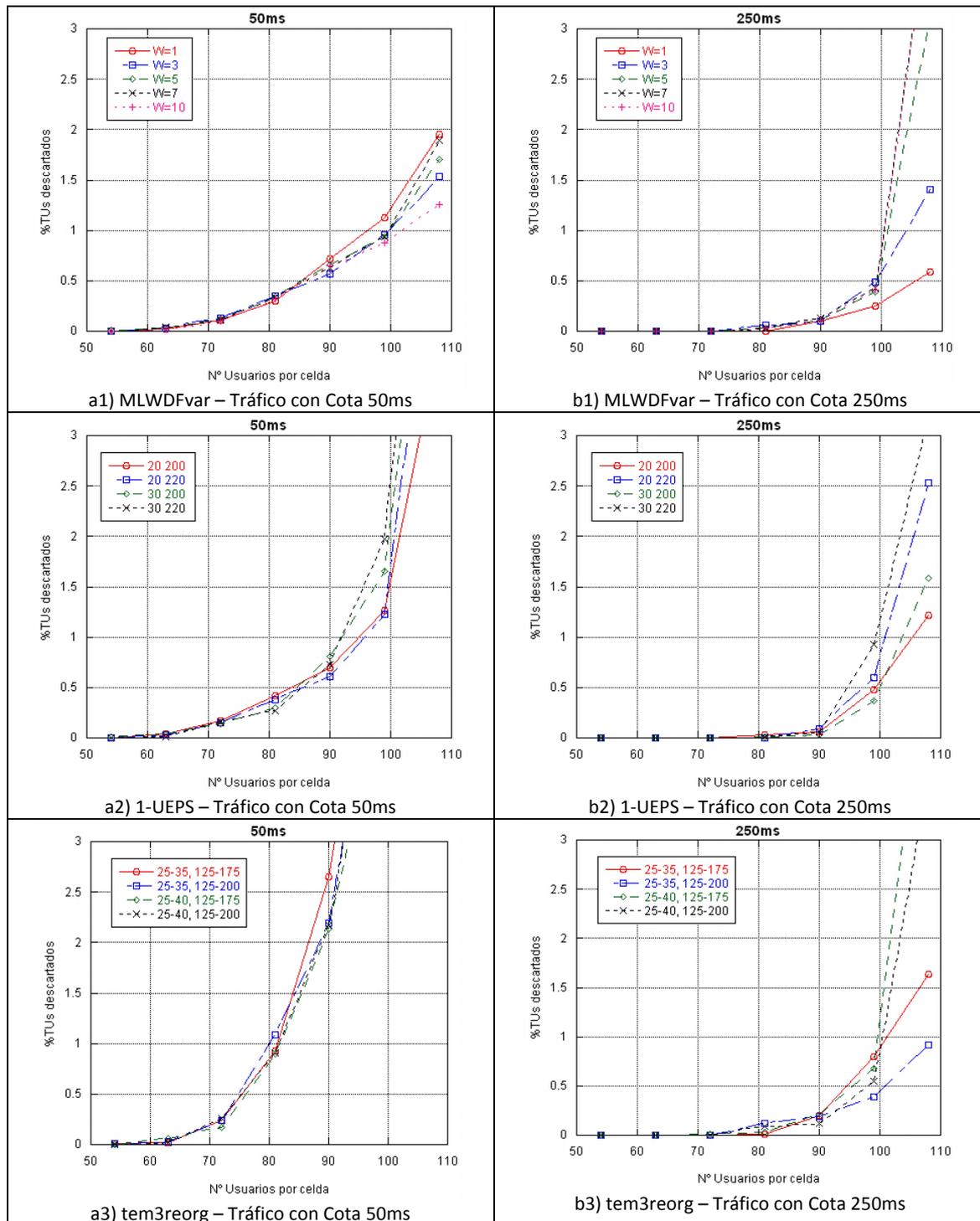


Figura 5.4.1: Resultados para los distintos métodos en un escenario multiservicio (50% $d_{max}=50ms$ +50% $d_{max}=250ms$) (rA). (I)

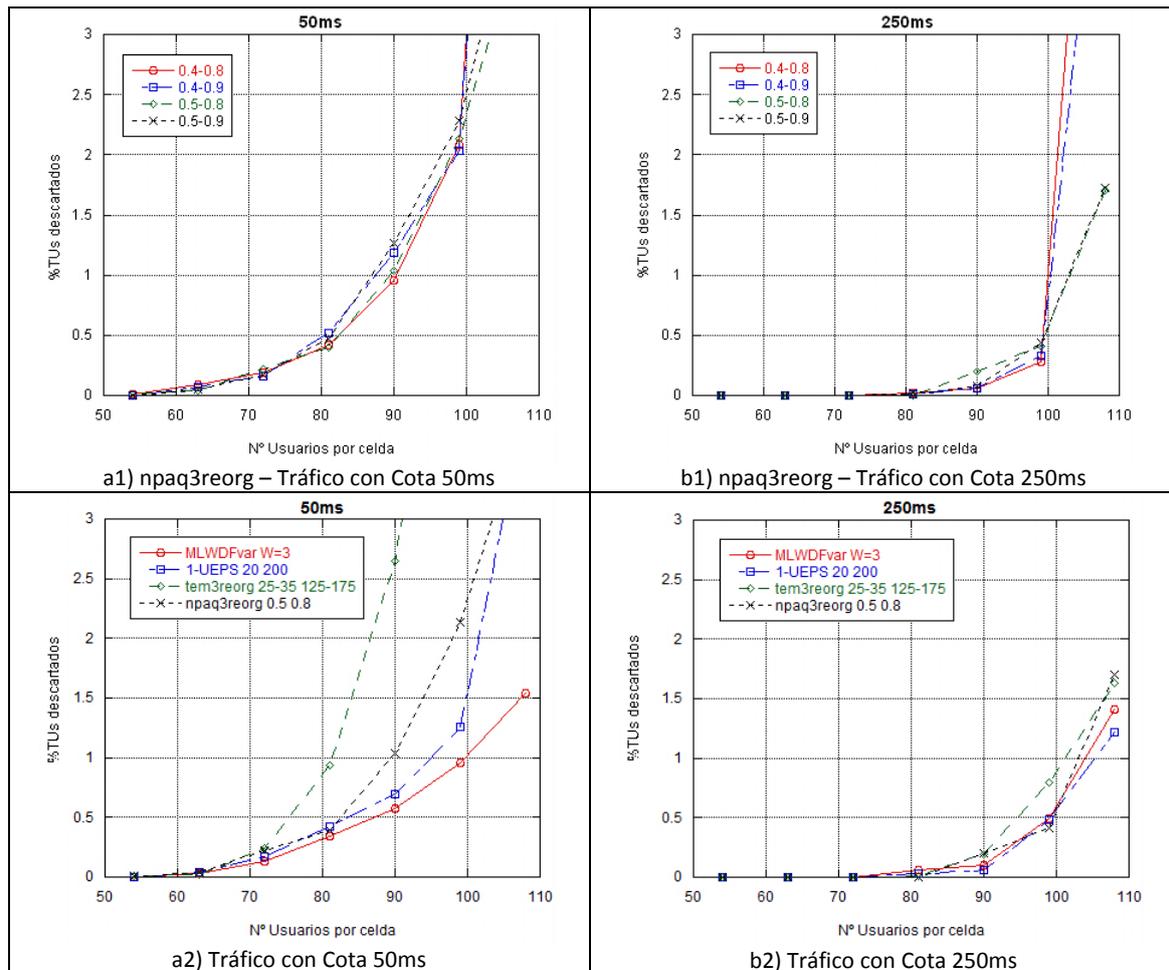


Figura 5.4.2: Resultados para los distintos métodos en un escenario multiservicio (50% $d_{max}=50ms$ +50% $d_{max}=250ms$) (rA). (II)

Como puede apreciarse en las figuras 5.4.1 y 5.4.2, de forma análoga a como se comentó en el apartado anterior de tráfico uniforme, los algoritmos que se ven afectados en menor medida por el valor de los parámetros escogidos son el npaq3reorg y el tem3reorg, teniendo un comportamiento muy similar en todos los casos estudiados. En el caso del 1-UEPS, se aprecia una diferencia entre los distintos valores de c para los distintos tipos de tráfico: el de $d_{max}=50ms$ funciona algo mejor con $c=20$, mientras que el de $d_{max}=250ms$ lo hace con $c=200$. De nuevo, el mejor comportamiento corresponde con el caso en que antes se empieza a variar los valores de $|1-U(t)|$. Con este tipo de tráfico, el algoritmo MLWDFvar depende en gran medida del parámetro W , un peso asignado para dar una mayor prioridad al tráfico con menor cota máxima de retardo, por lo que hay bastantes diferencias entre los resultados obtenidos para los distintos valores de este peso. Puede verse en las figuras 5.4.1a₁ y 5.4.1b₁ que el valor de W que acepta un mayor número de usuarios por debajo del 1% de descartes es $W=10$, pero su porcentaje se dispara en el caso del tráfico de $d_{max}=250ms$ cuando se incrementa el número de usuarios. Es por ello que escogimos el valor $W=3$, que evita este último inconveniente y que en el caso del tráfico con $d_{max}=50ms$ tiene unas prestaciones cercanas a las del $W=10$.

Scheduling y provisión de QoS en redes móviles 4G OFDMA

Se puede comprobar en las figuras que los métodos $tem3reorg$ (5.4.1a₃ y 5.4.1b₃) y $nraq3reorg$ (5.4.2a₁ y 5.4.2b₁), de igual forma que sucedía con el tráfico uniforme, tienen un peor comportamiento que el resto de algoritmos tanto para el tráfico de cota 50ms como para el de cota 250ms, por lo que no los incluiremos en la comparación posterior de la figura 5.5 con los algoritmos Retardo, PF, etc. Los que sí incluiremos en la esa figura son el 1-UEPS y el MLWDFvar, siendo este último el que ha experimentado una mejora considerable en su comportamiento respecto al caso del tráfico uniforme, gracias a la acción del peso W incrementando la prioridad del tráfico con una cota máxima de retardo menor.

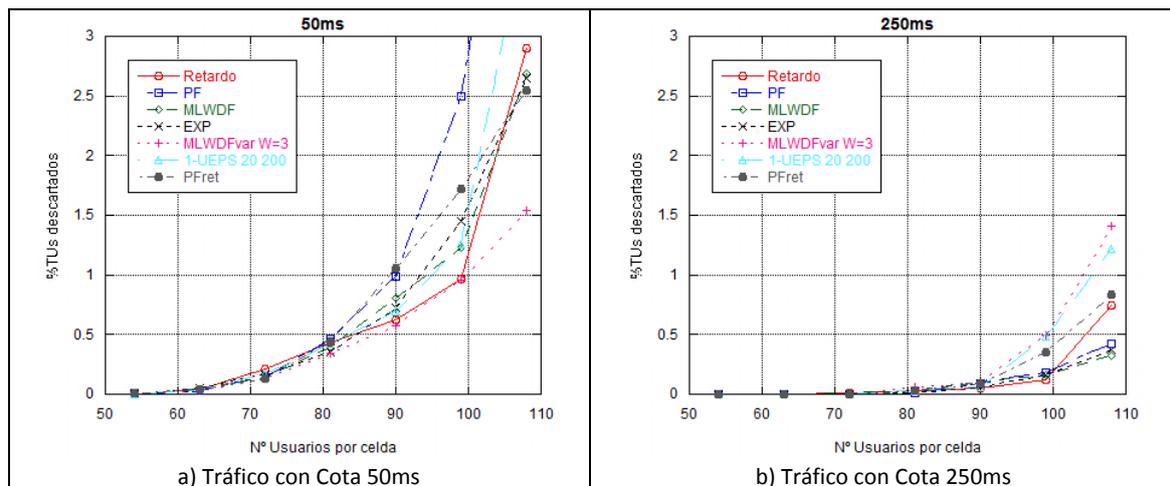


Figura 5.5: Comparación de los distintos métodos en un escenario multiservicio (50% $d_{max}=50ms$ +50% $d_{max}=250ms$) (rA).

En la figura 5.5 se puede ver como los algoritmos en general responden de una manera similar a como lo hacían en el apartado anterior, para el caso de tráfico uniforme. De nuevo PF es el que tiene un comportamiento algo peor por no tener en cuenta el retardo. El algoritmo Retardo funciona muy bien y da cabida a gran número de usuarios respetando la QoS, manteniéndose por debajo del 1% de descartes, pero de nuevo se dispara el porcentaje en un punto concreto al incrementarse el número de usuarios. De forma análoga, PFret, MLWDFvar y EXP funcionan de forma similar, llegando antes que Retardo a la cota del 1% de error pero manteniendo una progresión más lineal (no se dispara el porcentaje de descartes), aunque en este caso, al estar utilizando dos tipos de tráfico, sí que hay mayores diferencias entre ellos, haciendo que el PFret llegue antes al límite de descarte.

Por el contrario, hay dos algoritmos que ven modificado su comportamiento. El 1-UEPS se comporta de forma similar al MLWDF y el EXP tal y como sucedía con el tráfico uniforme pero sólo hasta 99 usuarios, ya que después su porcentaje de descartes se dispara subiendo por encima del 3%. Esto se debe a que este algoritmo es el que menos se diferencia del PF, es el menos influido por el retardo acumulado (depende del mismo por medio de una función exponencial que multiplica la ecuación dependiente de la tasa), mientras que el resto (MLWDF, EXP, etc.) dependen del

retardo más directamente (se multiplican por él o por una exponencial del mismo). Por ello el comportamiento del 1-UEPS tiende a igualarse más que el resto al del PF.

Mención especial merece la mejora conseguida con el algoritmo MLWDFvar. Consigue mejorar los resultados del Retardo en cuanto a que, aunque va muy parejo por debajo del 1% de descarte, su porcentaje no se dispara conforme aumenta el número de usuarios. Esta mejora es sin duda consecuencia de usar un peso (W) que prioriza más al tráfico con una menor cota máxima de retardo, consiguiendo retrasar las transmisiones de los UEs con mayor cota en favor de los primeros y consiguiendo un mayor número de usuarios en la celda respetando los parámetros de QoS fijados. Puede observarse en la figura 5.5b que esto influye de manera negativa en el porcentaje de descartes del tráfico de mayor cota (MLWDFvar es el que mayor porcentaje de descartes tiene), por lo que habrá que encontrar siempre el peso W que maximice la capacidad de la celda teniendo en cuenta los dos tipos de servicio.

5.2.3: Ráfagas B. Tráfico uniforme.

Este escenario es similar al del apartado 5.2.1 pero ahora todos los usuarios utilizan el perfil de tráfico de ráfagas B, en el que los tiempos medios de permanencia en los estados de actividad y silencio son respectivamente 300 y 900ms. Dado que las ráfagas en las que se generan datos son más largas, es esperable un aumento del descarte al no tener periodos de silencio tan frecuentemente en los que poder vaciar más rápidamente los buffers, como sucedía con las ráfagas A, de 30 y 90ms. Se genera tráfico uniforme; los paquetes de tráfico generados tienen una cota máxima de retardo de 50ms. La figura 5.6 muestra los resultados obtenidos para los algoritmos MLWDF var y 1-UEPS, en los que las diferencias son más significativas, dependiendo de la configuración de parámetros seleccionada. La leyenda en este caso es la misma que la utilizada en el apartado 5.2.1.

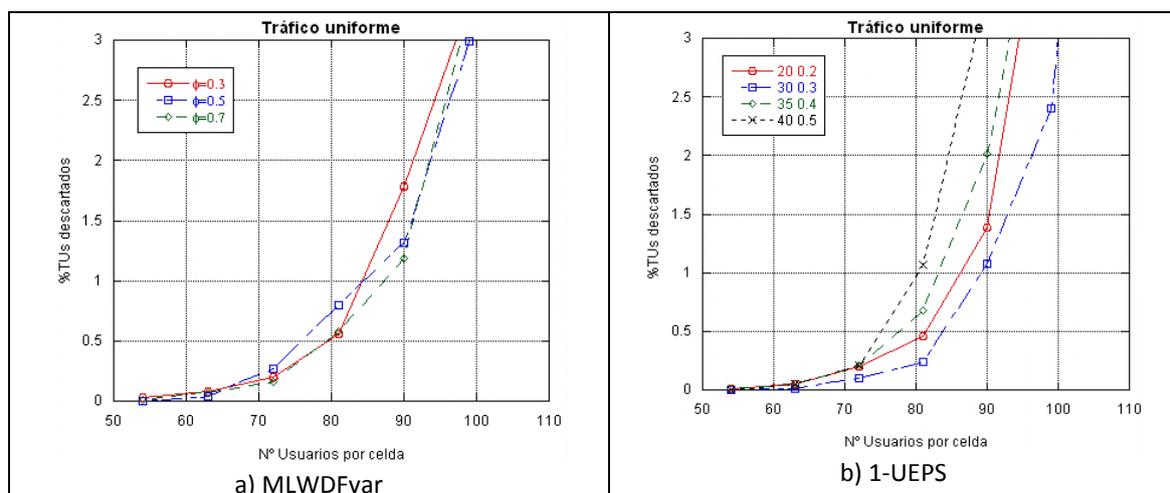


Figura 5.6: Resultados para los distintos métodos con tráfico uniforme (rB).

Scheduling y provisión de QoS en redes móviles 4G OFDMA

Como puede observarse, los valores escogidos para los parámetros afectan en mayor medida a los resultados obtenidos que en el caso de las ráfagas A. El algoritmo MLWDFvar sigue teniendo una mejor respuesta con $\phi=0.7$, en este caso aceptando hasta 3 o 4 usuarios más manteniendo los requerimientos de QoS que con los otros valores de ϕ . El 1-UEPS también aprecia mayores diferencias entre las diferentes configuraciones, con respecto a las apreciadas con las ráfagas A, sobre todo en los casos con un número no muy alto de usuarios. Para cargas bajas, en el caso de las ráfagas A los resultados eran muy similares para todos los parámetros. Ahora se aprecia un mejor funcionamiento con los valores de $c=30$ y $\alpha=0.3$. Esto es debido a que en este escenario es conveniente retrasar algo más el momento en que se comienza a incrementar el valor de la función $|1-U(t)|$, otorgando así más relevancia al factor dependiente del retardo en el algoritmo 1-UEPS.

En la figura 5.7 tenemos la comparación de las prestaciones de los algoritmos reflejados en la figura 5.6, con los parámetros que mejores resultados ofrecen, con las prestaciones de los algoritmos previos a este proyecto y la modificación PFret (figura 5.7a). Se incluye de nuevo una evaluación de estos algoritmos en términos del retardo acumulado para este tipo de ráfagas (figura 5.7b).

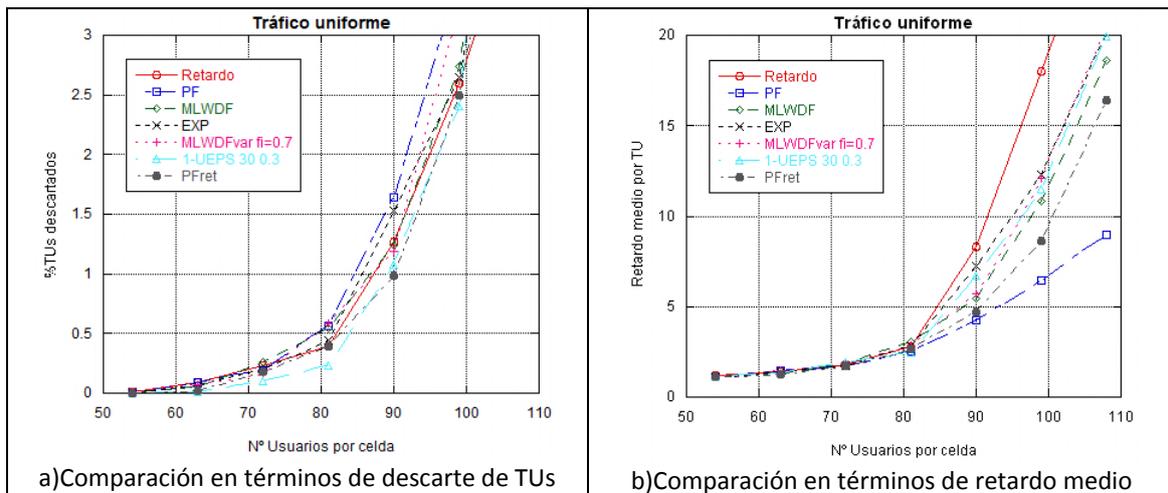


Figura 5.7: Comparación de los distintos métodos con tráfico uniforme (rB).

Puede verse en la figura 5.7a como las diferencias entre los algoritmos en este caso son poco significativas en lo que a descarte se refiere. Destaca que en este caso parecen funcionar algo mejor los algoritmos 1-UEPS y PFret, ya que Retardo empeora sus prestaciones en mayor medida que el resto de algoritmos ya que al haber ráfagas más largas, se acumulan retardos mayores. Al igual que en el caso de las ráfagas de 30 y 90ms, el PF es el algoritmo que peores resultados da, aunque en este caso con menores diferencias. De nuevo los algoritmos MLWDF, EXP y MLWDFvar tienen unas prestaciones que están entre las del PF y el Retardo.

Comparándolo con el primer escenario, el de ráfagas A, en este puede observarse un empeoramiento del descarte en todos los casos. Esto se debe a que al producirse unas ráfagas de datos más largas, los buffers no pueden vaciarse tan rápidamente, los paquetes acumulan retrasos mayores por norma general y, por tanto, tienen mayor probabilidad de llegar a ser descartados.

En la figura 5.7b se muestran las prestaciones en términos de retardo promedio para los distintos algoritmos. Con este tipo de ráfagas más largas se aprecia una división más clara entre los tres tipos de algoritmos utilizados y los retardos que se acumulan: por un lado está el Retardo, que al hacer depender sus priorizaciones exclusivamente del retardo acumulado, éste crece conforme crece la carga del sistema. De nuevo es el algoritmo que más retardo acumula. Por otro está el PF, que al depender exclusivamente de la tasa consigue no aumentar en demasía el retardo medio. Los paquetes con este algoritmo se priorizan en función de las condiciones del canal, la relación señal a ruido, etc., por lo que el aumento de la carga y del porcentaje de descartes no tienen porqué traer consigo un aumento del retardo. Por último tenemos el resto de algoritmos (MLWDF, 1-UEPS, etc.), que dependen al mismo tiempo, de un modo u otro, de la tasa y del retardo, y que por lo tanto ofrecen unos resultados intermedios a los de los algoritmos que funcionan exclusivamente con uno de los criterios. Alcanzan un mejor compromiso porcentaje de descartes/ retardo acumulado, pero ello no implica que ofrezcan mejores prestaciones, puesto que ha de garantizarse una QoS en todos los casos como condición *sine qua non*.

5.2.4: Ráfagas B. Escenario Multiservicio.

En este escenario, de forma análoga a lo realizado en el apartado 5.2.2, se estudian las prestaciones de los algoritmos analizados en el apartado anterior para dos tipos de tráfico distintos. De nuevo, el 50% de los usuarios de cada zona generan tráfico con cota de retardo de 50ms y el otro 50% con cota de retardo de 250ms, pero en este caso con $T_{ON}=300ms$ y $T_{OFF}=900ms$ (ráfagas B).

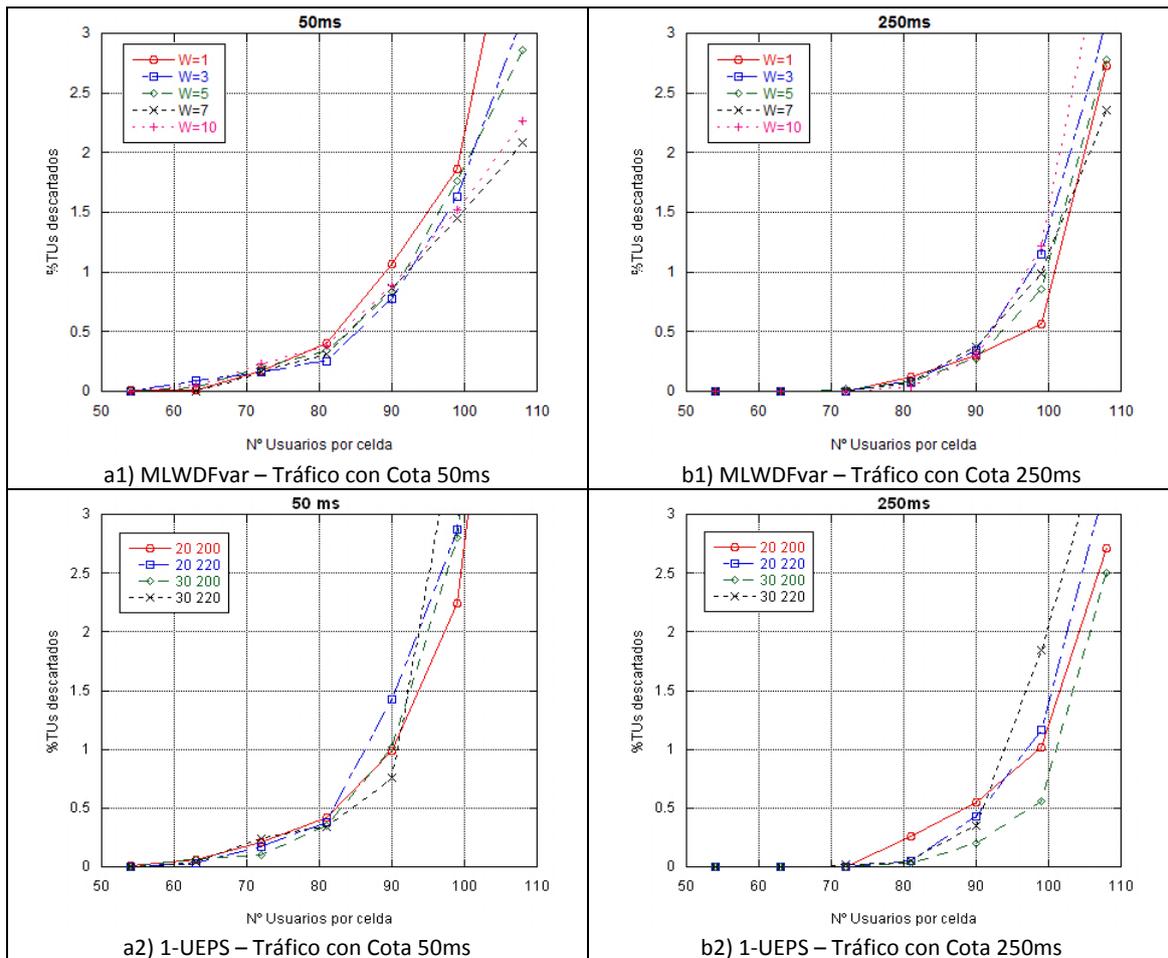


Figura 5.8: Resultados para los distintos métodos en un escenario multiservicio ($50\% d_{max}=50ms + 50\% d_{max}=250ms$) (rB).

En la figura 5.8 se muestran los resultados para MLWDFvar y 1-UEPS. Ahora la selección de la mejor opción entre los distintos valores de los parámetros para cada uno de los algoritmos resulta más difícil de determinar que en el caso del escenario de tráfico homogéneo. En el caso del MLWDFvar se tomó la decisión de usar las gráficas de $W=7$, ya que esta opción admite un número de usuarios cumpliendo el criterio de QoS requerida (por debajo del 1% de descartes) prácticamente igual que el de otros valores de W y además es el que, una vez superado ese margen, menos incrementa el descarte conforme crece el número de usuarios. Para el 1-UEPS se decidió como mejor caso el que tiene los valores del parámetro c de 30 y 200 para los tráficos con d_{max} 50ms y 250ms respectivamente. Esta opción proporciona una buena cota de descarte para el servicio con $d_{max}=50ms$ (su comportamiento está a la altura del obtenido para el resto de parámetros) y asegura un mejor comportamiento para el servicio con $d_{max}=250ms$.

Pueden apreciarse diferencias en los resultados de la figura 5.8 respecto al caso de tráfico mezclado con ráfagas A, apreciándose en este caso un mayor porcentaje de descartes, consecuencia de la mayor duración de las ráfagas, que provoca mayores acumulaciones de paquetes en los buffers. Resaltar sobre todo como el algoritmo

Scheduling y provisión de QoS en redes móviles 4G OFDMA

MLWDFvar experimenta crecimientos más rápidos en los casos de un número elevado de usuarios en la celda, circunstancia que no sucedía al usar ráfagas más cortas.

La figura 5.9 muestra las configuraciones elegidas de los algoritmos 1-UEPS y MLWDFvar con los algoritmos Retardo, MLWDF, EXP, etc.

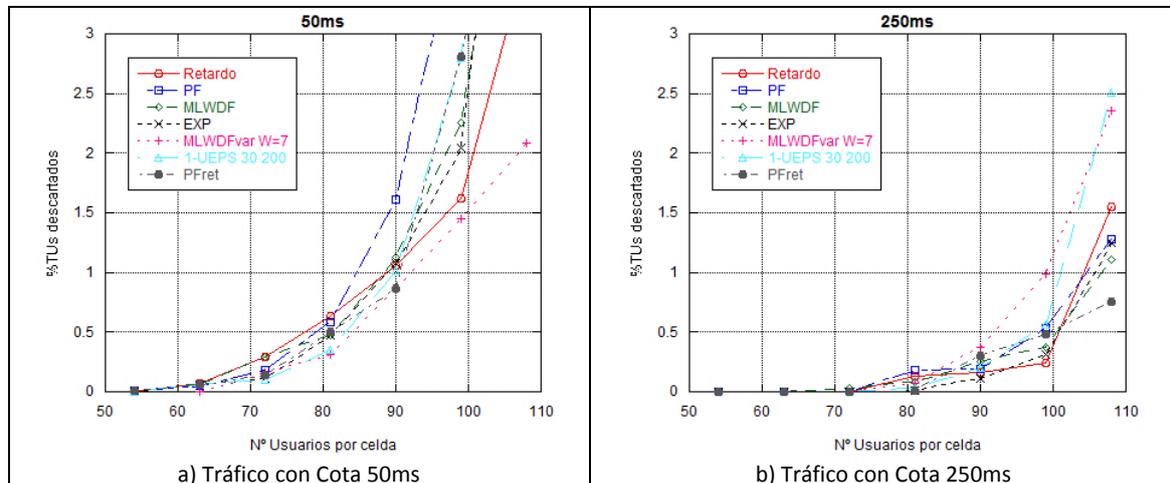


Figura 5.9: Comparación de los distintos métodos en un escenario multiservicio (50% $d_{max}=50ms$ +50% $d_{max}=250ms$) (rB).

En la figura 5.9 se aprecia que las diferencias relativas entre los algoritmos, son similares a las obtenidas en el apartado 5.2.2 (ráfagas de 30 y 90ms). Al igual que en aquel caso, el algoritmo PF es el que peores resultados da. Retardo sigue ofreciendo buenos resultados, pero al haber ráfagas de mayor duración se producen mayores esperas en los buffers y sus prestaciones en este caso se igualan a las ofrecidas por los algoritmos que combinan tasa y retardo en sus criterios de priorización: PFret, MLWDF, EXP, etc. Al igual que en aquel apartado, el mejor resultado es el obtenido por el algoritmo MLWDFvar. En este caso MLWDFvar no sólo consigue aguantar una progresión más lineal conforme crece el número de usuarios sino que además mejora el número de usuarios aceptables respetando los criterios de QoS. El ligero empeoramiento de las prestaciones del algoritmo Retardo por la mayor cantidad de tráfico y la utilización del peso W para dar mayor prioridad al tráfico con menor cota máxima de retardo, hace que para este escenario el algoritmo MLWDFvar nos ofrezca los mejores resultados. Observando también la figura 5.9b se aprecia como de nuevo el peso W afecta de manera negativa en el porcentaje de descartes del tráfico con $d_{max}=250ms$, siendo el MLWDFvar junto con el 1-UEPS los algoritmos que tienen un peor comportamiento con este tráfico. Pero esto no es preocupante puesto que es el tráfico de menor cota el que es más restrictivo ya que se debe cumplir el criterio de permanecer por debajo del 1% de descartes.

5.3: Ráfagas A y B. Bloqueo inicial.

En este apartado se va a analizar si resulta beneficioso para el global de la celda la aplicación de un bloqueo por condiciones de canal desde el inicio del proceso de asignación de recursos, de ahí el nombre de bloqueo inicial. El objetivo es analizar el impacto del procedimiento en función de la estrategia de scheduling considerada.

La aplicación de bloqueo a los usuarios con peores condiciones de canal evita que éstos monopolicen gran parte de los recursos. Así, se reservan más RBs para los usuarios con mejores condiciones de canal que pueden transmitir de manera más eficiente en ellos y, por tanto, el retardo promedio de la celda disminuye. De forma indirecta, este bloqueo está concentrando el retardo en aquellos usuarios con peores condiciones de canal.

En esencia, lo que se propone es asignar recursos primero a los usuarios con buenas condiciones de canal, que requerirán menos recursos para transmitir sus datos, y posteriormente repartir los RBs sobrantes entre los UEs con malas condiciones de canal hasta que no queden RBs libres o datos en los buffers. Sin embargo, la aplicación directa del bloqueo inicial supondría poner en riesgo el cumplimiento de los requisitos de QoS de los usuarios con malas condiciones de canal, ya que siempre verían reducida su accesibilidad a los recursos radio. Por ello, se propone aplicar este bloqueo inicial sólo en los casos en los que el usuario cumpla un determinado criterio que tenga en cuenta la satisfacción de sus requerimientos de QoS. En concreto, se propone utilizar dos criterios de retardo.

Con el primero el bloqueo inicial se aplicará exclusivamente a aquellos usuarios con malas condiciones de canal, es decir, los que requieran el uso de la MCS más baja del sistema (MCS0), y cuyo retardo (W_i) no sea superior a un determinado porcentaje δ de su cota máxima d_{max} , es decir:

$$\text{Bloqueo Inicial si } \frac{W_i}{d_{max}} \leq \delta \quad (\text{V.2})$$

El segundo se basa en el criterio ya utilizado en el apartado 4.3.2 para la reorganización en subgrupos. Éste toma en consideración la cantidad de datos que el usuario presenta en sus buffers, no solamente los que están en la cabecera de la cola. Empleamos de nuevo las ecuaciones IV.8 y IV.9, pero empleando, igual que en la ecuación V.2 un valor límite δ a partir del cual aplicaremos el bloqueo:

$$\text{Bloqueo Inicial si } \frac{n_TTIs_req}{(n_TTIs_req + (d_{max} - W_i)) \cdot (1 - \alpha)} \leq \delta \quad (\text{V.3})$$

Al igual que en el apartado 4.2.3, se emplea el valor $\alpha=0.1$.

Con este criterio, un usuario con malas condiciones de canal sufre el bloqueo inicial cuando todavía tiene margen para transmitir los paquetes almacenados en los buffers dentro de su tiempo de vida. Para un mismo número de datos en cola, el usuario cuyos paquetes tengan un mayor tiempo de vida será bloqueado con mayor probabilidad. Del mismo modo, a igualdad de condiciones de tiempo de vida, aquel usuario con un menor número de paquetes en cola es el que con mayor probabilidad será bloqueado.

Se decidió usar para ambos criterios los valores de $\delta = 0.8, 0.5$ y 0.2 , para evaluar las prestaciones obtenidas con el bloqueo inicial con criterio de retardo para las distintas estrategias de scheduling consideradas.

En los próximos subapartados se muestran los resultados más destacables de entre los obtenidos para los distintos algoritmos al aplicar bloqueo, comparándolos con los que se obtuvieron en el caso sin bloqueo. El resto de los resultados se incluyen en el Anexo B.

5.3.1: Bloqueo por retardo. Ráfagas A.

Empezaremos analizando el caso de emplear el bloqueo que únicamente tiene en cuenta el retardo acumulado por el paquete en cabecera usando ráfagas con tiempos de ON/OFF de 30 y 90 milisegundos respectivamente.

El primero de los casos para los que aplicar algún bloqueo tiene un efecto positivo es el algoritmo PF en el caso de tener tráfico uniforme de cota máxima de retardo 50ms (figura 5.10a). Al aplicar este bloqueo en función del retardo acumulado hacemos depender del retardo de alguna forma a la priorización, mejorando de esta forma los resultados del algoritmo acercándolo a los resultados de los algoritmos que empleaban ambos criterios (MLWDFvar, 1-UEPS,...) como puede verse en la figura 5.11b. Conviene destacar que en este caso es el hecho de aplicar el bloqueo en sí y crear cierta dependencia del retardo lo que causa la mejora, por lo que las diferencias entre los distintos tipos de bloqueo son mínimas, si bien el de $\delta = 0.2$ tiene una ligera ventaja sobre los otros. El algoritmo Retardo también experimenta cierta mejora en sus prestaciones, sobre todo en casos de carga elevada en el sistema, como puede verse en la figura 5.10b. En el caso de usar $\delta = 0.8$ se reduce el número de usuarios asumible por debajo del 1%, por lo que convendría no usar este valor de bloqueo con este método.

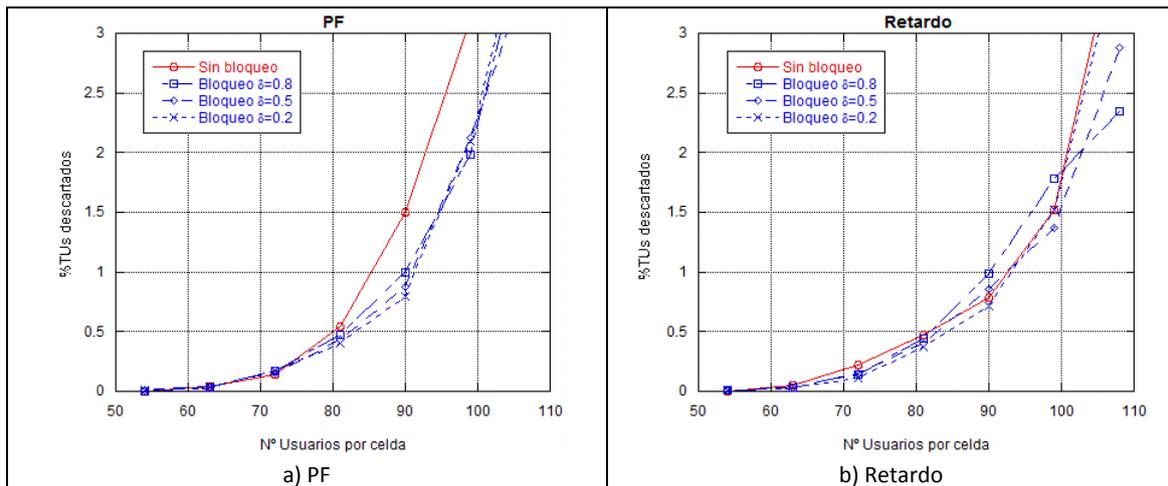


Figura 5.10: Algoritmos PF y Retardo con/sin bloqueo con tráfico uniforme (rA).

Para ver de qué forma afectan estos cambios en las prestaciones ofrecidas por algunos de los algoritmos al aplicar bloqueo, compararemos también los resultados para los distintos algoritmos entre sí con los que teníamos antes de aplicarlo en los casos en los que se aprecien diferencias significativas.

En la figura 5.11b podemos ver el efecto de aplicar bloqueo inicial en el caso del tráfico uniforme. Se observa como la mejora experimentada por el algoritmo PF gracias a ello hace que acerque sus prestaciones a las de los algoritmos EXP, 1-UEPS, etc., mientras que cuando no aplicábamos bloqueo funcionaba bastante peor que ellos, como podemos ver en la figura 5.11a. Ahora las diferencias entre los comportamientos de los distintos algoritmos son menores, pero ha de comentarse que este efecto positivo en el algoritmo PF no redundará en ventaja alguna para el objetivo de aumentar la capacidad de la celda, puesto que no consigue mejorar los resultados del resto de algoritmos en ningún caso.

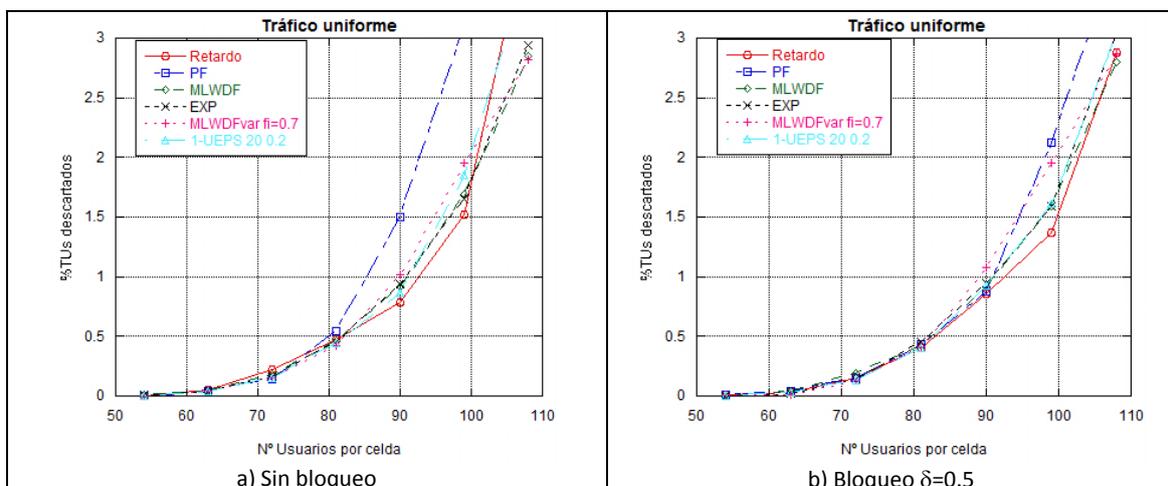


Figura 5.11: Comparación con/sin bloqueo para los distintos algoritmos (rA). Tráfico uniforme.

Scheduling y provisión de QoS en redes móviles 4G OFDMA

También experimenta cambios reseñables el Retardo cuando es empleado en un escenario multiservicio. En este caso, como vemos en la figura 5.12, al aplicar este tipo de bloqueo conseguimos que este algoritmo no experimente el aumento repentino de descartes al aumentar la carga del sistema como en el caso sin bloqueo (en concreto en la gráfica, al pasar de 99 a 108 usuarios). Sin embargo vemos que en algunos casos se alcanza el límite del 1% de descartes con menor número de usuarios que antes, consecuencia de bloquear a los que tienen malas condiciones de canal, que serán relegados al final y no les da tiempo a ser transmitidos. Aun así, si aplicamos el bloqueo sólo a aquellos paquetes que han consumido como mucho un 20% de su cota máxima ($\delta = 0.2$), mantenemos los niveles de descarte del caso sin bloqueo y evitamos que se disparen los descartes al alcanzar cierto nivel de carga. Sólo estaríamos bloqueando a los paquetes con malas condiciones de canal (han de emplear MCS0) y que aún les queda tiempo para alcanzar su cota máxima.

El algoritmo 1-UEPS experimenta también ciertos cambios en su comportamiento al aplicar bloqueo en el escenario con dos tipos diferentes de tráfico. Como puede verse en la figura 5.12a₂, al bloquear a los usuarios con peores condiciones de transmisión conseguimos un efecto similar al que se consiguió con el Retardo, evitando que el porcentaje de descartes crezca muy rápidamente cuando incrementamos la carga del sistema. Con este algoritmo vemos el efecto comentado con los valores de $\delta = 0.5$ y 0.8 , que son los que afectan a mayor cantidad de transmisiones, produciéndose también en este caso como en el del Retardo que se alcanza el límite de descartes antes de lo que sucedía en el caso sin bloqueo. Mientras tanto, con $\delta = 0.2$ vemos que al afectar el bloqueo a una parte muy pequeña de los paquetes, su comportamiento es muy similar al que se da cuando no existe bloqueo alguno. Por último, el PF empeora en general sus resultados, tanto para el tráfico de 50ms como para el de 250ms, por lo que no sería recomendable el uso de bloqueo en un escenario multiservicio si se empleara este algoritmo.

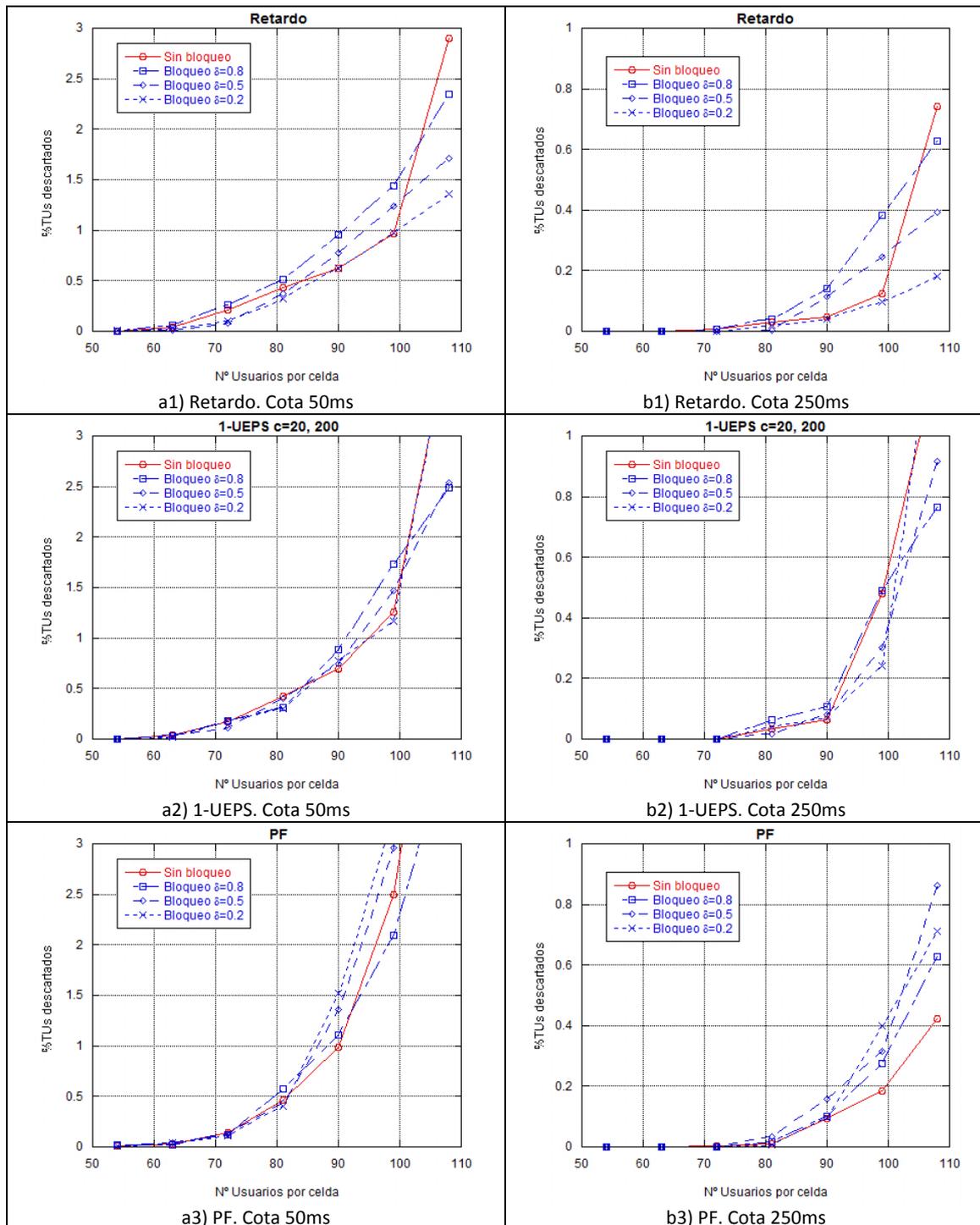


Figura 5.12: Algoritmos Retardo, 1-UEPS y PF con/sin bloqueo en un escenario multiservicio (50% $d_{max}=50ms$ +50% $d_{max}=250ms$) (rA).

En la figura 5.13 se observa un efecto del bloqueo interesante, gracias a la mejora del Retardo en el escenario multiservicio, uno de los algoritmos que da mejores resultados. Como puede observarse, al aplicar el algoritmo sin ningún tipo de bloqueo, su comportamiento es muy parecido al del MLWDFvar hasta llegar al límite de los requerimientos de QoS. Al seguir creciendo la carga en el sistema, se empieza a

acumular mucho retardo en los buffers y el algoritmo Retardo empieza a incrementar rápidamente su porcentaje de descartes. Gracias a la aplicación de bloqueo, si se observa la figura 5.13b, se aprecia como ahora el algoritmo Retardo ya no sufre este efecto, obteniéndose con él unos resultados incluso mejores que los que ofrece el algoritmo MLWDFvar. En este caso sí que podemos hablar de un efecto positivo del bloqueo, puesto que conseguimos que un algoritmo mejore las prestaciones que obteníamos con cualquiera otro en el caso sin bloqueo.

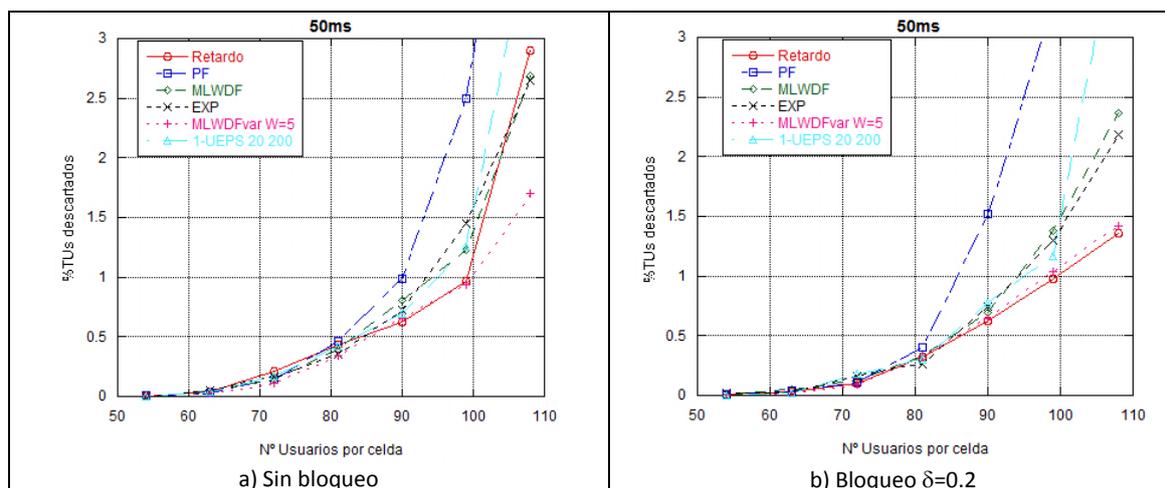


Figura 5.13: Comparación con/sin bloqueo para los distintos algoritmos (rA).
Escenario multiservicio (50% $d_{max}=50ms$ +50% $d_{max}=250ms$).

El bloqueo inicial de los paquetes con peores condiciones de canal afecta a algunos algoritmos en mayor medida que a otros, pero se hace interesante cuando mejora las prestaciones de los algoritmos que, sin aplicar este bloqueo, ya ofrecían unos buenos resultados. No obstante, en realidad las mejoras introducidas son despreciables si se tiene en cuenta que no es posible elevar el número de usuarios soportables, satisfaciendo la cota máxima de un 1% en la probabilidad de descarte.

5.3.2: Bloqueo por retardo. Ráfagas B.

En este apartado se analiza un escenario similar al del apartado anterior pero en este caso empleando ráfagas de tipo B, con tiempos actividad y silencio de 300 y 900 milisegundos respectivamente.

Uno de los casos para los que aplicar bloqueo afecta positivamente en esta situación es el algoritmo Retardo. Éste mejorará en ambos casos, tanto con tráfico uniforme como con el mezclado, ya que al haber periodos de actividad más largos se producen mayores acumulaciones de retardo y el bloqueo ayuda en mayor medida eliminando a los usuarios con peores condiciones de canal para que no bloqueen las salidas del resto. En la figura 5.14a vemos los resultados obtenidos con los distintos bloqueos para el caso de tráfico uniforme. Vemos como en todos los casos conseguimos

Scheduling y provisión de QoS en redes móviles 4G OFDMA

aumentar la capacidad de la celda respetando las condiciones impuestas de calidad de servicio, aunque vemos como con este tipo de tráfico el bloqueo no es capaz de frenar la repentina crecida del porcentaje de descartes con el aumento de usuarios. El bloqueo también tiene resultados positivos al aplicarse sobre otros algoritmos, como sucede con el EXP con tráfico uniforme. En la figura 5.17 se observa como en general los porcentajes de descarte para los distintos valores de δ del bloqueo son inferiores a los que teníamos sin aplicarlo, consiguiendo incluso en el caso de $\delta=0.8$ aumentar considerablemente la capacidad del sistema.

En cambio, el bloqueo tiene cierto efecto negativo sobre el algoritmo PF con baja carga, especialmente con $\delta=0.8$ y $\delta=0.5$, aunque esto no afecta negativamente a la capacidad del sistema ni a las situaciones con carga elevada. El algoritmo que sí ve reducidas sus prestaciones al aplicar bloqueo en este escenario es el 1-UEPS. Los resultados empeoran considerablemente, por lo que no sería nada recomendable la aplicación de bloqueo con este algoritmo. En el peor caso, con $\delta=0.2$, podría llegar a reducirse la capacidad de 89 a 82 usuarios aproximadamente (fig. 5.14d, 1% descarte), lo cual supone una pérdida de capacidad del 8%.

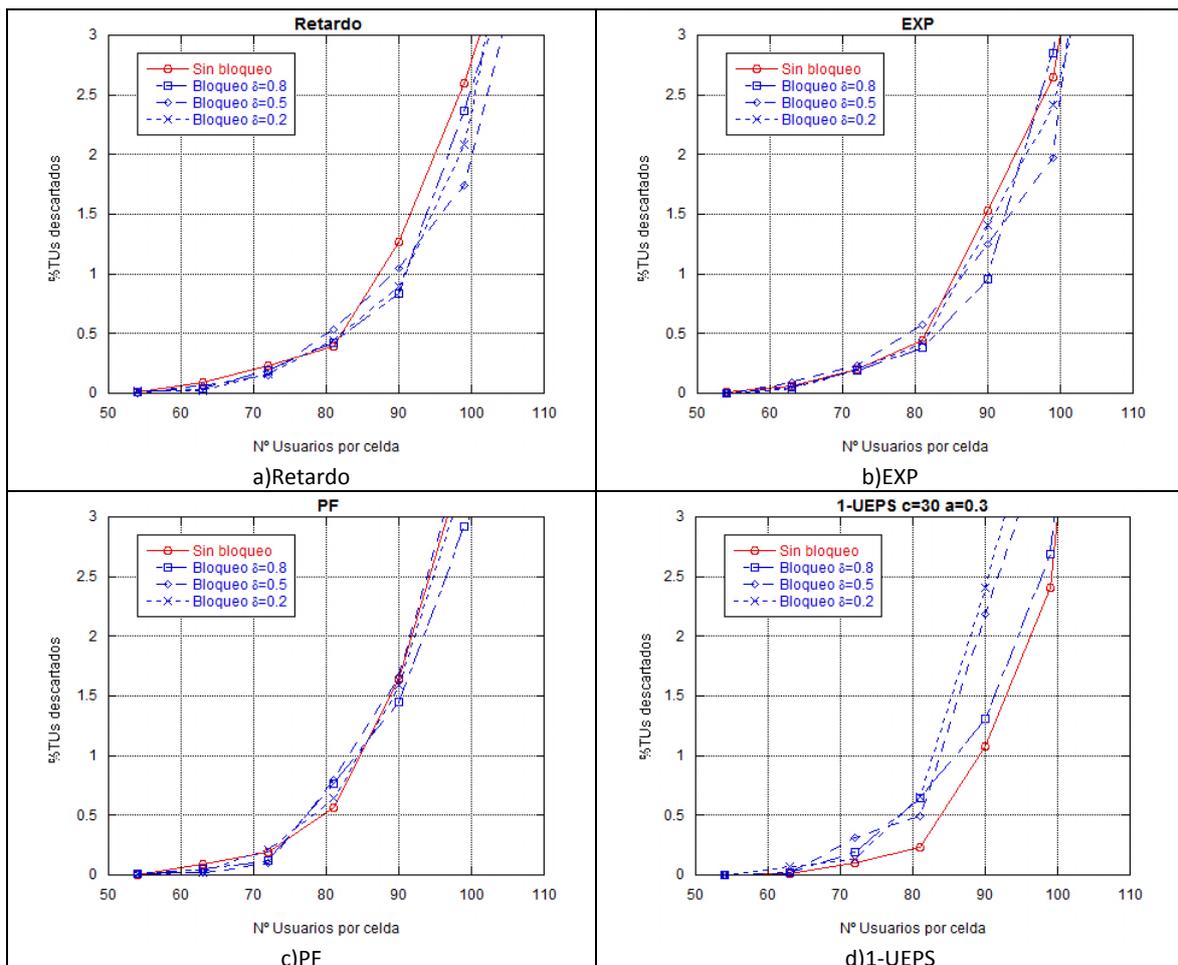


Figura 5.14: Algoritmos Retardo, EXP, PF y 1-UEPS con/sin bloqueo con tráfico uniforme (rB).

Scheduling y provisión de QoS en redes móviles 4G OFDMA

De nuevo compararemos los resultados con los que existían antes de aplicar ningún bloqueo en los casos en los que se aprecien diferencias significativas.

Vemos en la figura 5.15 que, en el caso del tráfico uniforme, la aplicación de bloqueo inicial conlleva la mejora de los resultados de los algoritmos Retardo y EXP, que se posicionan como aquellos que mejores prestaciones ofrecen. Se siguen viendo pequeñas diferencias entre los resultados de las distintas estrategias, pero la mejora obtenida, sobre todo en el caso del Retardo, hace que el algoritmo de scheduling con mejores prestaciones ahora consiga una mayor capacidad de la conseguíamos sin aplicar bloqueo inicial. Por el contrario, el 1-UEPS empeora considerablemente sus resultados, y pasa de ser el que mejores prestaciones ofrecía en el caso sin bloqueo a ser uno de los algoritmos que peor funcionan con el bloqueo aplicado.

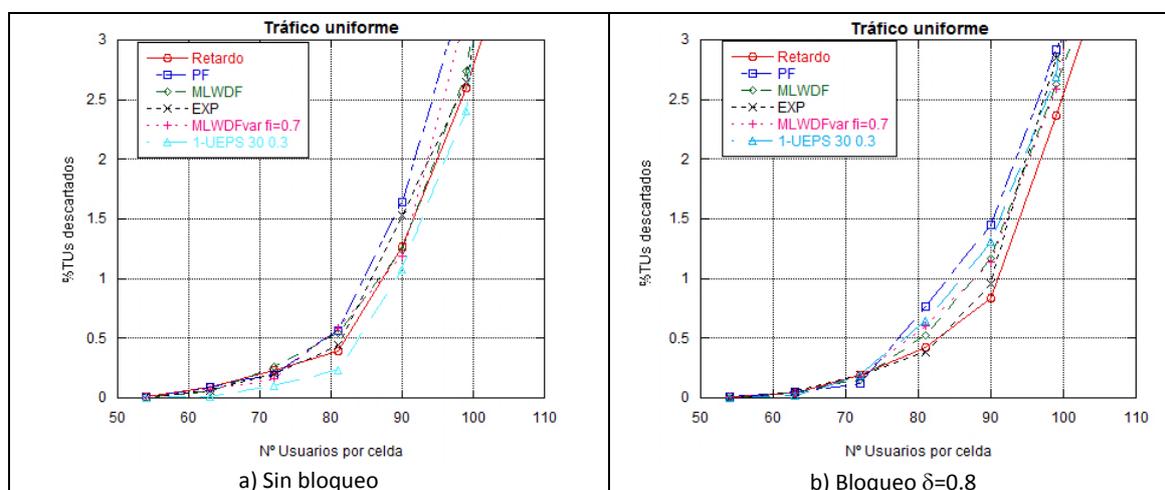


Figura 5.15: Resultados con/sin bloqueo para los distintos algoritmos (rB). Tráfico uniforme.

Si evaluamos ahora la mejora obtenida en el escenario multiservicio (figura 5.16), vemos que el efecto del bloqueo es bastante positivo en el algoritmo Retardo, puesto que conseguimos frenar la crecida del porcentaje de descartes cuando existe un nivel elevado de carga en la celda y aumentamos la capacidad de la misma admitiendo aproximadamente 6 usuarios más manteniendo el descarte de TUs por debajo del 1%. En este caso se juntan los efectos positivos del bloqueo sobre ráfagas largas y los que tiene también sobre el tráfico mezclado. Otro caso que mejora sus prestaciones es el algoritmo MLWDFvar con tráfico mezclado. Puede verse en la figura 5.16a₂ como se aumenta el número de usuarios admisibles por debajo del 1% de descartes, y teniendo en cuenta que éste es uno de los métodos que mejores resultados da, se estaría logrando una mejora ciertamente relevante. En la figura se aprecia que en algunos casos el bloqueo hace crecer el descarte por encima del que existía sin bloqueo para un número muy alto de usuarios, pero $\delta=0.5$ logra evitar esta circunstancia, siendo este valor el que se seleccionaría si se decidiera utilizar este algoritmo en un escenario de características similares a las aquí evaluadas. Habría que tener cuidado con la crecida que se aprecia en el porcentaje de descartes cuando se incrementa el número de usuarios para el tráfico de 250ms, como se aprecia en la figura 5.16b₂, aunque

Scheduling y provisión de QoS en redes móviles 4G OFDMA

afecta en poca medida a la capacidad del sistema ya que por debajo del 1% prácticamente no empeora. Al igual que sucedía con las ráfagas A, el algoritmo PF empeora sus resultados al aplicar bloqueo con cualquiera de los dos tipos de tráfico. Empleando el valor $\delta = 0.8$ en este caso conseguimos cierta mejora, por lo que emplearíamos este valor para el bloqueo en caso de utilizar este algoritmo de scheduling en un escenario de este tipo.

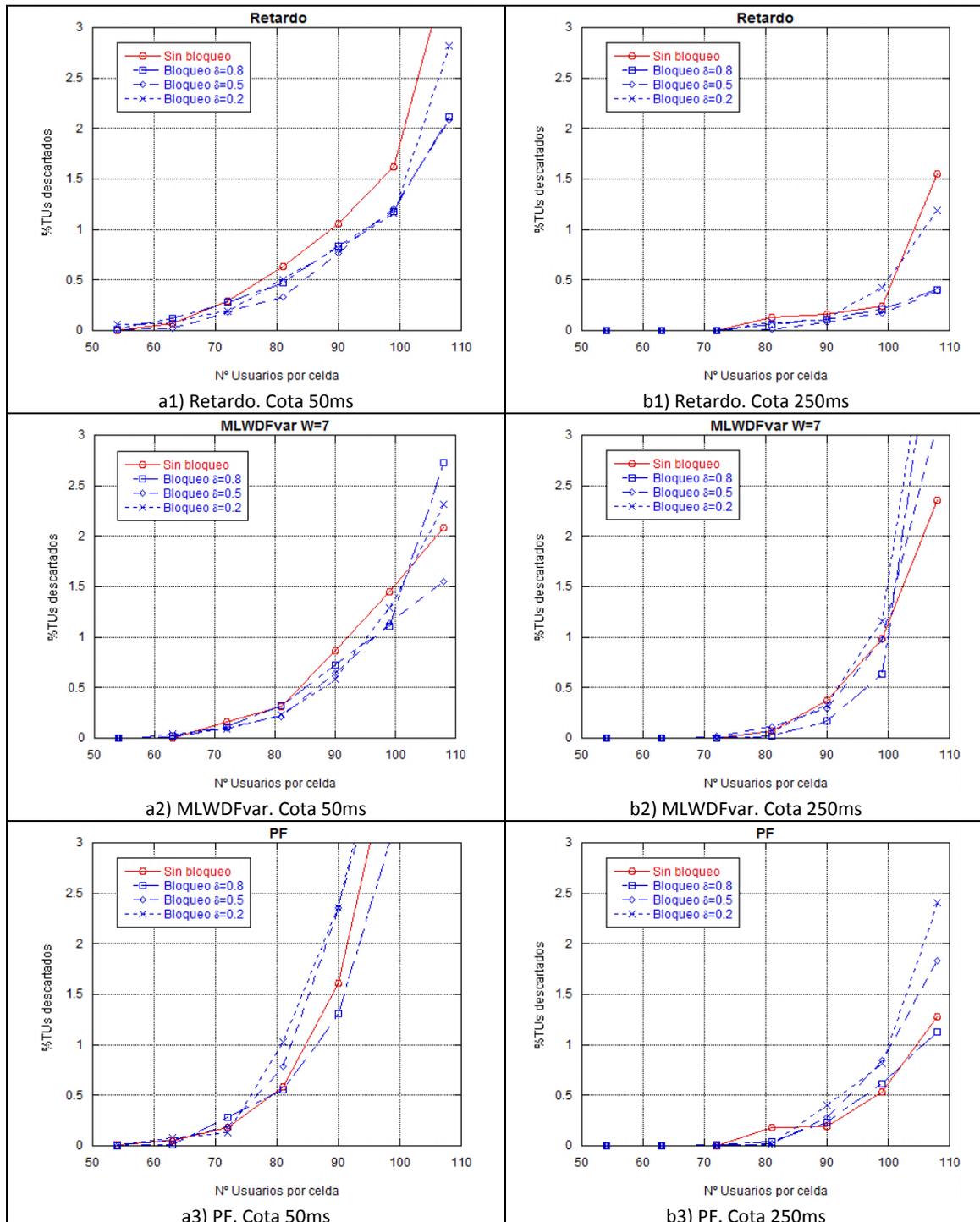


Figura 5.16: Algoritmos Retardo, MLWDFvar y PF con/sin bloqueo en un escenario multiservicio (50% $d_{max}=50ms$ +50% $d_{max}=250ms$) (rB).

Scheduling y provisión de QoS en redes móviles 4G OFDMA

En la figura 5.17b se aprecia cómo, al mejorar con el bloqueo las prestaciones de MLWDFvar y Retardo, se incrementan las diferencias entre los resultados de los distintos algoritmos, puesto que éstos ya eran los que mejor funcionaban en el caso sin bloqueo. A ello también colabora el peor funcionamiento con $\delta=0.5$ del PF, ya que hemos querido usar este valor para la comparación. En el distanciamiento de los resultados del PF con el resto de algoritmos también influye cierta mejora en los algoritmos MLWDF y EXP, como puede observarse en las figuras adjuntas en el anexo B.

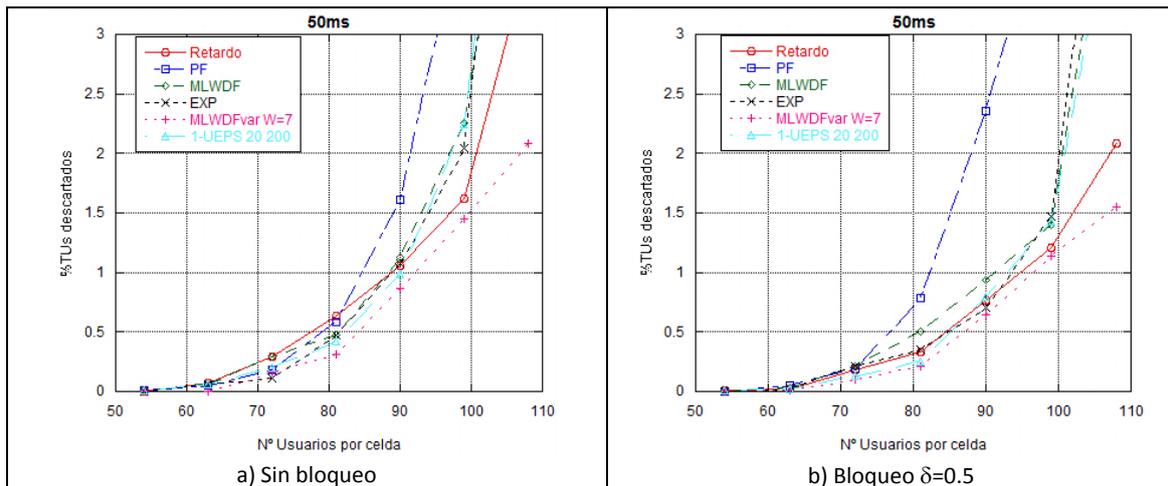


Figura 5.17: Comparación con/sin bloqueo para los distintos algoritmos (rB). Escenario multiservicio (50% $d_{max}=50ms$ +50% $d_{max}=250ms$).

En cualquier caso, de nuevo la aplicación de bloqueo con tráfico mezclado tiene efectos positivos, puesto que conseguimos aumentar la capacidad del sistema gracias a la mejora de los algoritmos que mejor funcionan en este tipo de escenario. En este caso, al mejorar el MLWDFvar al mismo tiempo que el Retardo, el primero mantiene la ventaja que tenía en el caso sin bloqueo y sigue siendo el algoritmo con mejores prestaciones.

5.3.3: Bloqueo por estado de los buffers. Ráfagas A.

Estudiaremos ahora el efecto del bloqueo realizado en función del tiempo estimado que sería necesario para el vaciado de los buffers. Empezaremos poniendo nuestra atención en las ráfagas A, las de más corta duración.

Tal y como sucedía al aplicar bloqueo teniendo en cuenta únicamente el paquete de cabecera del buffer, el algoritmo PF con tráfico uniforme es uno de los que mejora (fig. 5.18), habiendo pocas variaciones entre los resultados obtenidos con los distintos valores de δ . También experimenta cierta mejora el algoritmo Retardo, sobre todo cuando el sistema tiene un elevado nivel de carga (fig. 5.18b). Si observamos en

Scheduling y provisión de QoS en redes móviles 4G OFDMA

concreto para $\delta=0.5$, no incrementa el porcentaje de descarte cuando hay baja carga (lo que sí sucede con $\delta=0.8$) y conseguimos la mejora comentada.

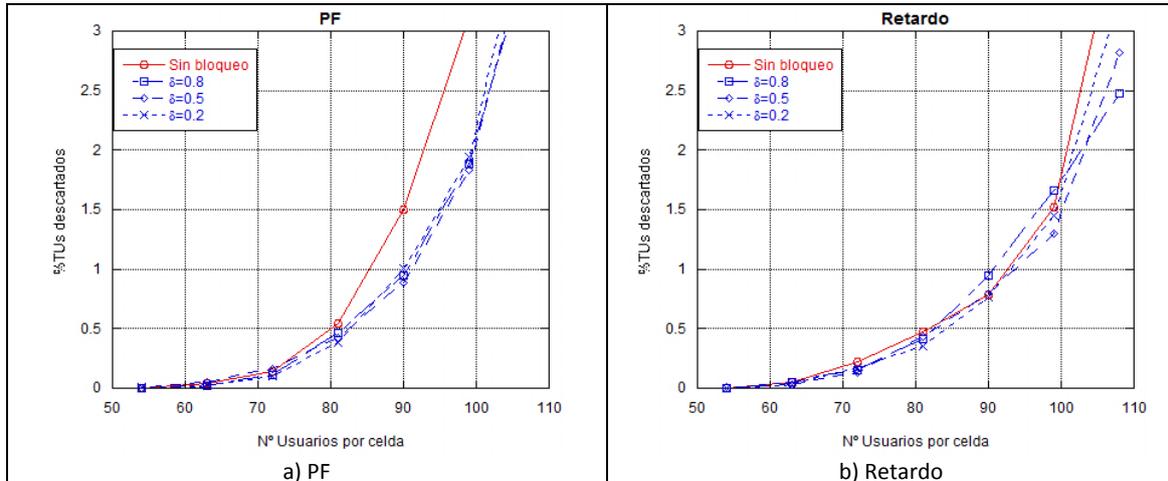


Figura 5.18: Algoritmos PF y Retardo con/sin bloqueo con tráfico uniforme (rA).

La figura 5.19 muestra las diferencias en el caso del tráfico uniforme, para un bloqueo con $\delta=0.5$ respecto a la evaluación de los algoritmos sin bloqueo. Como puede apreciarse, el algoritmo PF, a pesar de la mejora, sigue obteniendo unos resultados peores que el resto de algoritmos. En el resto de algoritmos, excepto Retardo, no se aprecian mejoras significativas. Por el contrario, la mejora experimentada por el Retardo hace que este método mejore las prestaciones del resto para condiciones de carga más altas, lo cual no sucedía cuando no aplicábamos bloqueo alguno.

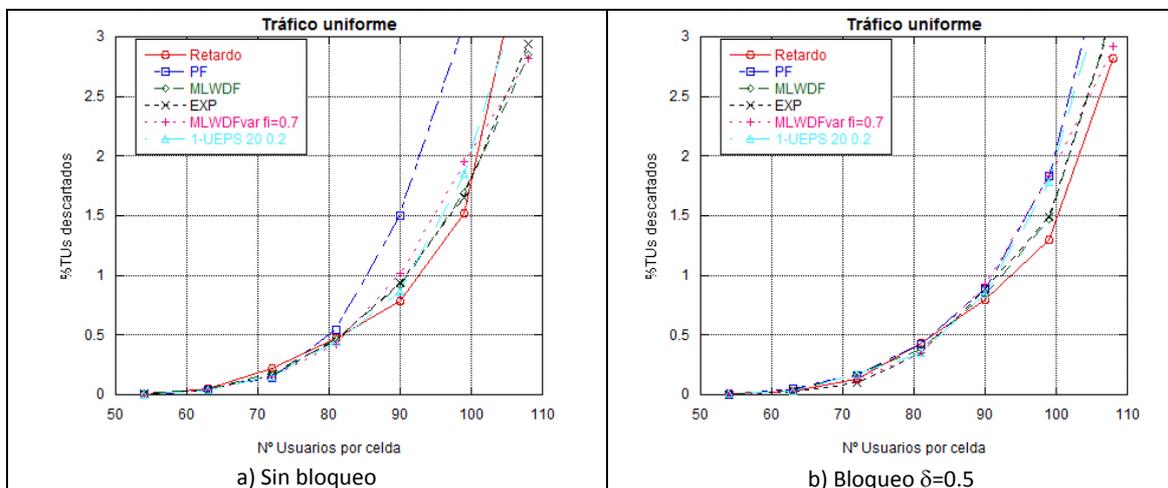


Figura 5.19: Comparación con/sin bloqueo para los distintos algoritmos (rA). Tráfico uniforme.

En el caso de un escenario multiservicio con dos tipos de tráfico ($d_{max1}=50ms$, $d_{max2}=250ms$), el algoritmo que experimenta un cambio más reseñable es el Retardo (figura 5.20). Como se aprecia en la figura, conseguimos evitar que se dispare el porcentaje de descartes al incrementar mucho la carga del sistema, de nuevo a costa de reducir algo la capacidad por debajo del 1%. En este caso este incremento en los

Scheduling y provisión de QoS en redes móviles 4G OFDMA

descartes con baja carga es menor que con el otro tipo de bloqueo, ya que en este caso se evita parte de este empeoramiento añadiendo el factor del estado del buffer.

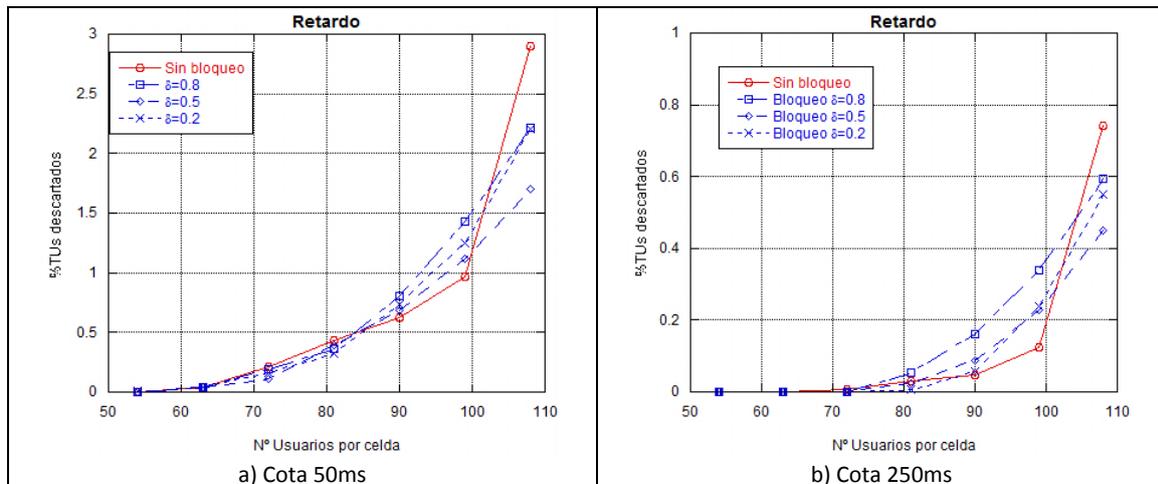


Figura 5.20: Algoritmo Retardo con/sin bloqueo en un escenario multiservicio (50% $d_{max}=50ms$ +50% $d_{max}=250ms$) (rA).

En la figura 5.21 se observan las diferencias para el servicio $d_{max}=50ms$, en el escenario multiservicio, y las distintas estrategias. Se ha escogido el valor de $\delta=0.5$ puesto que es el que mejor resultado ofrece para el algoritmo Retardo, que es el que sufre cambios más evidentes. Podemos ver que a causa de ese empeoramiento sufrido por el Retardo en condiciones de baja carga, destaca algo más el resultado del MLWDFvar, que prácticamente no cambia respecto al obtenido sin bloqueo. En cambio, al aumenta el número de usuarios (ej: 108) se igualan los resultados de los dos algoritmos. También es destacable la reducción del descarte para 108 usuarios de los algoritmos MLWDF y EXP, mostrando que el bloqueo inicial es sensiblemente más eficaz cuando la carga del sistema es alta. Mientras no se permita transmitir a los usuarios con peores condiciones hasta que no lo hayan hecho el resto, se evita que éstos monopolicen el uso del canal y se permite transmitir a otros usuarios. En general, en condiciones de carga elevada, disminuye la probabilidad de descarte. En lo que respecta al tráfico de 250ms, los cambios son muy ligeros excepto en caso de carga elevada, situación en la que los algoritmos MLWDFvar y Retardo mejoran ligeramente sus resultados mientras que el resto los empeoran.

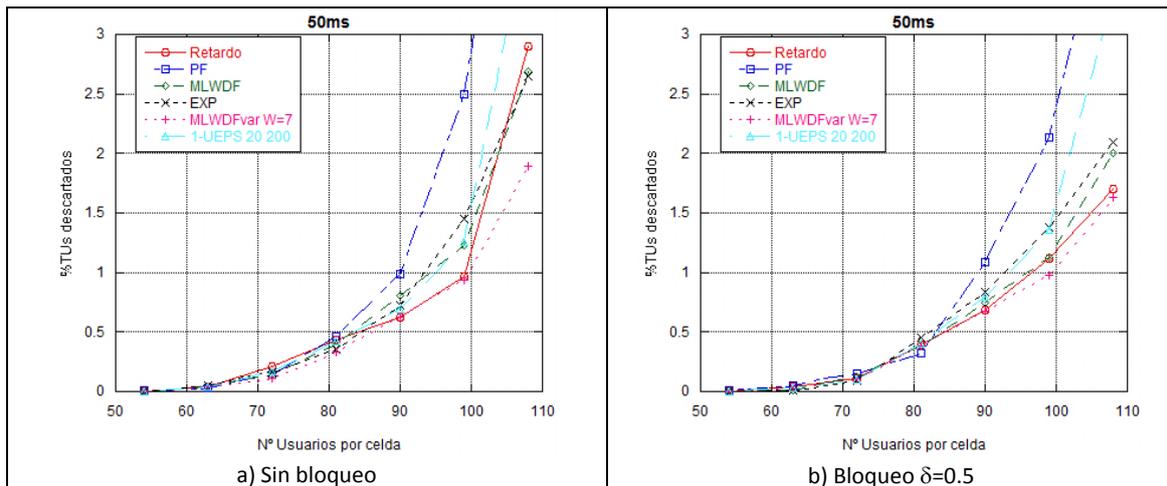


Figura 5.21: Comparación con/sin bloqueo para los distintos algoritmos (rA).
Escenario multiservicio (50% $d_{max}=50ms$ +50% $d_{max}=250ms$).

5.3.4: Bloqueo por estado de los buffers. Ráfagas B.

En este apartado se mantiene el mismo bloqueo que en el anterior pero se cambian el tipo de ráfagas empleadas por las de $T_{ON}=300ms$, $T_{OFF}=900ms$.

En este escenario destaca, entre otras, la mejora experimentada por el algoritmo MLWDF cuando existe un único tipo de tráfico. En la figura 5.22b puede apreciarse como, para el caso de $\delta =0.8$, MLWDF mejora sensiblemente sus resultados, admitiendo a varios usuarios más antes de alcanzar el umbral del 1% de descartes. Dejando para el final las transmisiones de los usuarios con peores condiciones conseguimos que el resto transmita a tiempo. Para ese mismo valor del parámetro δ conseguimos también un aumento considerable de la capacidad de la celda con el algoritmo EXP, como puede apreciarse en la figura 5.22c. También el algoritmo Retardo experimenta cierta mejoría, al igual que sucedía empleando el otro tipo de bloqueo, pero esta no es tanta como la del MLWDF o el EXP, tal y como puede verse en la figura 5.22a. En este caso cualquier valor de δ mejora a los resultados obtenidos sin bloqueo.

Por el contrario, como puede verse en la figura 5.22d, el algoritmo 1-UEPS experimenta un aumento de los porcentajes de descarte en situaciones de baja carga, causando una reducción de la capacidad del sistema al aplicar bloqueo. Puede verse como para $\delta=0.5$ perderíamos unos tres usuarios para mantener los niveles exigidos de QoS.

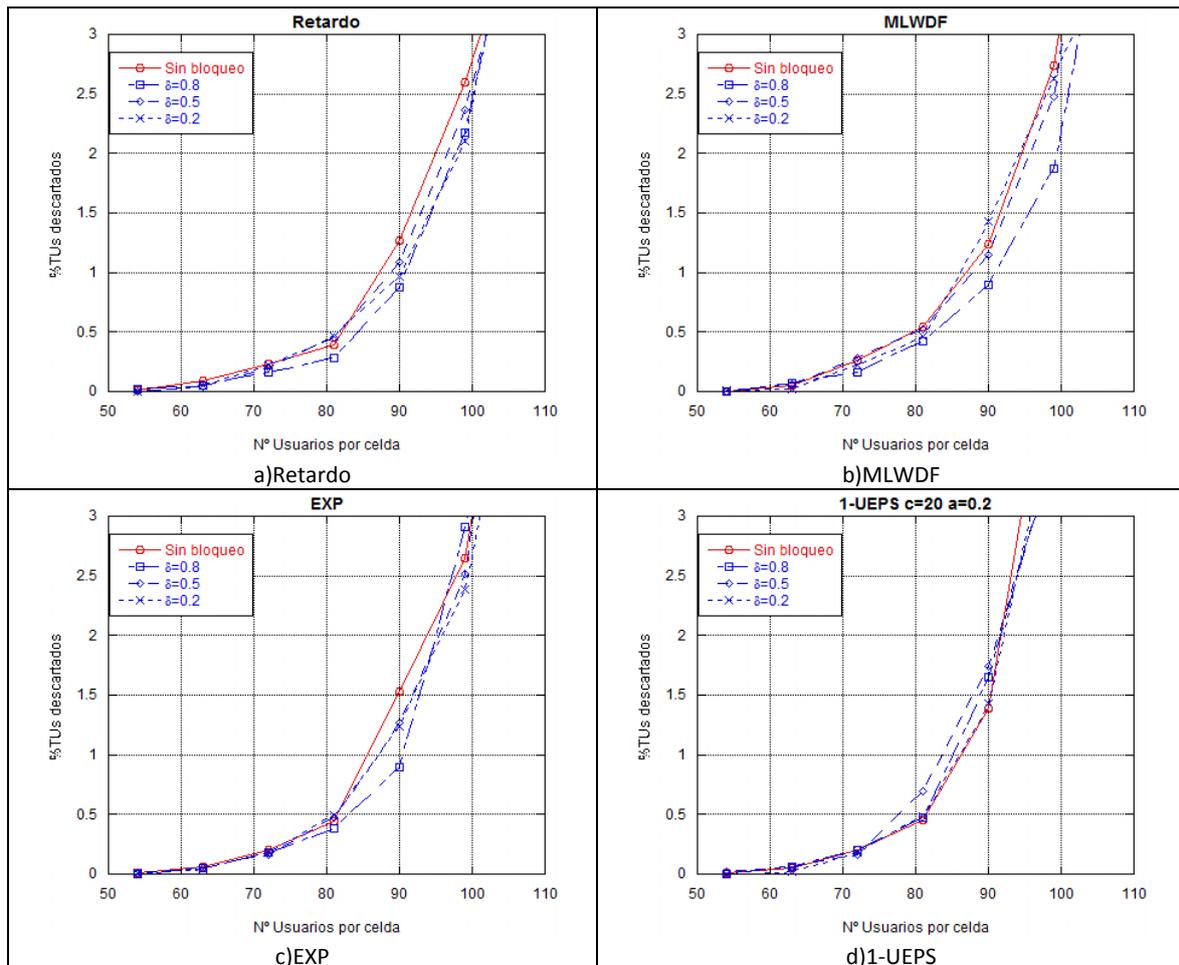


Figura 5.22: Algoritmos Retardo, MLWDF, EXP y 1-UEPS con/sin bloqueo con tráfico uniforme (rB).

La figura 5.23 muestra una comparación de resultados para los distintos algoritmos. Se aprecia un aumento de las diferencias entre ellos. Son claramente destacables los buenos resultados que ofrece en este caso el MLWDF. Los tres algoritmos que experimentaban mejora, MLWDF, EXP y Retardo tienen unas prestaciones similares hasta alcanzar el 1%, pero es el MLWDF el que después crece más lentamente.

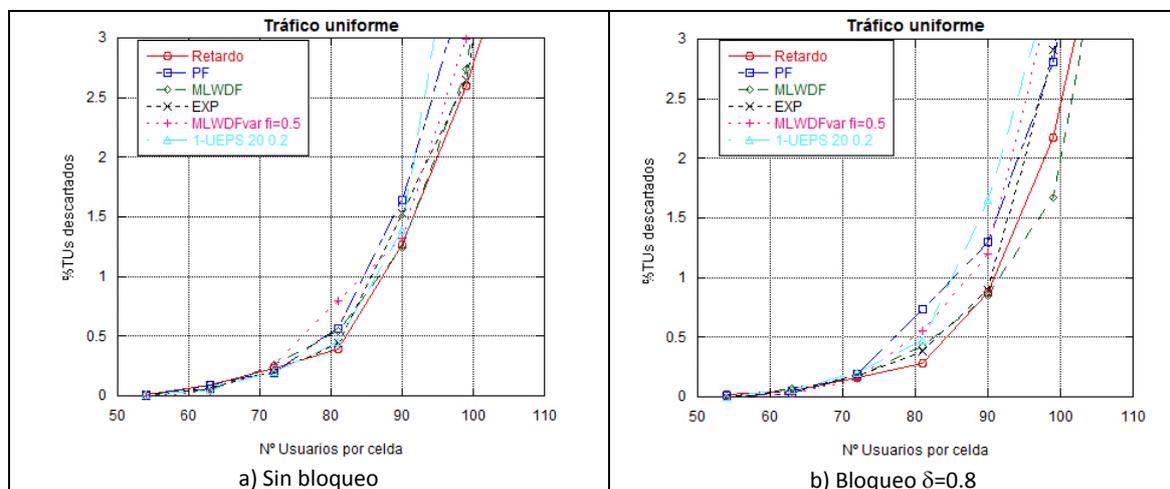


Figura 5.23: Resultados con/sin bloqueo para los distintos algoritmos (rB). Tráfico uniforme.

Scheduling y provisión de QoS en redes móviles 4G OFDMA

En el escenario multiservicio, el algoritmo que garantiza el máximo número de usuarios (con probabilidad de descarte inferior al 1%) es MLWDFvar, ya que mejora sus prestaciones cuando ya tenía las mejores en comparación con el resto de algoritmos. Como puede apreciarse en la figura 5.24, al aplicar el bloqueo conseguimos aumentar la capacidad del sistema garantizando las exigencias de QoS, a costa de una subida del número de descartes cuando incrementamos la carga (108 usuarios). Habría que tener especial cuidado con el efecto del bloqueo sobre el tráfico de 250ms, que empeora sus porcentajes de descarte como puede verse en la figura 5.24b₁, y que podría convertirse en el determinante a la hora de decidir la capacidad de la celda. En este caso esa circunstancia no llega a producirse, aunque el descarte para este tipo de tráfico crece demasiado cuando hay un número de usuarios elevado. También se observa cierta mejora en los resultados del algoritmo Retardo. Para todos los valores de δ mejora ligeramente la capacidad, pero destaca especialmente que para $\delta=0.8$ consigue frenar en gran medida el rápido incremento del porcentaje de descartes cuando sobrepasamos los 100 usuarios por celda. En este caso el tráfico de 250ms también se ve afectado positivamente, evitando también la fuerte crecida cuando hay mucha carga de datos en el sistema.

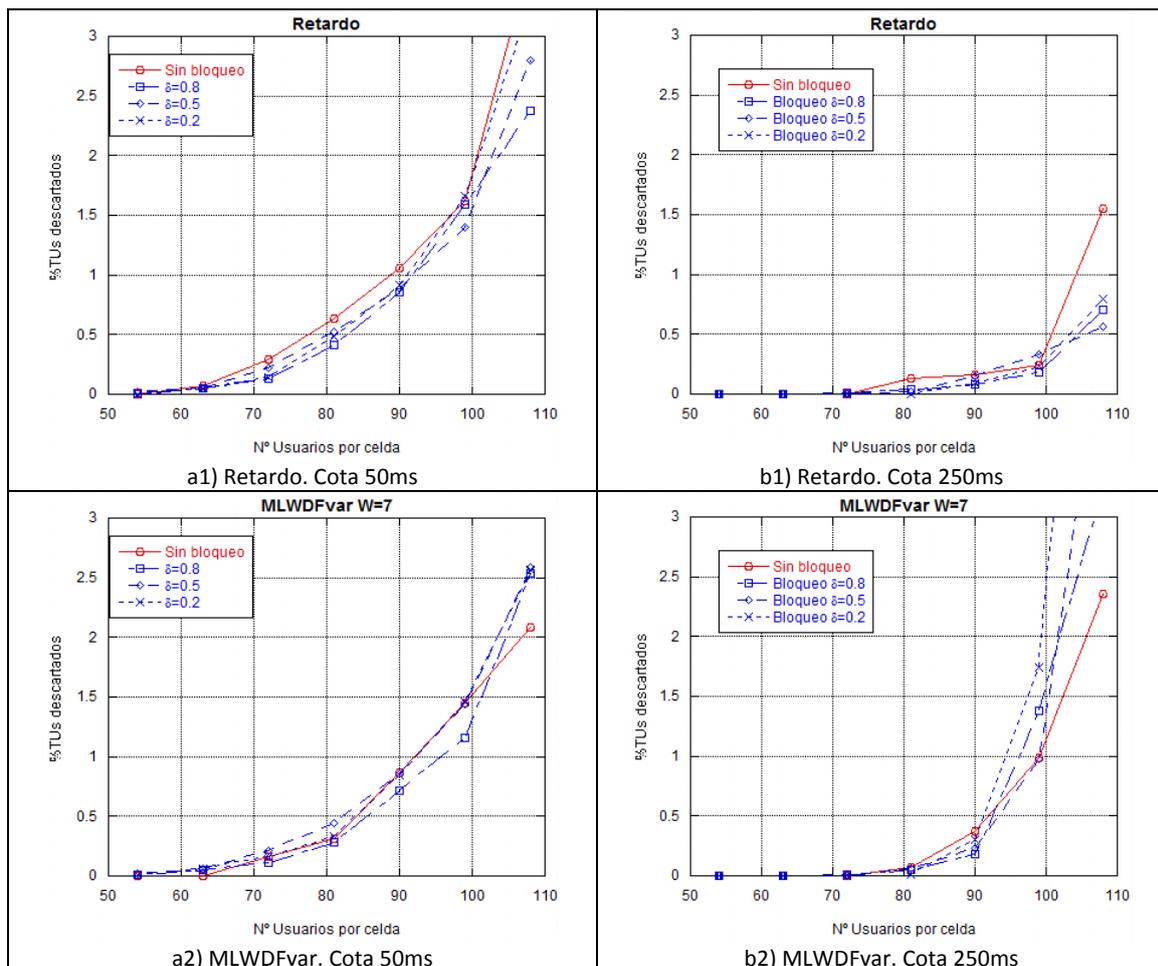


Figura 5.24: Algoritmos Retardo y MLWDFvar con/sin bloqueo en un escenario multiservicio (50% $d_{max}=50ms$ +50% $d_{max}=250ms$) (rB).

Scheduling y provisión de QoS en redes móviles 4G OFDMA

Al comparar entre si las distintas estrategias, en la figura 5.25b vemos que al mejorar con el bloqueo los resultados del MLWDFvar, podemos dar cabida a 4 o 5 usuarios más que con cualquiera otro algoritmo antes de alcanzar el 1% de descarte de TUs, consiguiendo una diferencia mayor que la existente en el caso sin bloqueo.

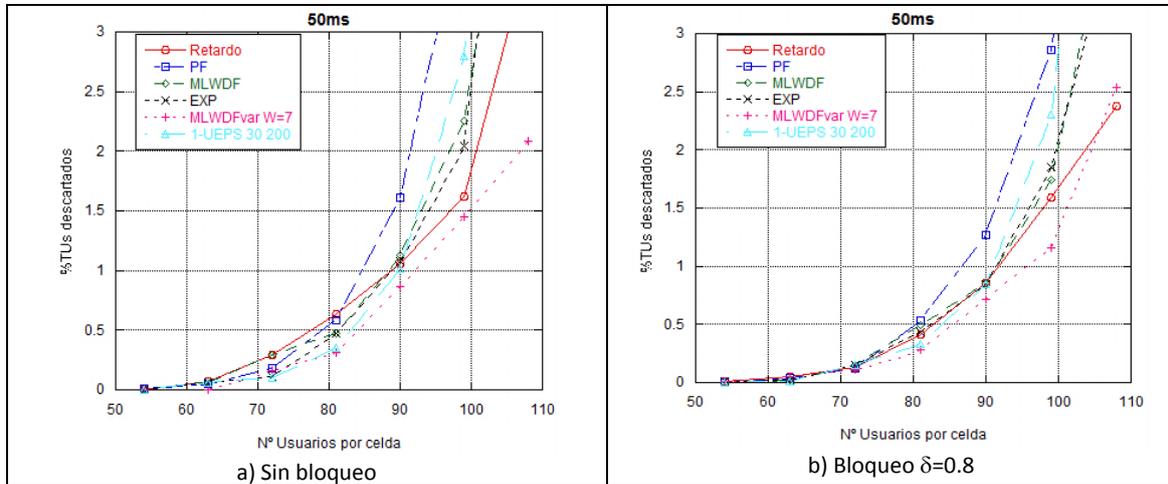


Figura 5.25: Comparación con/sin bloqueo para los distintos algoritmos (rB).

Escenario multiservicio (50% $d_{max}=50ms$ +50% $d_{max}=250ms$).

En este escenario de ráfagas de tipo B y con dos tipos de tráfico, la aplicación de este bloqueo resulta claramente positiva, ya que conseguimos mejorar los resultados del que ya era el algoritmo con mejores prestaciones, de forma similar a como sucedía en el apartado 5.3.2, pero en este caso además incrementando la diferencia respecto al resto de algoritmos. Habríamos de tener cuidado con el crecimiento del descarte para el tráfico de $d_{max}=250ms$, pero ciñéndonos al criterio de QoS exigida se estaría logrando una mejora.

5.4: Conclusiones sobre los resultados.

Al respecto de los resultados que se han ido obteniendo con las pruebas realizadas con los distintos tipos de ráfagas y tráficos y aplicando o no bloqueo, podemos concluir que los algoritmos de scheduling han de tener en cuenta, de un modo u otro, el tiempo que llevan en espera los distintos paquetes para ser transmitidos y las condiciones del canal por el que serán transmitidos.

La mayoría de los algoritmos utilizados durante esta memoria cumplen ambos requisitos, pero hay dos que no. El algoritmo PF sólo tiene en cuenta las condiciones del canal, fijándose en la tasa alcanzable en la transmisión. Esto es causa del pobre comportamiento que tiene en cuanto a número de usuarios que podría haber en una celda cumpliendo los requerimientos de QoS (porcentaje de descarte de TUs por debajo del 1%).

Scheduling y provisión de QoS en redes móviles 4G OFDMA

El algoritmo Retardo ignora estas condiciones de transmisión y se fija únicamente en el retardo acumulado respecto a la cota máxima permitida para el tipo de tráfico empleado. Al contrario de lo que sucedía con el PF, los resultados del algoritmo Retardo en cuanto a capacidad del sistema son bastante buenos, siendo habitualmente los mejores cuando se trata de tráfico uniforme y compitiendo con el MLWDFvar cuando existen varios tipos de tráfico. El problema que tiene es que en caso de haber un alto nivel de carga en el sistema los usuarios empiezan a acumular grandes cantidades de retardo y el porcentaje de descartes crece rápidamente, pudiendo alcanzar una situación de colapso del sistema en la que el retardo medio fuera igual o muy cercano a la cota máxima de retardo y se descartaran la mayoría de los paquetes. Una buena solución a este problema del algoritmo Retardo es la aplicación de bloqueo inicial, limitando el acceso a los recursos de los usuarios con peores condiciones de canal para evitar que bloqueen el sistema con sus lentas transmisiones. De esta forma le estamos haciendo depender de las condiciones del canal, y puede verse en los resultados obtenidos que con esto conseguimos limitar ese crecimiento de los descartes cuando la carga del sistema crece.

MLWDFvar obtiene también muy buenos resultados en el escenario multiservicio, consiguiendo en muchos casos los mejores resultados. Esto se debe al uso por parte de este algoritmo de un peso W para diferenciar entre los tipos de tráfico y otorgar una mayor prioridad a aquellos paquetes cuya cota máxima de retardo es más pequeña. Esto lleva consigo un incremento de los descartes que se producen para el tráfico de cota mayor, pero se observa en los resultados que, si sus requisitos de QoS son los mismos para ambos tipos de tráfico, en ningún caso llegaría a ser más restrictivo el descarte producido en el de mayor cota.

6. CONCLUSIONES Y LÍNEAS FUTURAS

En este PFC se han evaluado distintos algoritmos de scheduling para el enlace descendente en un sistema basado en la tecnología LTE. Se necesita un diseño eficaz en la gestión de recursos radio (RRM) para que la provisión de calidad de servicio sea eficiente, y para ello se emplearon estrategias de scheduling diversas que priorizan el acceso a estos recursos considerando criterios de tasa y/o retardo y condiciones del canal.

El propósito del análisis realizado ha sido maximizar la capacidad del sistema en cuanto a número de usuarios asumibles por celda respetando los requerimientos de QoS, siendo para ello necesaria la búsqueda de algoritmos que minimizaran el porcentaje de descartes de TUs.

Apoyándonos en una plataforma de simulación previamente desarrollada, se ha estudiado el comportamiento de varios algoritmos de scheduling en función de los distintos valores asignados a sus parámetros: Urgency and Efficiency-based Packet Scheduling (UEPSder y 1-UEPS), una variante del Modified Largest Weighted Delay First (MLWDFvar) y los algoritmos que realizan una organización previa en subgrupos (temp3reorg y npaq3reorg). Optimizaremos el funcionamiento de un algoritmo si encontramos los parámetros adecuados para el escenario en el que será utilizado.

Se han incluido también comparaciones con algunas estrategias que ya se encontraban en el simulador (Retardo, PF, MLWDF, EXP) y con una combinación de los algoritmos PF y Retardo (PFret). Para un mejor estudio se emplean diferentes escenarios de evaluación. Se compararán los porcentajes de descartes de cada uno de los algoritmos en situaciones de tráfico uniforme con cota máxima de retardo 50ms o en un escenario multiservicio (50% $d_{max}=50ms$ +50% $d_{max}=250ms$), para ambos casos con dos tipos de ráfagas: cortas ($T_{ON}=30ms$ y $T_{OFF}=90ms$) y largas ($T_{ON}=300ms$ y $T_{OFF}=900ms$).

Para evitar que los usuarios con peores condiciones de servicio monopolicen el uso del canal, se ha propuesto aplicar un bloqueo a éstos, permitiéndoles transmitir sólo en caso de que ya se hayan transmitido el resto de paquetes, lleven cierto retardo acumulado (bloqueo por retardo) o tengan un número determinado de paquetes esperando ser transmitidos en el buffer (bloqueo por estado del buffer). Este bloqueo afecta de forma distinta dependiendo del algoritmo de scheduling y el escenario de evaluación en el que se aplique, por lo que debería estudiarse previamente la conveniencia o no de su aplicación.

Scheduling y provisión de QoS en redes móviles 4G OFDMA

Por todos los resultados obtenidos en las distintas circunstancias evaluadas se concluye que un buen algoritmo de scheduling ha de tener en cuenta tanto el retardo que van acumulando los paquetes en espera como la tasa a la que puede ser transmitido por las condiciones de canal. Si bien el criterio a satisfacer es el de la calidad de servicio garantizada, han de tenerse en cuenta las acumulaciones de retardo y los rápidos crecimientos del descarte de algunos algoritmos, y tratar de evitarlos en la medida de lo posible.

El estudio del efecto de dar distintos pesos a los distintos tipos de tráfico (como sucede en el MLWDFvar) en las diferentes estrategias de scheduling empleadas podría ser un interesante estudio a realizar, pretendiendo priorizar los paquetes con una menor cota máxima de retardo frente a aquellos que cuentan con más tiempo para ser servidos.

También podrían probarse nuevos algoritmos existentes en la literatura, buscando aumentar la capacidad del sistema entre aquellos que compatibilicen un bajo porcentaje de descartes con la menor acumulación posible de retardo conforme crece la carga del sistema.

Otra posibilidad sería evaluar los resultados con otros esquemas de asignación de recursos. En este proyecto sólo se empleó FFRopa ya que en un estudio anterior mejoraba en todos los casos los resultados del FFR.

Además las simulaciones se han realizado para una velocidad de los terminales (UEs) de 3km/h, sería interesante obtener el comportamiento escenarios con distintas velocidades de desplazamiento de los UEs, para observar la evolución de los distintos algoritmos y en qué medida se ven afectados por la movilidad de los usuarios.

REFERENCIAS

- [1] A. Hernández, I. Guío, A. Valdovinos; *“Downlink Scheduling for Intercell Interference Fluctuation Mitigation in Partial-Loaded Broadband Cellular OFDMA Systems”*. Proc. IEEE International Conf. on Ultra Modern Telecommunications (ICUMT'09), pp.1-6, Oct. 2009.
- [2] 3GPP TSG RAN WG1#42 R1-050764. *“Inter-cell interference handling for E-UTRA”*. Technical report, Ericsson, September 2005.
- [3] A. Pokhariyal, K.I. Pedersen, G. Monghal, I.Z. Kovacs, C. Rosa, T.E. Kolding, P.E. Mogensen, *“HARQ Aware Frequency Domain Packet Scheduler with Different Degrees of Fairness for the UTRAN Long Term Evolution”*, IEEE 65th Vehicular Technology Conference, 2007. VTC2007-Spring.
- [4] A. Pokhariyal, K.I. Pedersen, G. Monghal, I.Z. Kovacs, C. Rosa, T.E. Kolding, P.E. Mogensen, *“HARQ Aware Frequency Domain Packet Scheduler with Different Degrees of Fairness for the UTRAN Long Term Evolution”*, IEEE 65th Vehicular Technology Conference, 2007. VTC2007-Spring.
- [5] Suk-Bok Lee, Ioannis Pefkianakis, Adam Meyerson, Shugong Xu, and Songwu Lu, *“Proportional Fair Frequency-Domain Packet Scheduling for 3GPP LTE Uplink”*, IEEE INFOCOM 2009 mini-symposium.
- [6] M. Andrews, K.Kumaran, K.Ramanan, A.Stolyar, and P.Whiting, *“Providing quality of service over a shared wireless link,”* IEEE Communications Magazine, pp. 150-154, February 2001.
- [7] Jong-Hun Rhee, Jack M. Holtzman, and Dong-Ku Kim, *“Performance analysis of the adaptive EXP/PF channel scheduler in an AMC/TDM system”*, IEEE Communications Letters, vol.8, No. 8, Aug. 2004.
- [8] Seungwan Ryu, Byunghan Ryu, Hyunhwa Seo, and Mooyong Shin, *“Urgency and efficiency based packet scheduling algorithm for OFDMA wireless system”*, ICC 2005, vol.4, pp.2779-2785, May, 2005.

- [9] Seungwan Ryu, Byung-Han Ryu, Hyunhwa Seo, Muyong Shin, and SeiKwon Park, "Wireless Packet Scheduling Algorithm for OFDMA System based on Time-utility and Channel State", ETRI Journal, vol.27, no.6, Dec. 2005, pp.777-787.
- [10] Haipeng Lei, C. Fan, X. Zhnag, D. Yang, "QoS Aware Packet Scheduling Algorithm for OFDMA Systems", IEEE 66th Vehicular Technology Conference, 2007 Fall.
- [11] Ángela Hernández, Antonio Valdovinos, Femando Casadevall, "Performance Analysis of Packet Scheduling Strategies for Multimedia Traffic in WCDMA", Vehicular Technology Conference, 2002. VTC Spring 2002. IEEE 55th.
- [12] 3GPP TR 36.942, "Evolved Universal Terrestrial Radio Access (E-UTRA); Radio Frequency (RF) system scenarios"

ANEXO A: DESCRIPCIÓN DEL SIMULADOR

El simulador de sistema que aquí se describe pretende modelar la capa física y determinados aspectos de la capa MAC (Medium Access Control) con el fin de dar soporte para el análisis y desarrollo de mecanismos de control de interferencias y scheduling. Para la capa física, basada en tecnología OFDM, se han utilizado los parámetros físicos considerados en E-UTRA (Evolved UMTS Terrestrial Radio Access network), modelándose la cadena completa de transmisión de información: modulador, transmisor, canal, receptor y demodulador. Sobre estas clases de capa física se programarán los algoritmos y procesos a nivel MAC que se deseen simular.

El simulador ha sido programado en lenguaje C++ haciendo uso de la librería SPUC (Signal Processing Using C++). Esta librería es una potente herramienta para el procesamiento de señal en cualquier campo, ya que dispone de un gran abanico de funciones. De hecho, en nuestro caso sólo se va a utilizar un subconjunto de éstas, las relativas a operaciones con complejos, operaciones matriciales y generación de números aleatorios. Una de sus mayores ventajas es su tratamiento eficiente de la memoria en las operaciones con vectores y complejos. La definición de una clase vector genérica para cualquier tipo de datos con una gran cantidad de operaciones ya programadas permite manejar grandes bloques de datos de forma rápida y eficiente.

A.1: Estructura del simulador

El simulador consta, en su versión actual, de un total de 14 clases programadas en lenguaje C++ y un *namespace* con funciones matemáticas de propósito general. La Figura A.1 muestra un esquema general de las relaciones existentes entre las distintas clases del simulador, donde las clases Fraccion, Histograma, HistogramaDiscreto y el *namespace* Matemat son implementadas por varias clases.

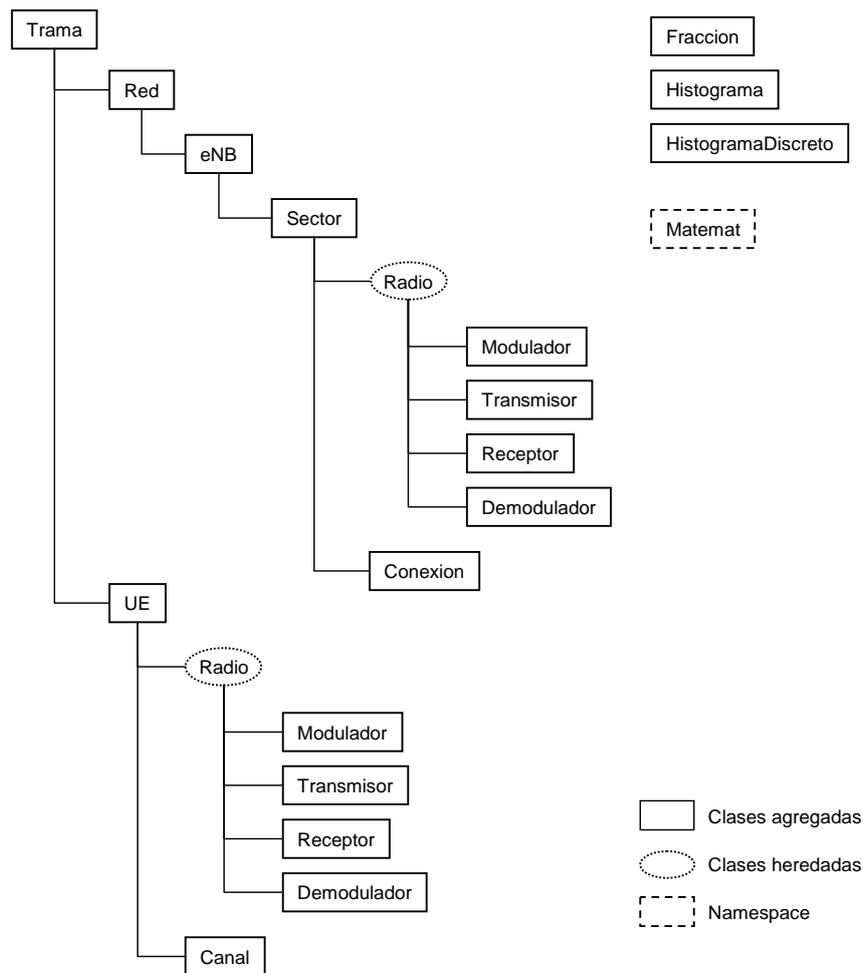


Figura A.1: Estructura de clases del simulador

A continuación se incluye una breve explicación de cada una de las clases del sistema, de modo que resulte más sencillo entender la organización del simulador:

- **Trama:** la clase Trama es la clase principal del simulador. Supone una referencia temporal común para todos los objetos del sistema, generando el entramado de la comunicación y configurando adecuadamente la transmisión de cada símbolo OFDM. Esta clase configura la simulación a realizar de acuerdo con los parámetros indicados por el fichero de parámetros y recopila las estadísticas al finalizar la simulación. La clase Trama crea un objeto Red, que aglutina todas las estaciones base (eNB) de la red, y un vector con todos los móviles (UE) presentes en el sistema, de modo que a través del objeto Trama global, las estaciones base tienen acceso a los móviles y viceversa.
- **Red:** la clase Red se encarga de definir el despliegue de eNBs para realizar la simulación, indicando a cada una de ellas sus coordenadas espaciales así como la lista de sus eNBs vecinas. Para evitar los efectos de borde en los eNBs exteriores, el despliegue se ha simulado con la técnica del wrap-around, consistente en emular que celdas opuestas en los extremos del escenario son

celdas vecinas y por tanto generan interferencia mutua (Figura A.2). De este modo se da continuidad a la interferencia a lo largo del escenario y se evita que los resultados queden falseados por una simulación en condiciones irreales.

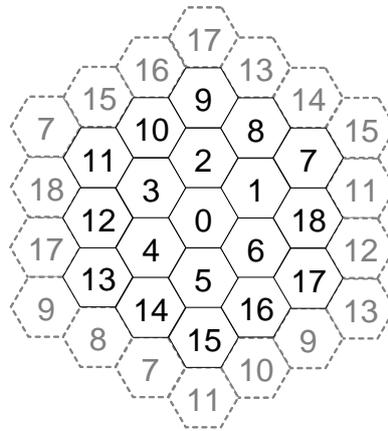


Figura A.2: Ejemplo de despliegue con wrap-around

- **eNB:** la clase eNB aglutina todas las funciones relativas a la estación base en su conjunto, tales como los algoritmos de scheduling a nivel de celda y la transmisión de señales en el DL. Un vector contiene los punteros a los objetos Sector que conforman la celda.
- **Sector:** la clase Sector puede verse como la unidad mínima de gestión de recursos radio en la red, manteniendo una lista de los UEs registrados en él y de sus conexiones activas, desarrollando las decisiones de scheduling cuando éstas se llevan a cabo de forma independiente para cada sector y generando su señal OFDM a transmitir.
- **Radio:** ésta es una clase que agrupa las clases que implementan la cadena de transmisión: Modulador, Transmisor, Receptor y Demodulador. Acepta varios modelos para la cadena de transmisión, lo que permite configurar los parámetros de la misma.
- **Modulador:** esta clase implementa las funciones necesarias para modular una secuencia de bits en símbolos complejos QAM o QPSK. Dependiendo de la configuración de la simulación, la etapa de modulación puede incluir a su vez un ensanchado de los datos con códigos CDMA, una transmisión con diversidad frecuencial (replicación de los símbolos QAM en varias portadoras OFDM) o la aplicación de una secuencia de scrambling.
- **Transmisor:** el objetivo principal de esta clase es la transformación de una secuencia de símbolos complejos de datos en un símbolo OFDM para su posterior transmisión.
- **Receptor:** además de la obtención de los símbolos complejos de datos a partir de la señal temporal recibida, la clase Receptor realiza la ecualización Zero-

forcing de la señal recibida de acuerdo a la respuesta frecuencial estimada del canal.

- **Demodulador:** la clase Demodulador toma la secuencia de símbolos ecualizados que proporciona Receptor y aplica la demodulación (y en su caso, el desensanchado) para obtener la secuencia de bits recibida.
- **Conexión:** la clase Conexión modela la existencia de una conexión lógica activa entre el eNB y el UE. Esta clase genera paquetes de datos de acuerdo con el modelo de tráfico elegido, trocea dichos paquetes en unidades mínimas de transmisión (TUs) y las almacena en los buffers de salida. Cuando se produce una asignación de recursos para una determinada conexión, los TUs a transmitir se almacenan en un vector y, una vez que son recibidos por el UE, se eliminan del buffer de salida si se recibieron correctamente o se mueven al buffer de ARQ para una posterior retransmisión si se produjeron errores.
- **UE:** la clase UE implementa las funciones relativas al terminal móvil, como son el modelo de desplazamiento, la recepción de señales y la medida de calidad del enlace. Cada objeto UE contiene a su vez un objeto Canal, que permite seguir la evolución de los distintos canales relativos al UE e introducir las pérdidas correspondientes a la señal transmitida.
- **Canal:** el objeto Canal mantiene un total de 7 canales en paralelo para cada UE (un canal al eNB local y 6 más para los eNBs interferentes del primer anillo, como se muestra en la Figura A.3), modelando el comportamiento de los desvanecimientos y del ruido térmico.

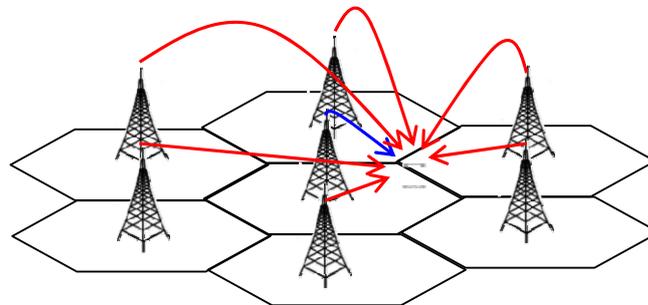


Figura A.3: Representación de los 7 canales de cada UE

- **Fracción:** la clase Fracción se utiliza para almacenar estadísticas de tipo cociente, como por ejemplo el BER o la tasa cursada. El numerador y el denominador de la fracción se guardan por separado, con lo que se pueden añadir continuamente valores en ambos términos de forma sencilla.
- **Histograma:** la clase Histograma permite almacenar directamente la estadística de una variable sin necesidad de guardar sus valores concretos a lo largo del tiempo, reduciéndose así el tamaño de los ficheros de resultados generados.

- **HistogramaDiscreto:** se trata de un histograma pero cuyos posibles valores de entrada están definidos a priori, y son iguales a los distintos niveles considerados en el histograma.

Finalmente, el *namespace* *Matemat* proporciona una serie de funciones matemáticas generales que son utilizadas por otras clases del sistema, tales como la FFT, generadores de variables aleatorias con distribuciones concretas, funciones de ordenación de vectores, etc

ANEXO B: RESULTADOS COMPLETOS AL APLICAR BLOQUEO INICIAL

En este anexo se encuentran los resultados de la evaluación de los distintos algoritmos en función de los valores de sus parámetros y sus posteriores comparaciones entre ellos para todos los escenarios planteados en la memoria: tipos de ráfagas A o B, con tráfico uniforme o mezclando varios tipos de servicios. Todos estos escenarios se probaron con los dos tipos de bloqueo explicados en el apartado 5.3 de la memoria, y cada uno de ellos con los distintos valores seleccionados para el parámetro δ : 0.8, 0.5 y 0.2.

En la memoria se destacan los resultados más relevantes de los obtenidos, por lo que aquí mostraremos todos los que se obtuvieron y comentaremos brevemente sus comportamientos, explicando por qué sus variaciones respecto al caso sin bloqueo son más o menos relevantes.

B.1: Bloqueo por retardo

B.1.1: Ráfagas A. Tráfico uniforme.

En primer lugar se estudia un escenario correspondiente al caso en el que todos los usuarios de la red generan el mismo tipo de tráfico, con ráfagas que admiten un retardo máximo $d_{max}=50ms$ y tiempos de ON/OFF de 30 y 90 milisegundos respectivamente.

En primer lugar se evalúa la mejora o el empeoramiento de cada uno de los algoritmos al aplicar bloqueo con los distintos valores de δ . En la figura B.1 se observa cómo los cambios más significativos al aplicar bloqueo los sufren los algoritmos 1-UEPS, Retardo y PF, siendo estos dos últimos los incluidos en la memoria. El 1-UEPS no se incluyó puesto que, aun siendo relevante su pérdida de capacidad con $\delta =0.8$, en los otros casos prácticamente no cambia sus prestaciones. El resto de algoritmos sufre menos cambios al aplicar el bloqueo y con distinta suerte, dependiendo del valor de δ .

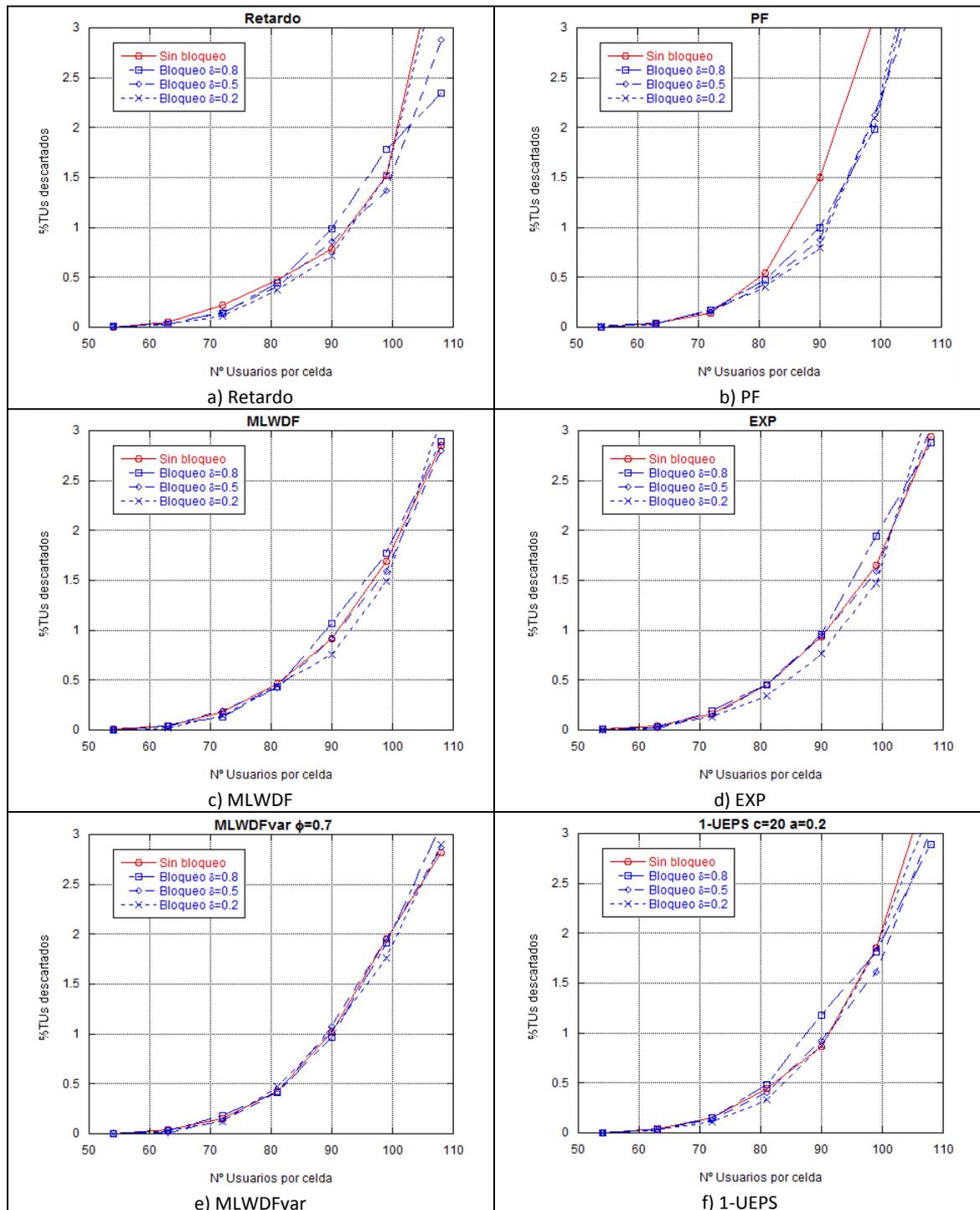


Figura B.1: Comparación entre algoritmos con/sin bloqueo con tráfico uniforme (rA).

Ahora compararemos los resultados que se obtienen con los distintos algoritmos cambiando los parámetros que controlan las métricas y viendo cómo responden a los distintos tipos de bloqueo. En la figura B.2 vemos lo que sucede con el algoritmo MLWDFvar. Tal y como sucedía cuando no aplicábamos bloqueo, con este tipo de ráfagas y tráfico uniforme, la influencia del parámetro ϕ en los resultados es mínima.

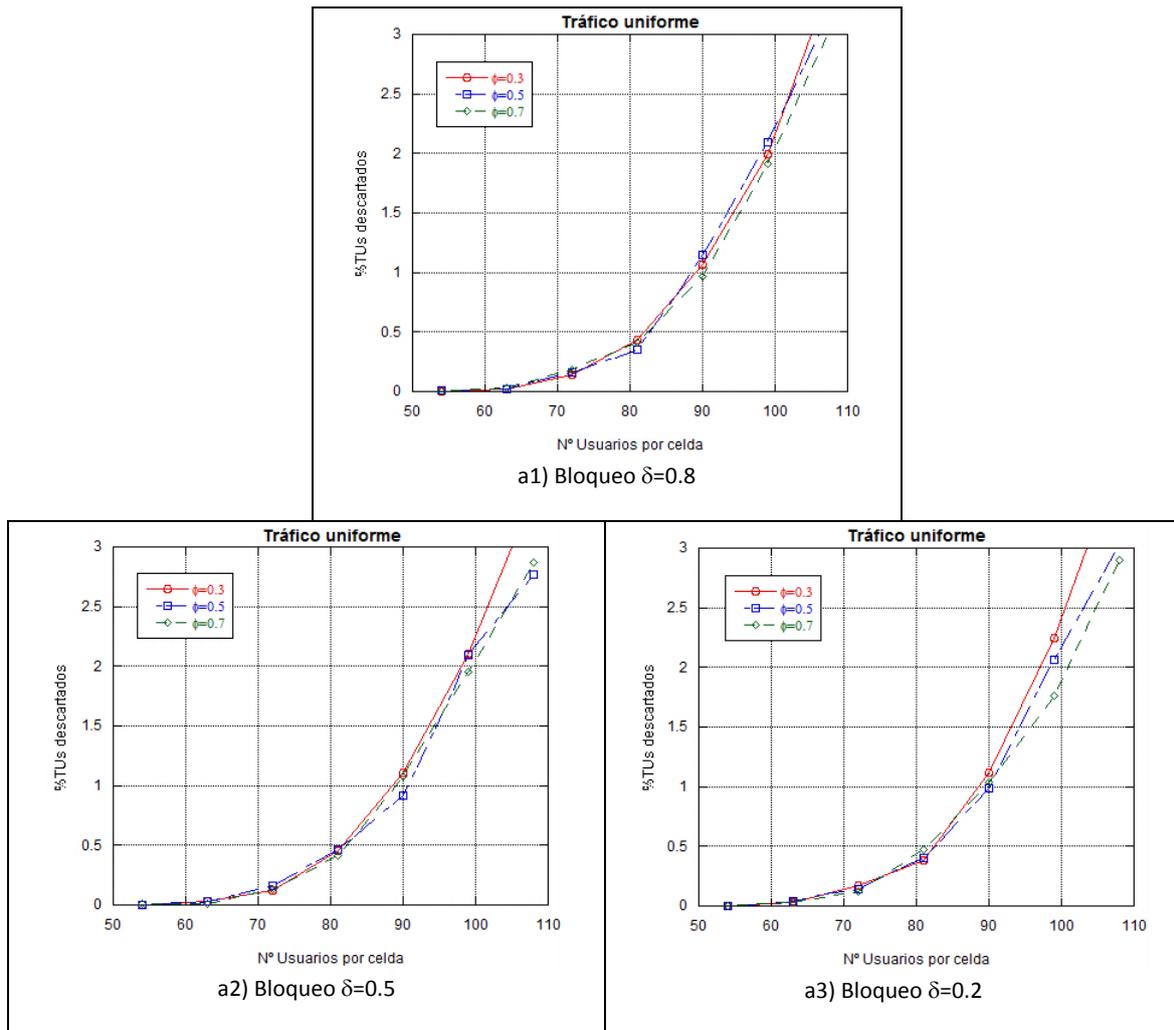


Figura B.2: Algoritmo MLWDFvar con tráfico uniforme (rA).

En la figura B.3 vemos de qué forma afecta el bloqueo al algoritmo 1-UEPS. Observamos que dependiendo de los valores de los parámetros afecta de una forma u otra, ya que con $c=20$, $\alpha=0.2$ se obtenían los mejores resultados si no aplicábamos bloqueo, y para $\delta=0.8$ (fig. B.3a₁) empeora en situaciones de baja carga hasta ser el que menos capacidad soportaría. Para los otros valores de δ la situación prácticamente no variaría.

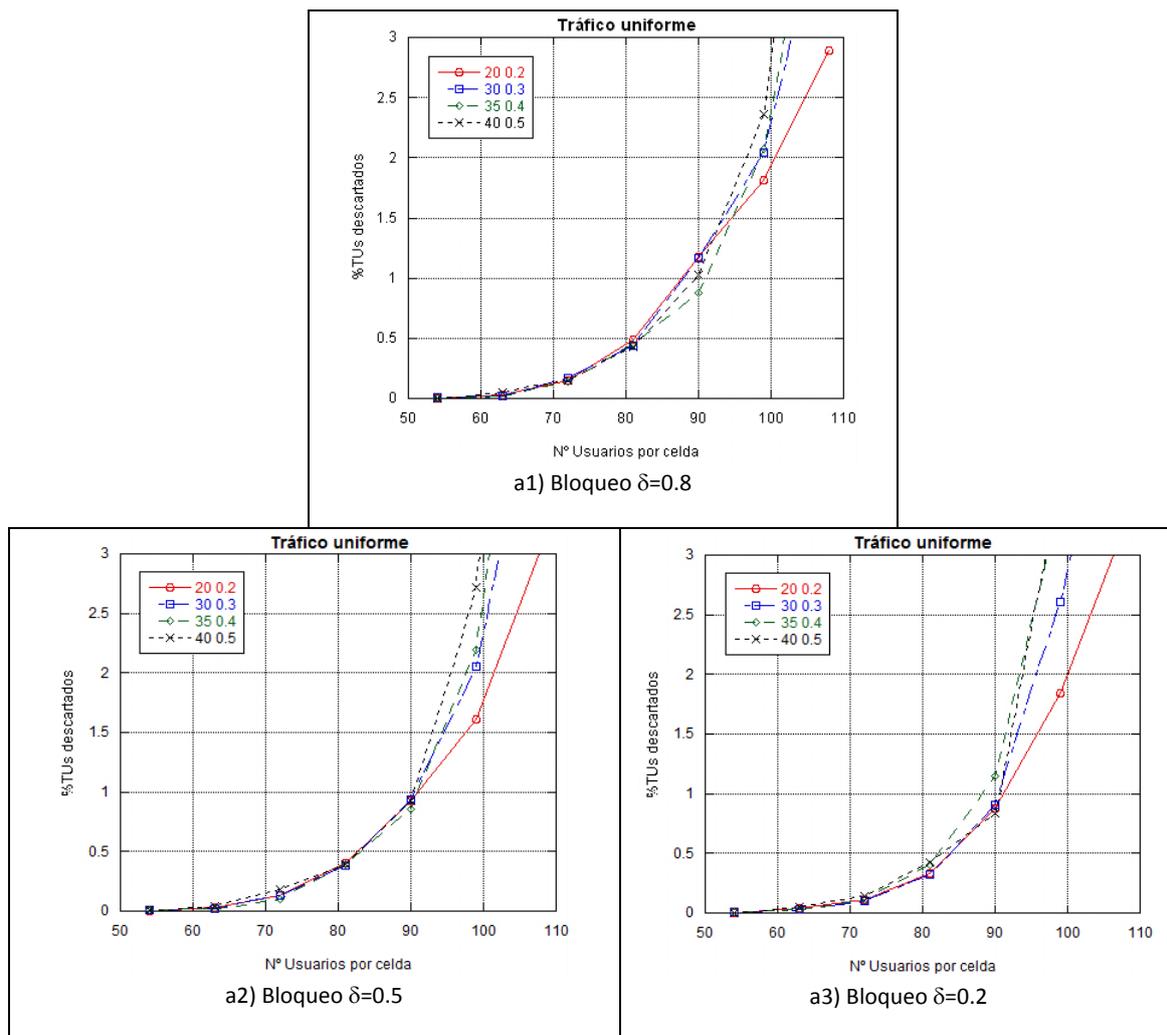


Figura B.3: Algoritmo 1-UEPS con tráfico uniforme (rA).

Una vez evaluada la afeción del bloqueo a los distintos parámetros, compararemos los algoritmos MLWDFvar y 1-UEPS con los Retardo, PF, MLWDF y EXP, tal y como hacíamos en la memoria para el caso sin bloqueo. Se observa en la figura B.4 que al aplicar los algoritmos con cierto bloqueo a los usuarios con peores condiciones de canal se obtienen unos resultados similares para casi todos los algoritmos, que tienen un cruce por el 1% de tasa de descarte similar, en este caso incluido el PF, que tiene una mayor mejora que el resto al aplicar bloqueo (anteriormente funcionaba peor que los demás algoritmos). La excepción la encontramos en el caso de $\delta=0.2$ (fig. B.4a₃), donde el MLWDFvar parece funcionar algo peor. Escogimos para la memoria la figura B.4a₂ porque en ella se aprecian más fácilmente las mejoras conseguidas con el bloqueo para el Retardo y el PF.

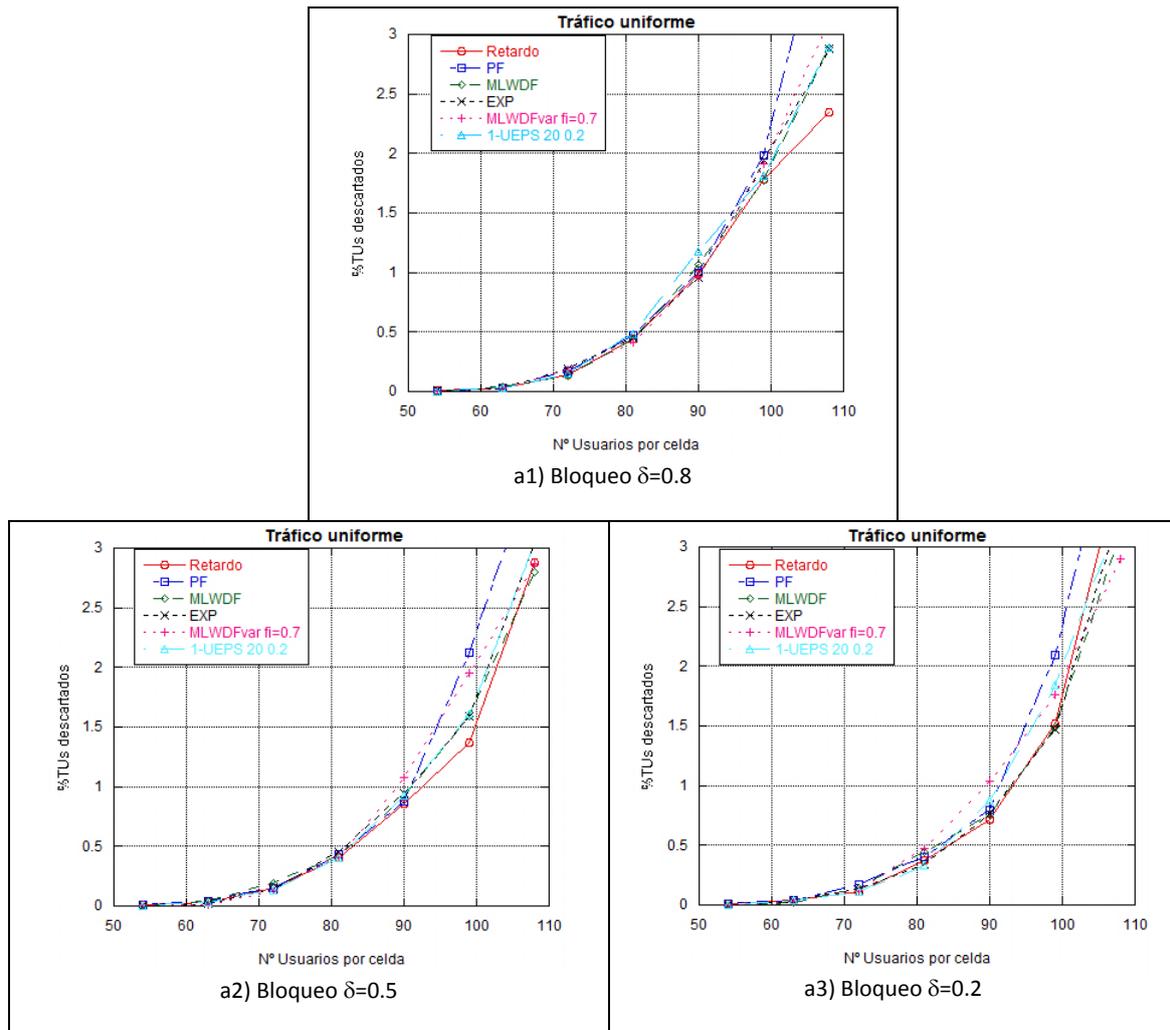


Figura B.4: Comparación bloqueos para los distintos algoritmos (rA). Tráfico uniforme.

B.1.2: Ráfagas A. Escenario multiservicio.

Evaluaremos ahora un escenario que mantiene las ráfagas del anterior pero cambia el tráfico en el sistema, teniendo ahora de los dos tipos ya conocidos: $d_{max1}=50ms$ y $d_{max2}=250ms$.

En la figura B.5 se observa cómo, en la mayoría de los casos, el bloqueo afecta negativamente a este escenario, aumentando el porcentaje de descartes y disminuyendo por tanto la capacidad de la celda. Sin embargo, si nos fijamos concretamente en el valor $\delta=0.2$, veremos que en la mayoría de los casos tiene un funcionamiento muy parejo al caso sin bloqueo excepto en dos casos: el PF, que empeora notablemente sus prestaciones, y el Retardo, que aunque mantiene la capacidad, consigue evitar la gran crecida del porcentaje de descargas cuando el número de usuarios es elevado y la carga del sistema alta.

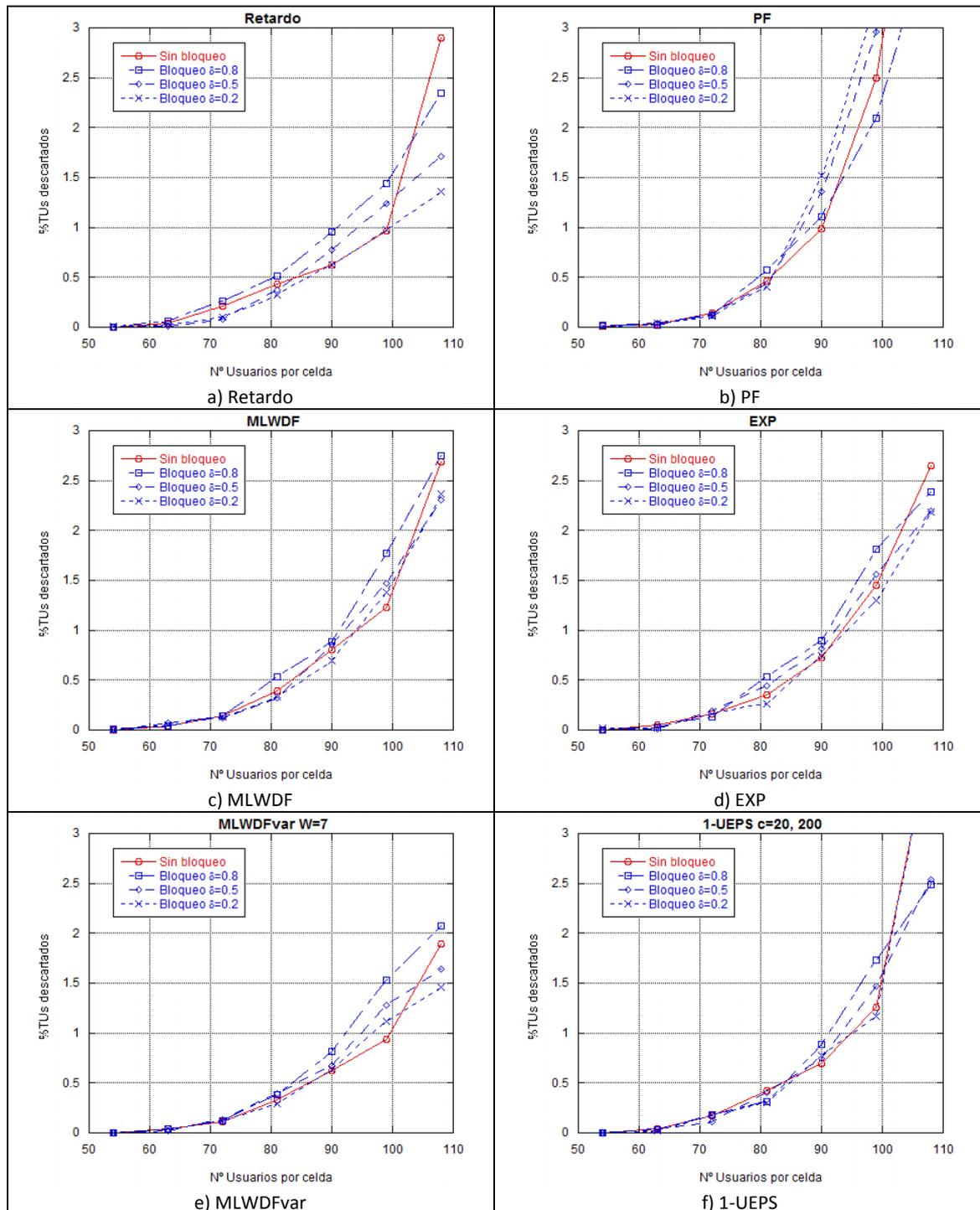


Figura B.5: Comparación entre algoritmos con/sin bloqueo en un escenario multiservicio (50% $d_{max}=50ms$ +50% $d_{max}=250ms$) (rA). Tráfico de cota 50ms

En la figura B.6 vemos que el efecto en el tráfico de 250ms es muy similar al del tráfico de 50ms; empeora en la mayoría de los casos, excepto con algunos valores de δ para los algoritmos MLWDFvar y 1-UEPS.

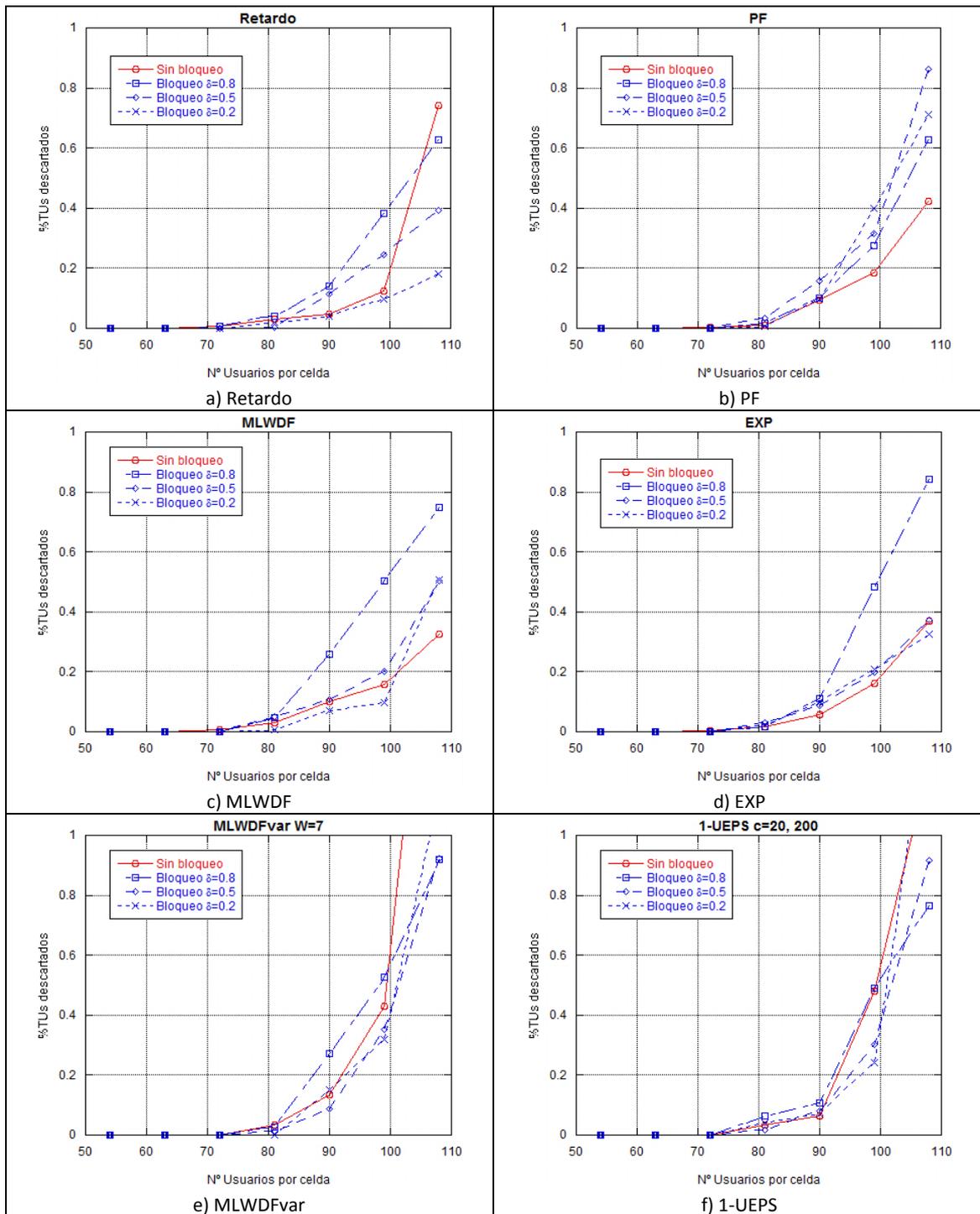


Figura B.6: Comparación entre algoritmos con/sin bloqueo en un escenario multiservicio (50% $d_{max}=50ms$ +50% $d_{max}=250ms$) (rA). Tráfico de cota 250ms

En la figura B.7 podemos ver como al aplicar bloqueo inicial las diferencias de resultados obtenidos al cambiar el parámetro W son muy pequeñas, reduciéndose las diferencias existentes respecto al caso sin bloqueo. Todos los casos siguen un dibujo muy parecido.

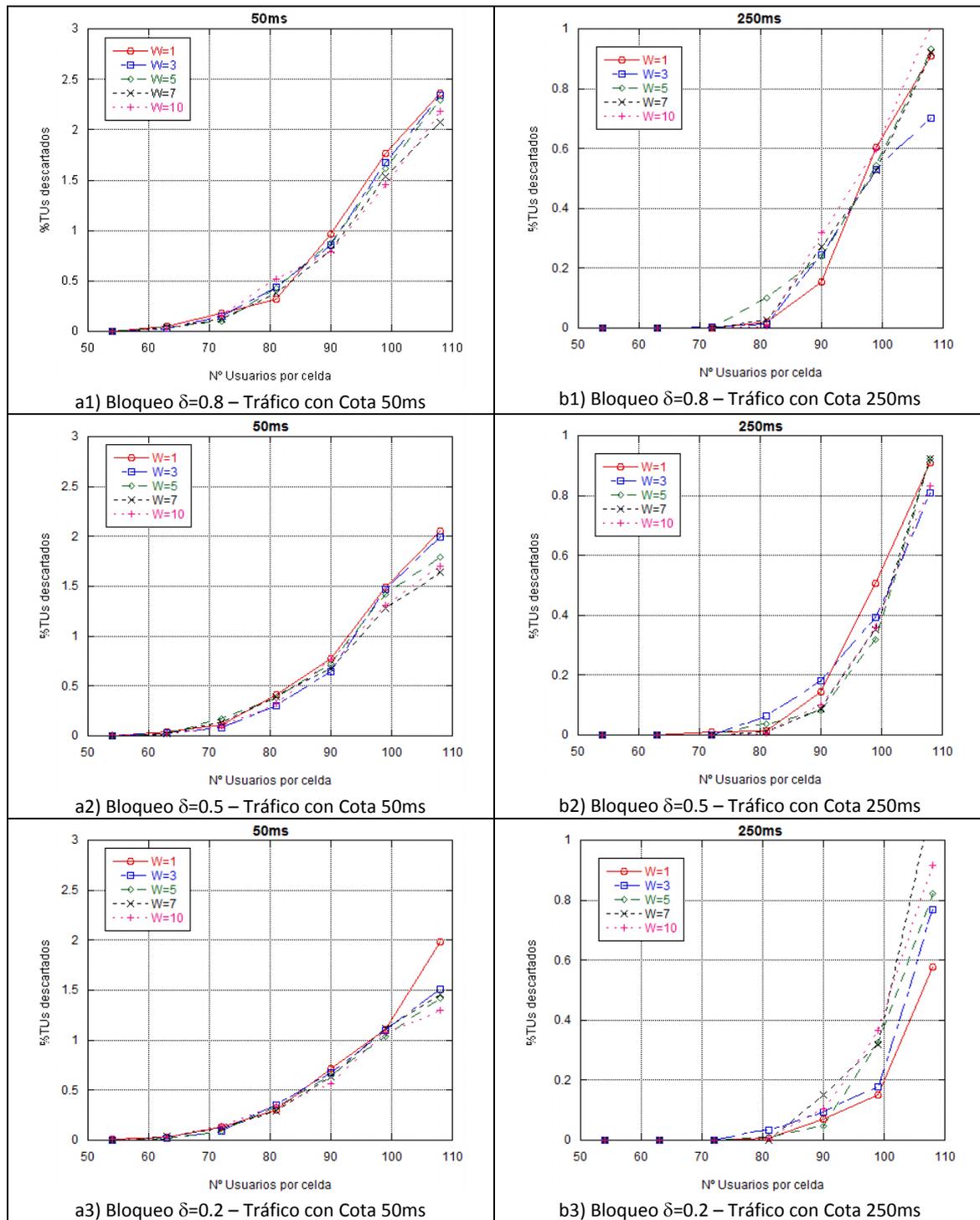


Figura B.7: Algoritmo MLWDFvar en el escenario multiservicio (rA).

Respecto al algoritmo 1-UEPS, que podemos ver en la figura B.8, hemos de decir que ocurre algo parecido a lo del MLWDFvar, y las diferencias entre los resultados de los distintos parámetros se han visto reducidas.

Scheduling y provisión de QoS en redes móviles 4G OFDMA

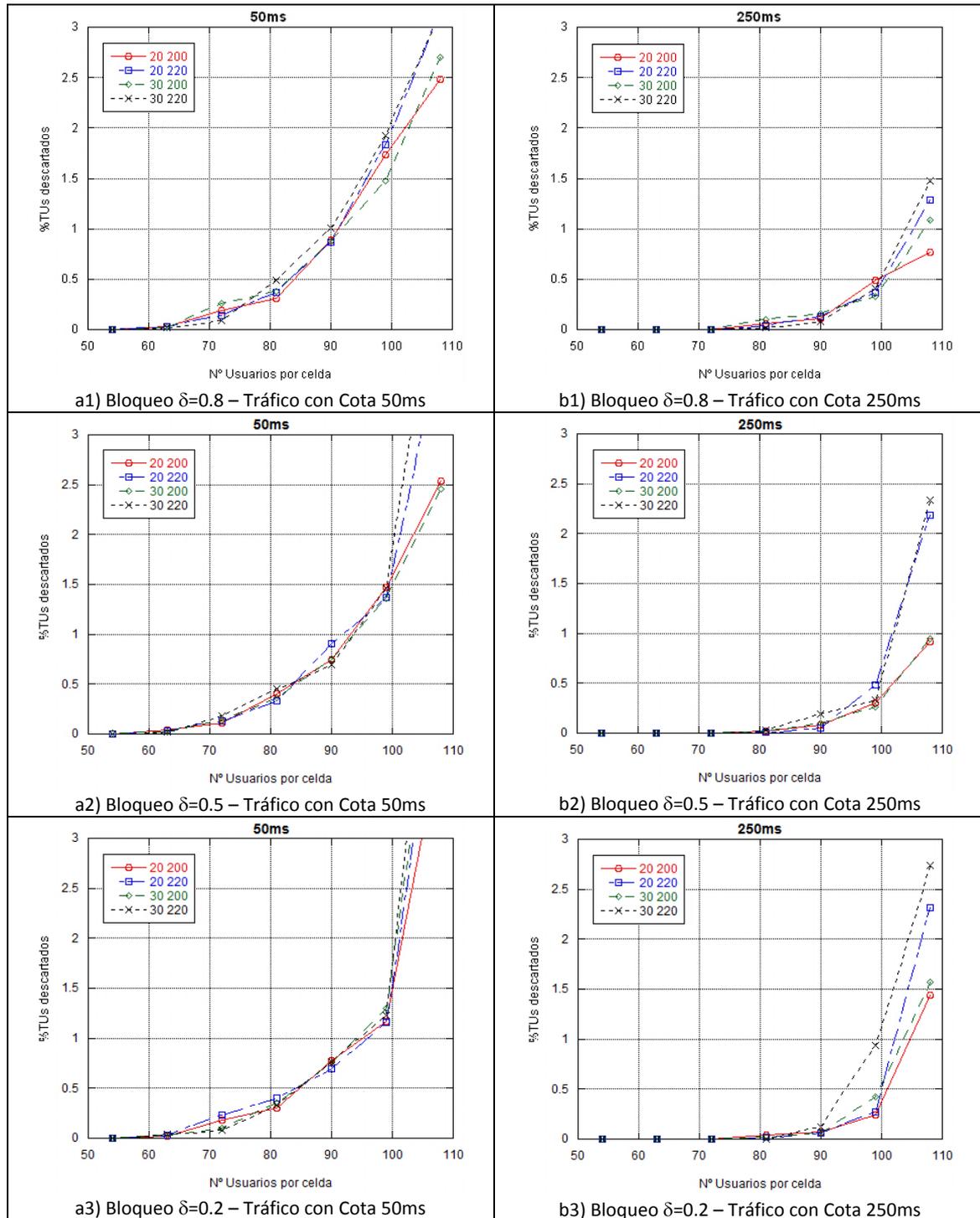


Figura B.8: Algoritmo 1-UEPS en el escenario multiservicio (rA).

En la figura B.9 tenemos las comparaciones de los resultados de todos los algoritmos evaluados, y cabe reseñar que, aunque no se consigue una mejora de la capacidad del sistema en general, sí que es bastante positiva la mejora del algoritmo Retardo en el caso de $\delta = 0.2$, ya que con la aplicación de este bloqueo conseguimos que este algoritmo sea el que mejor funcione para todos los tráficos en este escenario.

Scheduling y provisión de QoS en redes móviles 4G OFDMA

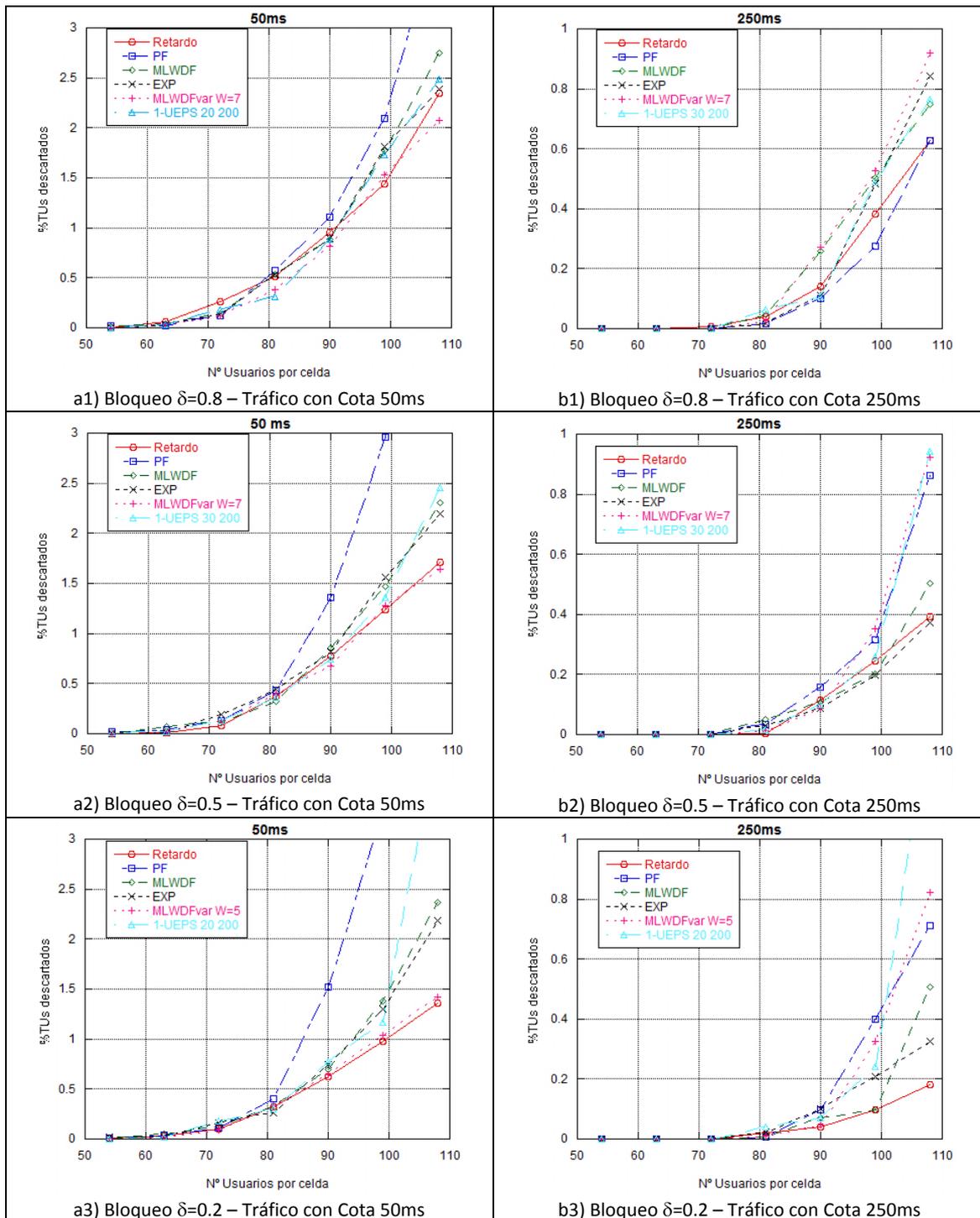


Figura B.9: Comparación bloqueos para los distintos algoritmos (rA). Escenario multiservicio.

B.1.3: Ráfagas B. Tráfico uniforme.

Se estudia en este apartado un escenario correspondiente al caso en el que todos los usuarios de la red generan el mismo tipo de tráfico, con ráfagas que admiten un retardo máximo $d_{max}=50ms$ y ráfagas de tipo B, con periodos de actividad y silencio de 300 y 900 milisegundos respectivamente.

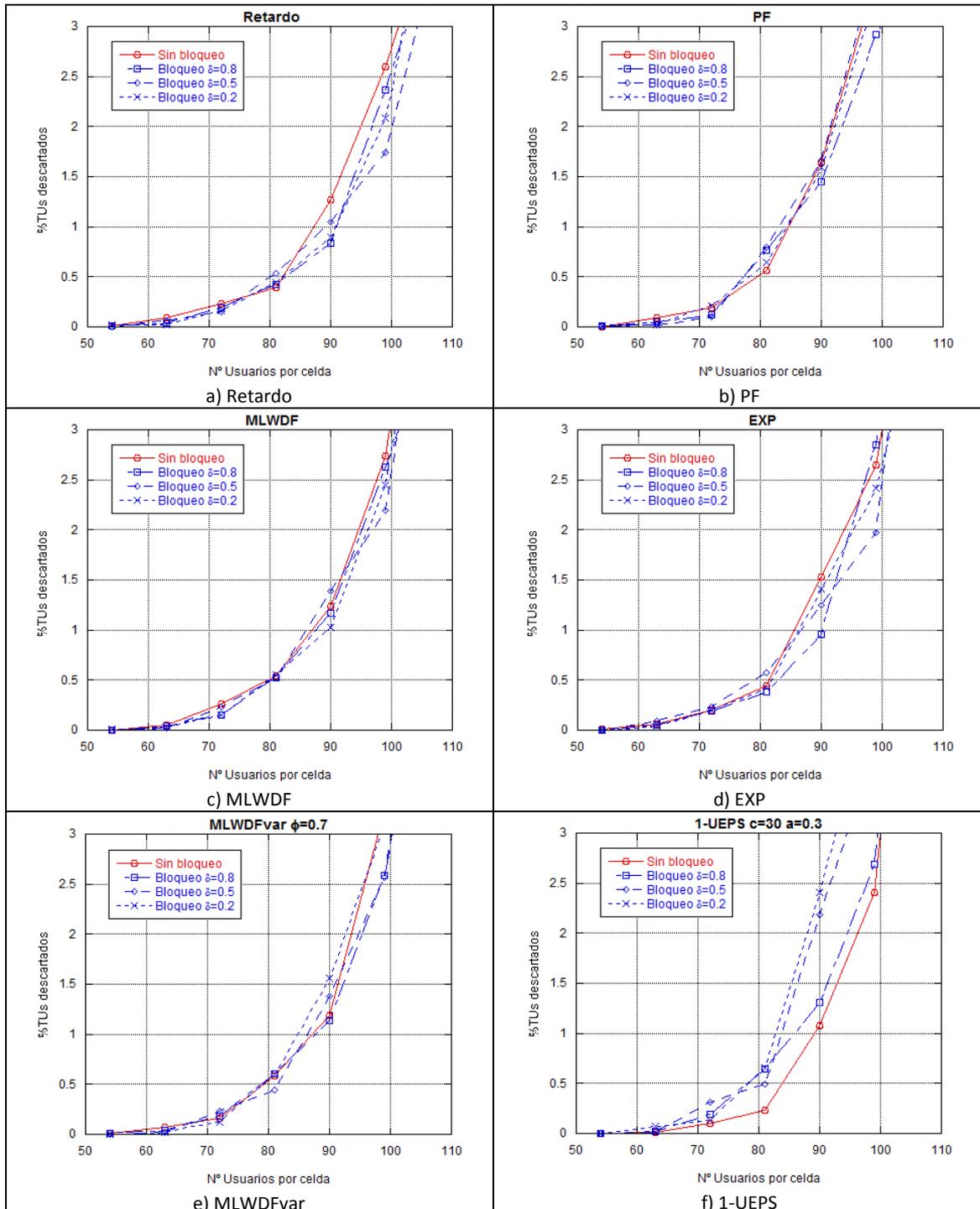


Figura B.10: Comparación entre algoritmos con/sin bloqueo con tráfico uniforme (rB).

Scheduling y provisión de QoS en redes móviles 4G OFDMA

En la figura B.10 se observa cómo los cambios más significativos al aplicar bloqueo los sufren los algoritmos 1-UEPS, Retardo y EXP. Las prestaciones del primero sufren un empeoramiento considerable, mientras que Retardo y EXP experimentan un aumento de la capacidad de la celda con la aplicación de bloqueo. También es reseñable que el algoritmo PF empeora ligeramente sus prestaciones en situaciones de baja carga del sistema.

Siguiendo el esquema utilizado con ráfagas A, en la figura B.11 vemos lo que sucede con el algoritmo MLWDFvar. En este caso vemos que la influencia del parámetro ϕ es mayor que en cuando usábamos el otro tipo de ráfagas, excepcionalmente en el caso de $\delta=0.8$, donde $\phi=0.7$ da unos resultados bastante mejores que el resto.

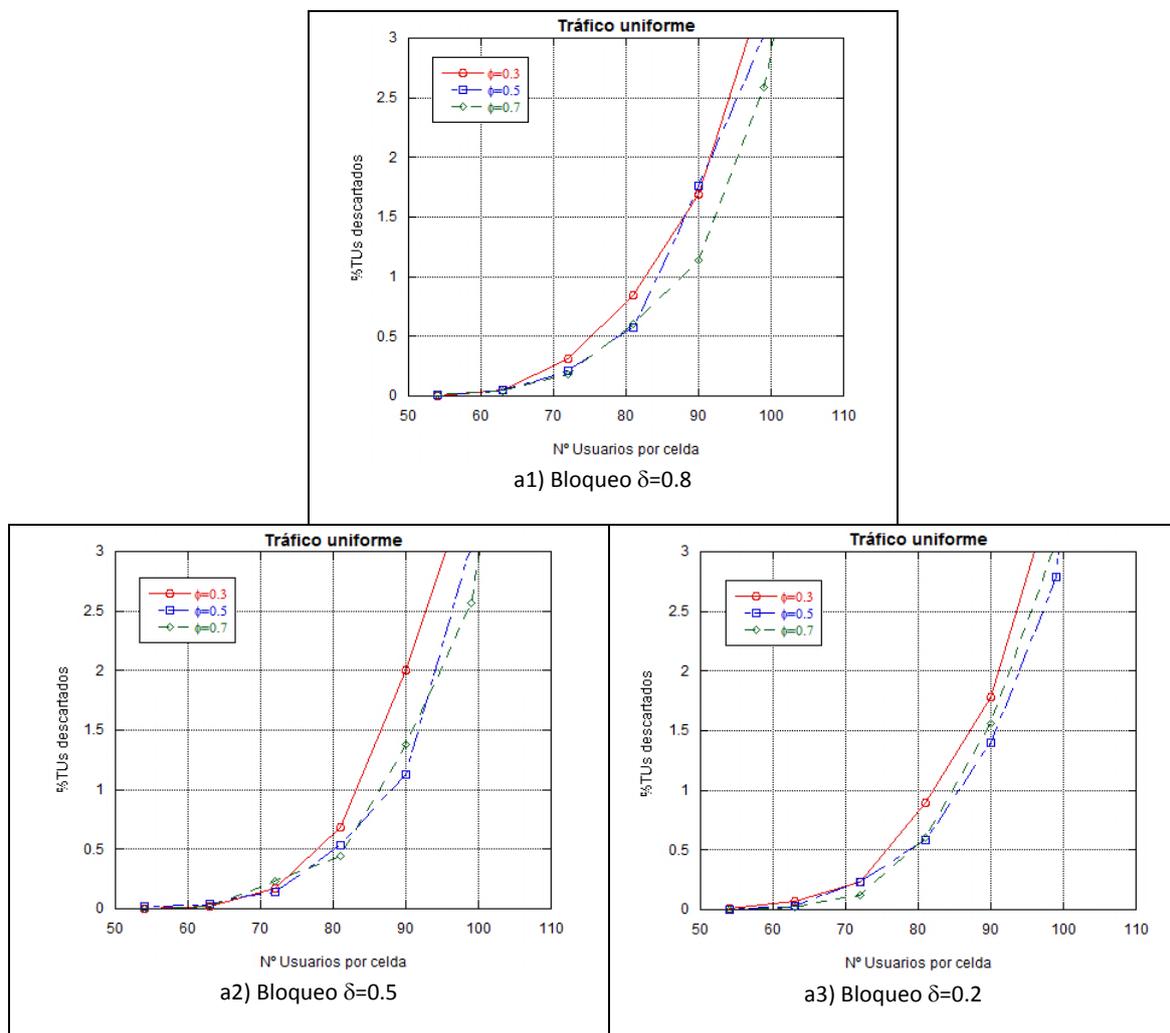


Figura B.11: Algoritmo MLWDFvar con tráfico uniforme (rB).

En la figura B.12 vemos el algoritmo 1-UEPS. En este caso, con $c=20$, $\alpha=0.2$ se obtienen en general los mejores resultados, especialmente para $\delta=0.2$ (fig. B.12a₃), por lo que tendríamos una situación que cambia respecto del caso sin bloqueo, donde conseguíamos las mejores prestaciones con $c=30$, $\alpha=0.3$.

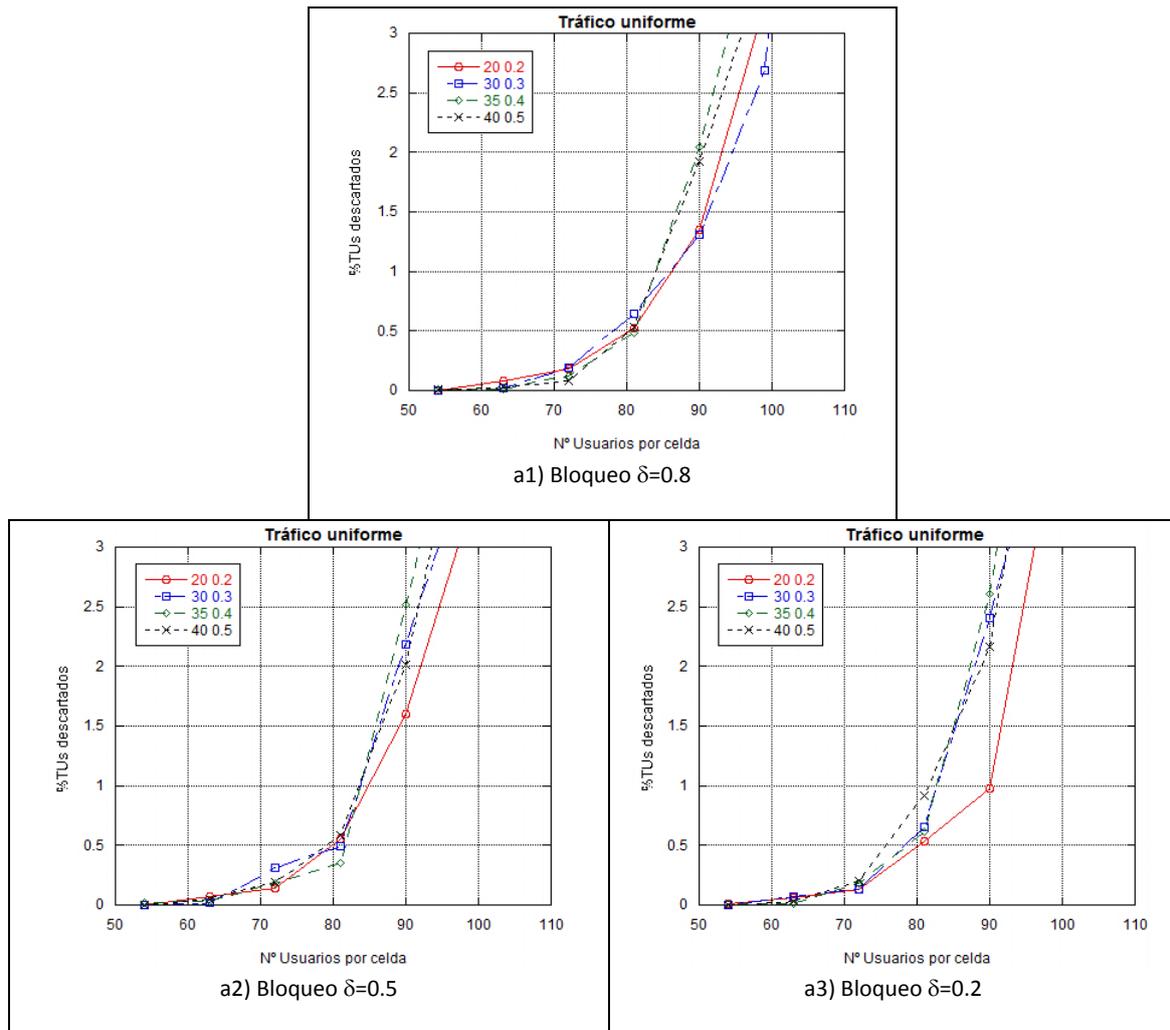


Figura B.12: Algoritmo 1-UEPS con tráfico uniforme (rB).

En la figura B.13 volvemos a comparar los resultados que se obtienen para los distintos algoritmos. Se observa que gracias a la mejora conseguida con algunos algoritmos, principalmente el Retardo, las prestaciones mejoran ligeramente respecto al caso sin bloqueo. El algoritmo 1-UEPS, al empeorar sus resultados, pasa de tener unas buenas prestaciones en el caso sin bloqueo a verse superado por la mayoría de los algoritmos.

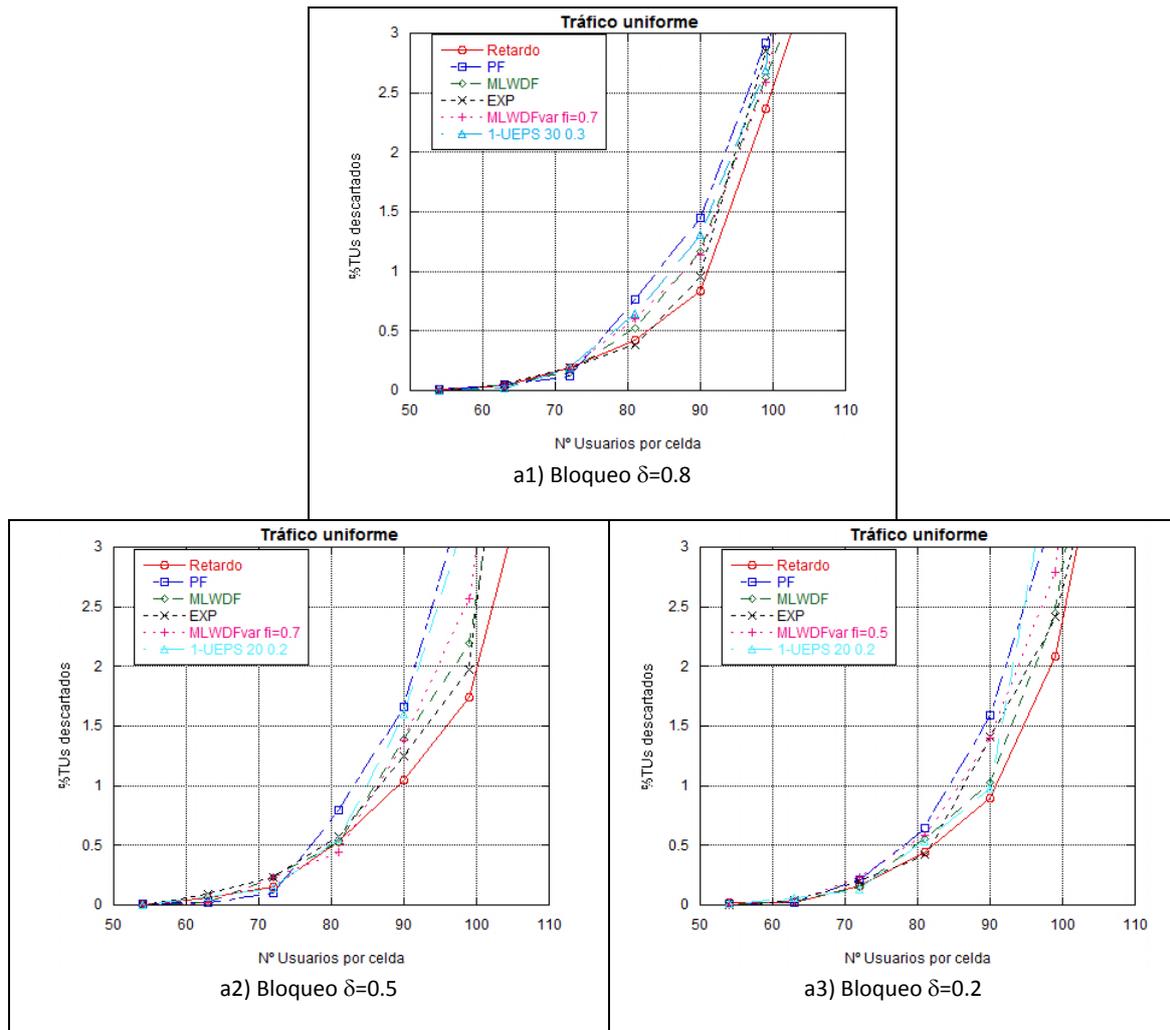


Figura B.13: Comparación bloqueos para los distintos algoritmos (rB). Tráfico uniforme.

B.1.4: Ráfagas B. Escenario multiservicio.

Evaluaremos ahora un escenario que mantiene las ráfagas del apartado B.1.3 pero cambia el tráfico en el sistema, teniendo ahora de los dos tipos ya conocidos: $d_{max1}=50ms$ y $d_{max2}=250ms$.

En la figura B.14 se observa que, en la mayoría de los casos, el bloqueo tiene un efecto positivo al aplicarse sobre los algoritmos en este escenario. El único caso que empeora sus resultados es el PF, por lo que se destaca en la memoria junto con los algoritmos Retardo y MLWDFvar, destacados por mejorar sus prestaciones cuando ya eran los que mejor funcionaban en el caso sin bloqueo.

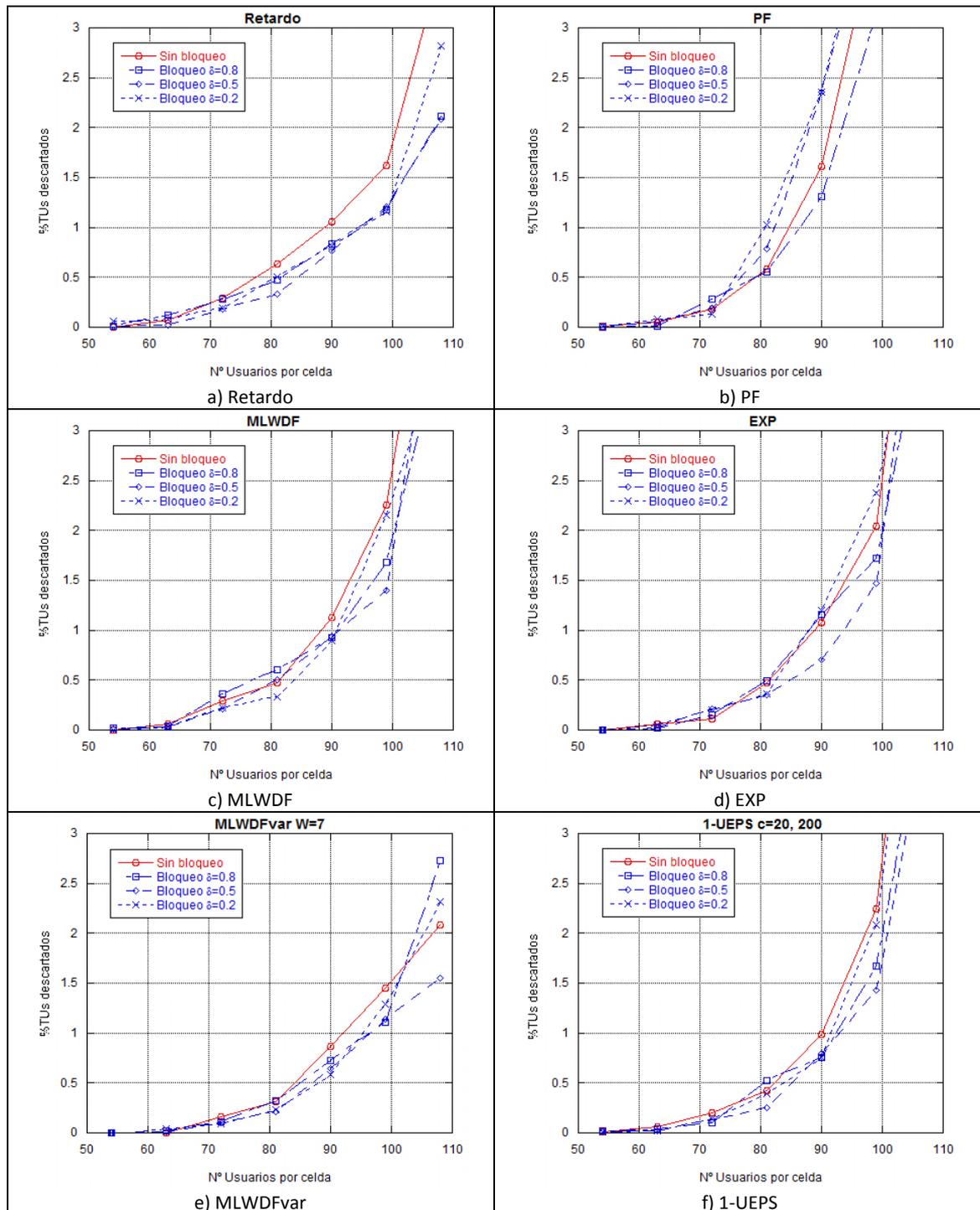


Figura B.14: Comparación entre algoritmos con/sin bloqueo en un escenario multiservicio ($50\% d_{max}=50ms + 50\% d_{max}=250ms$) (rB). Tráfico de cota 50ms

En la figura B.15 vemos que el efecto en el tráfico de 250ms es muy similar al del tráfico de 50ms; empeorando especialmente en el caso de MLWDFvar, que aun así mantiene el tráfico de 50ms como el más restrictivo en cuanto a número de usuarios admisibles manteniendo el descarte por debajo del 1% para ambos tipos de tráfico.

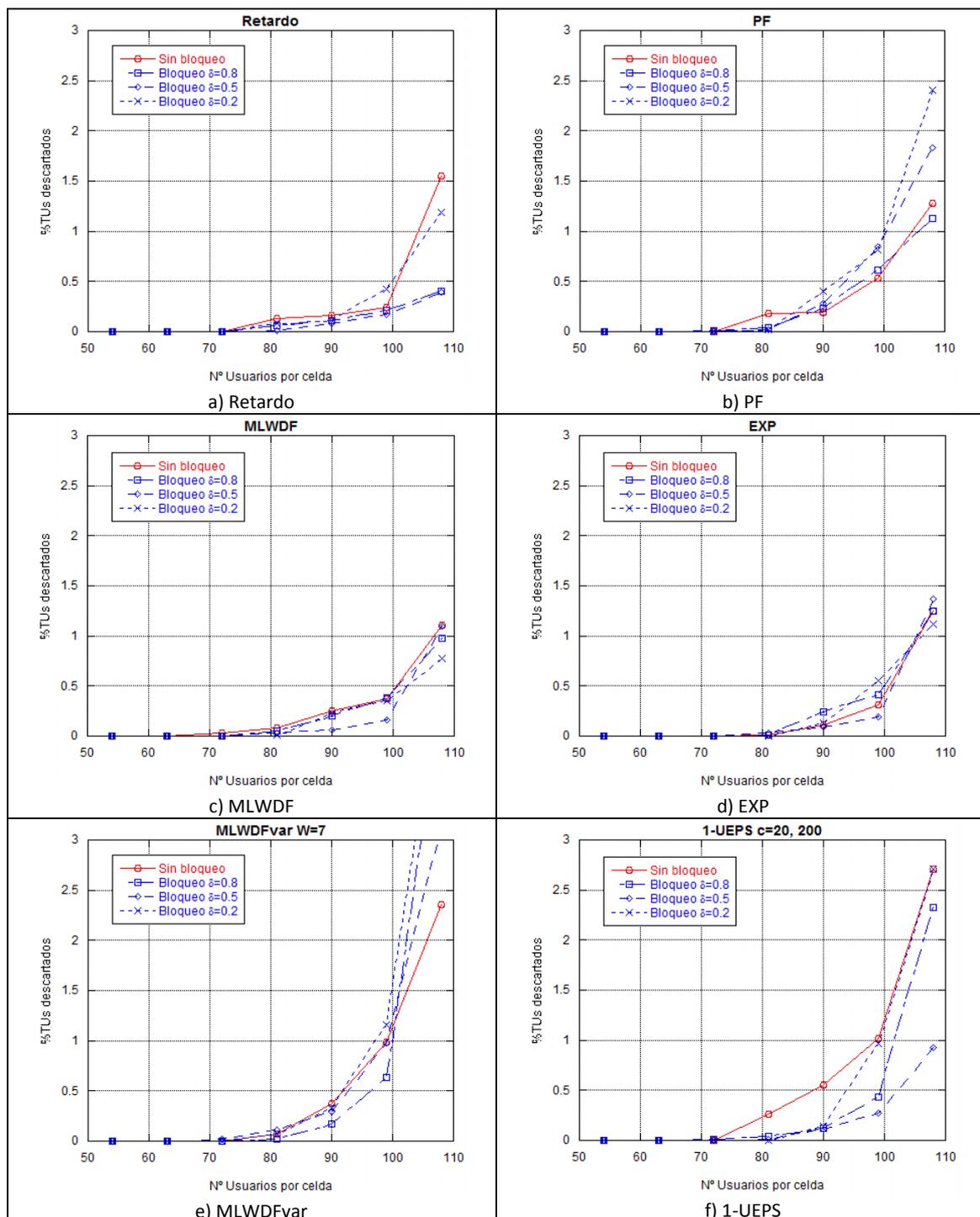


Figura B.15: Comparación entre algoritmos con/sin bloqueo en un escenario multiservicio (50% $d_{max}=50ms$ +50% $d_{max}=250ms$) (rB). Tráfico de cota 250ms

Con ráfagas B y aplicando bloqueo, existen diferencias entre los resultados obtenidos para MLWDFvar con los distintos pesos W, como puede verse en la figura B.16. Se observa en ella que para el tráfico de 250ms, gran parte de los casos sufren un muy elevado crecimiento del descarte cuando la carga del sistema aumenta, algo que deberíamos tratar de evitar en la medida de lo posible.

Scheduling y provisión de QoS en redes móviles 4G OFDMA

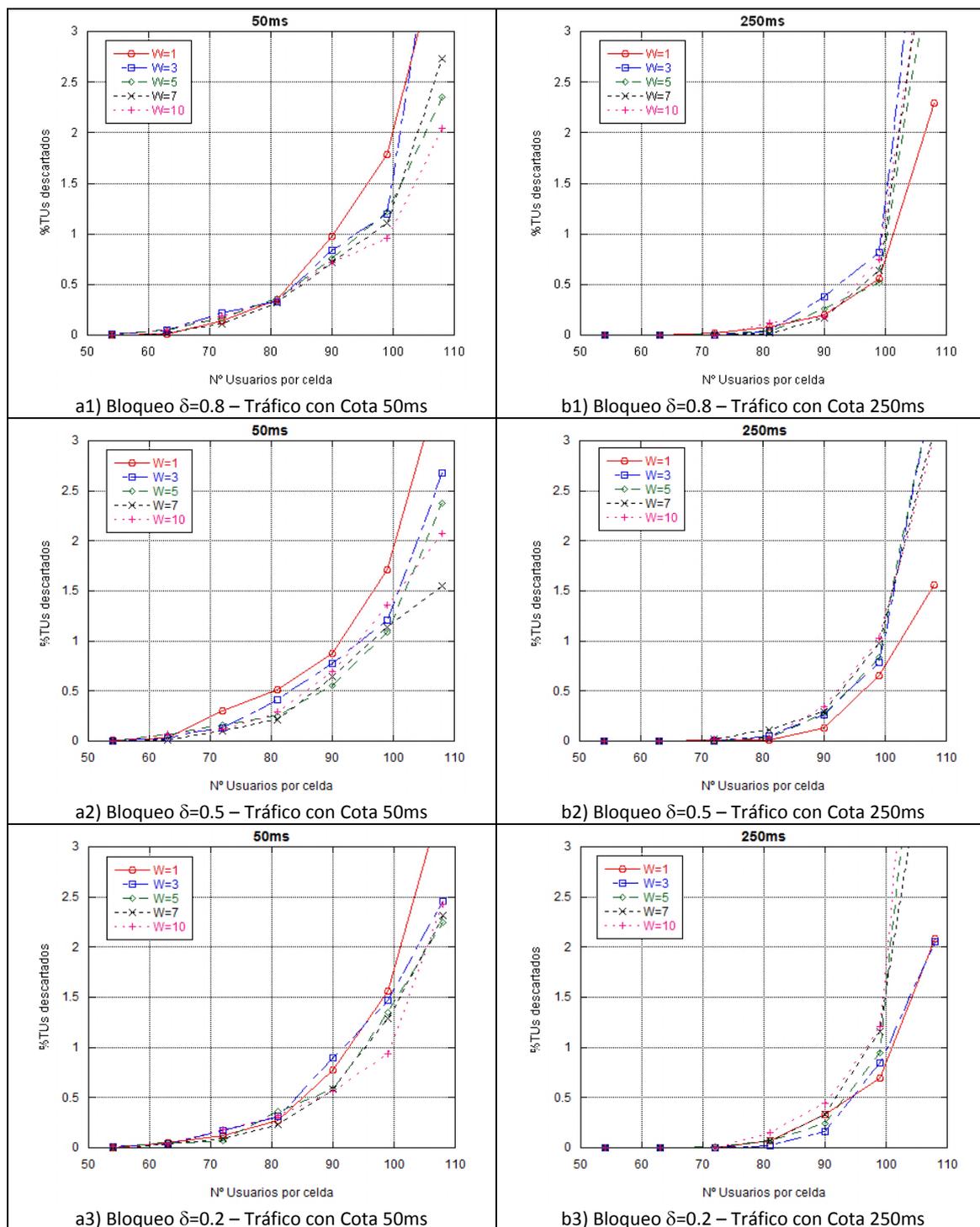


Figura B.16: Algoritmo MLWDFvar en el escenario multiservicio (rB).

En la figura B.17 puede verse lo que sucede para el algoritmo 1-UEPS, que mejora ligeramente sus prestaciones para el tráfico de 50ms. Para el de 250ms vemos que el porcentaje de descartes depende fuertemente de los parámetros utilizados, por lo que habría que seleccionarlos cuidadosamente.

Scheduling y provisión de QoS en redes móviles 4G OFDMA

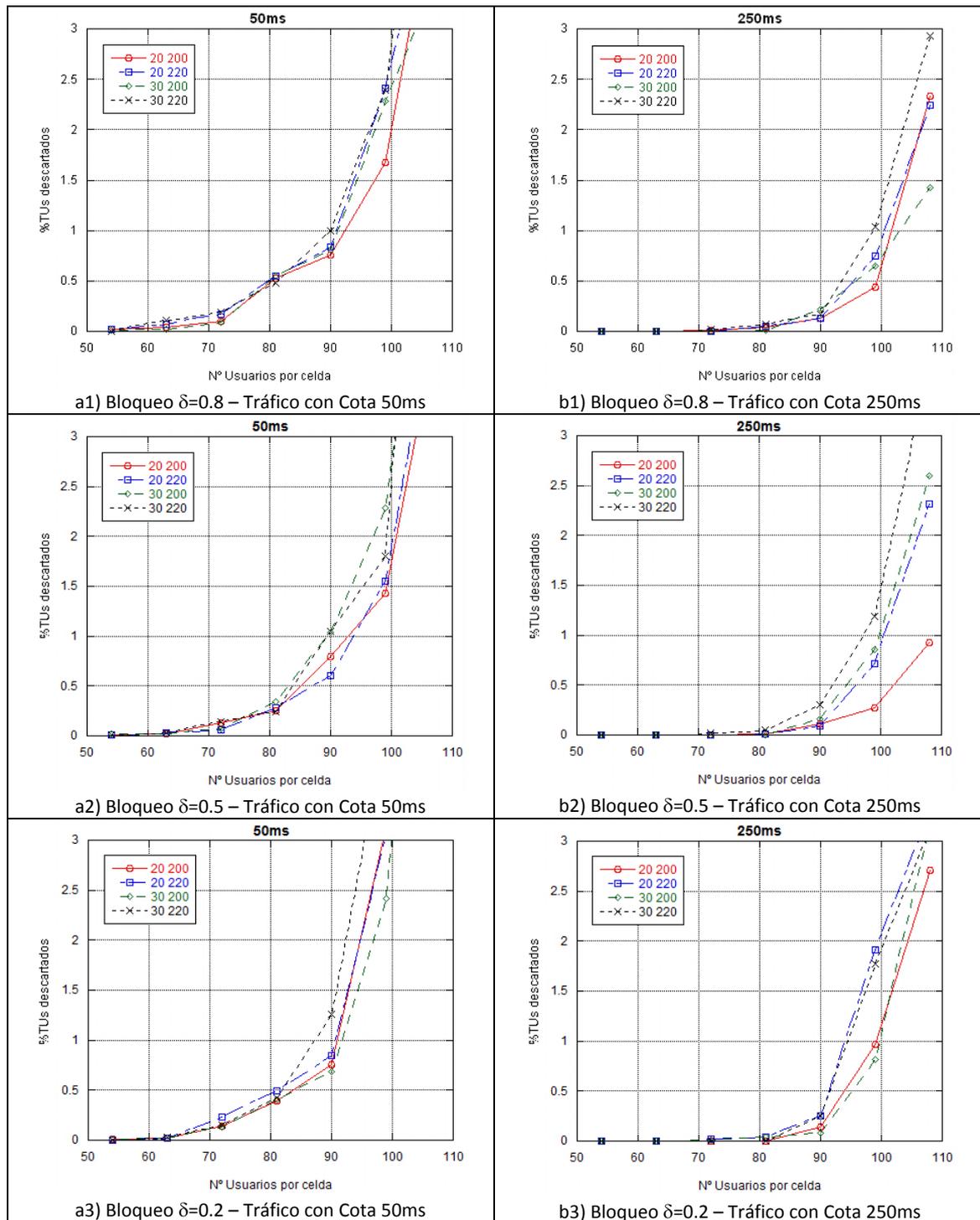


Figura B.17: Algoritmo 1-UEPS en el escenario multiservicio (rB).

Finalmente comparamos los resultados de todos los algoritmos evaluados en la figura B.18. En este caso sí que se obtiene una mejora de las prestaciones y un aumento de la capacidad gracias a la aplicación del bloqueo, ya que mejoran los dos algoritmos con mejores prestaciones, el MLWDFvar y el Retardo. En este caso sería especialmente importante aplicar bloqueo a los usuarios con peores condiciones de canal.

Scheduling y provisión de QoS en redes móviles 4G OFDMA

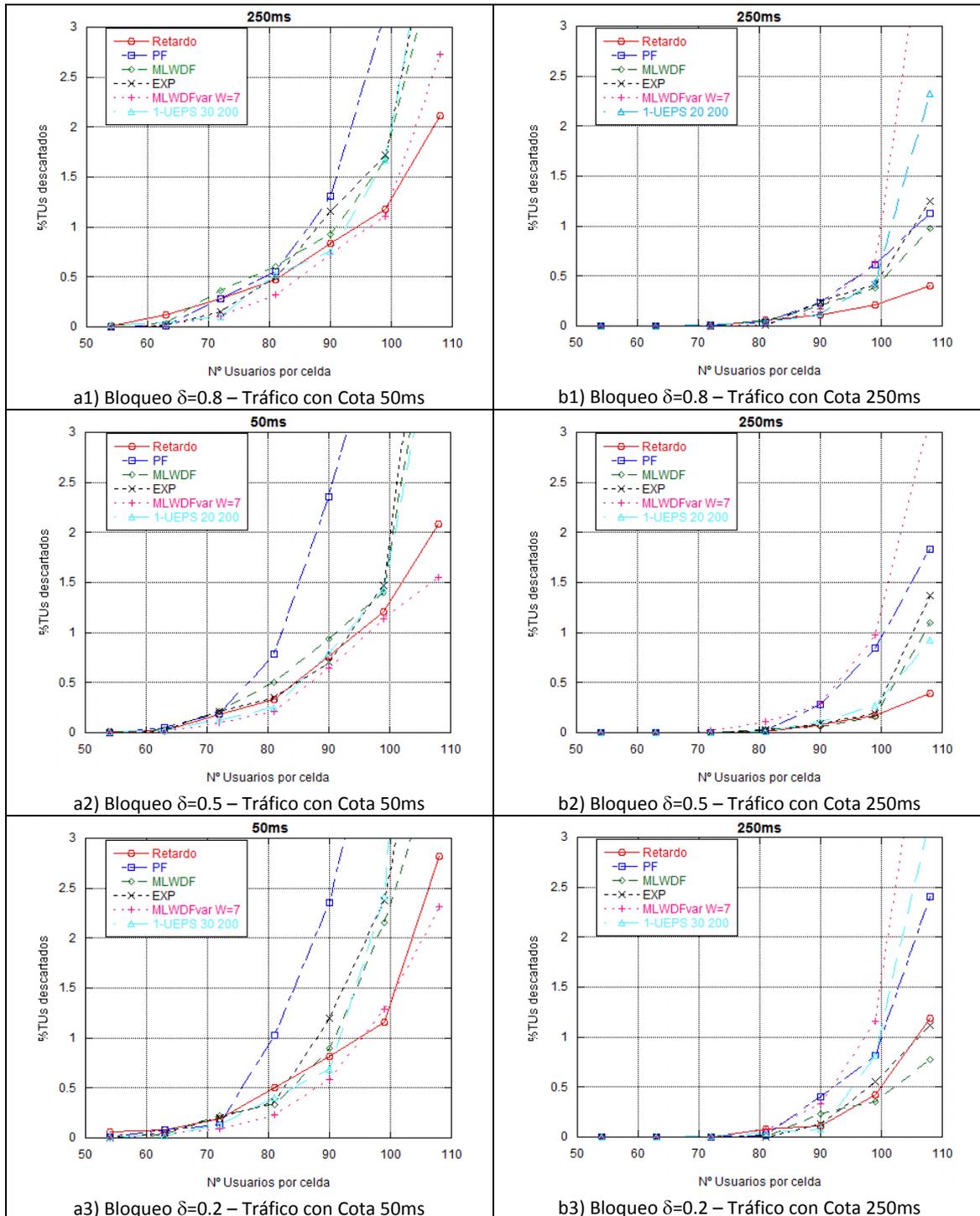


Figura B.18: Comparación bloqueos para los distintos algoritmos (rB). Escenario multiservicio.

B.2: Bloqueo por estado de los buffers

B.2.1: Ráfagas A. Tráfico uniforme.

Volvemos al escenario correspondiente al caso en el que todos los usuarios de la red generan el mismo tipo de tráfico de retardo máximo $d_{max}=50ms$, con ráfagas de tiempos de ON/OFF de 30 y 90ms respectivamente.

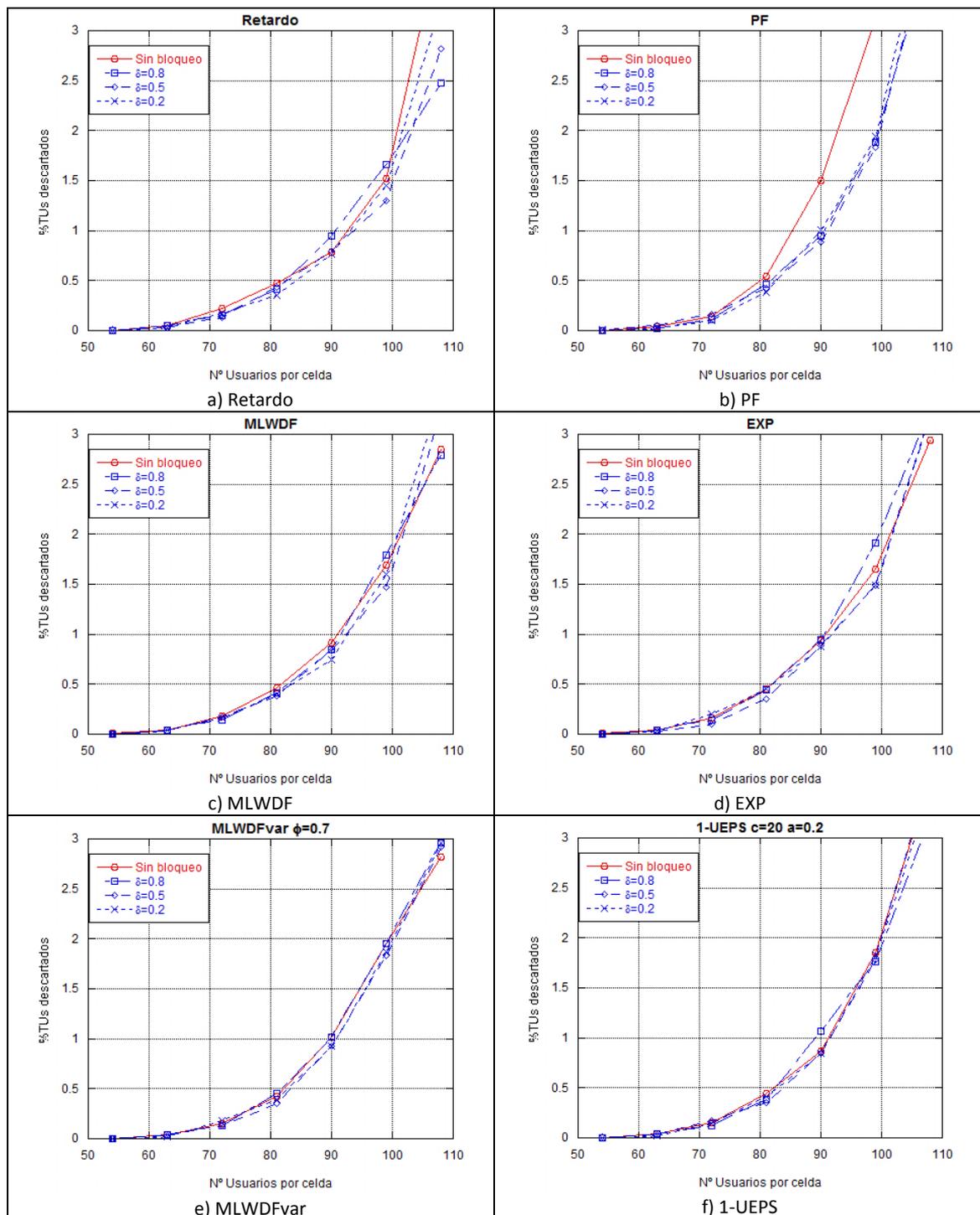


Figura B.19: Comparación entre algoritmos con/sin bloqueo con tráfico uniforme (rA).

Scheduling y provisión de QoS en redes móviles 4G OFDMA

Vemos en la figura B.19 que la mayoría de los algoritmos no sufren prácticamente variaciones en sus resultados, a excepción del PF, que manifiesta la mejora más significativa al aplicar bloqueo, y el Retardo, que mejora en general a pesar de que con $\delta = 0.8$, ve reducido el número de usuarios aceptables por celda cumpliendo los criterios de QoS. Con este mismo valor de δ , el algoritmo 1-UEPS sufre también esta reducción de capacidad, pero con los demás prácticamente no modifica su comportamiento respecto al caso sin bloqueo.

A continuación se muestran los resultados que se obtienen cambiando los parámetros de las ecuaciones de los algoritmos con los distintos tipos de bloqueo. En la figura B.20 vemos lo que sucede con el algoritmo MLWDFvar. La influencia del parámetro ϕ en los resultados es de nuevo escasa.

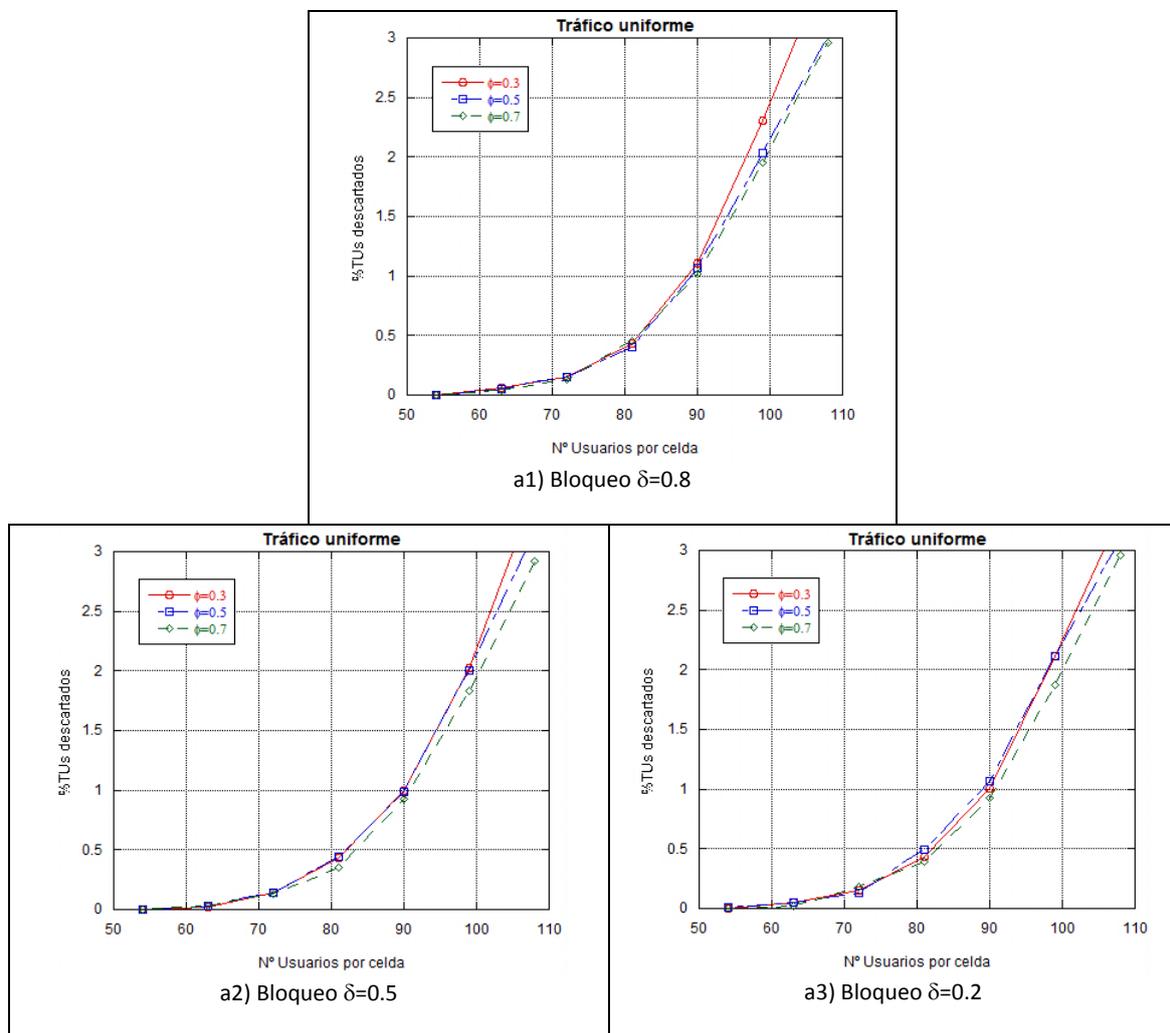


Figura B.20: Algoritmo MLWDFvar con tráfico uniforme (rA).

En la figura B.21 vemos de qué forma afecta el bloqueo al algoritmo 1-UEPS. Con este tipo de bloqueo se mantiene el buen funcionamiento de algoritmo con $c=20$, $a=0.2$, tal y como sucedía en el caso sin bloqueo, aunque con $\delta = 0.8$ no son los valores que

Scheduling y provisión de QoS en redes móviles 4G OFDMA

obtienen mayor capacidad para el sistema por su peor comportamiento en situaciones de poca carga en el sistema.

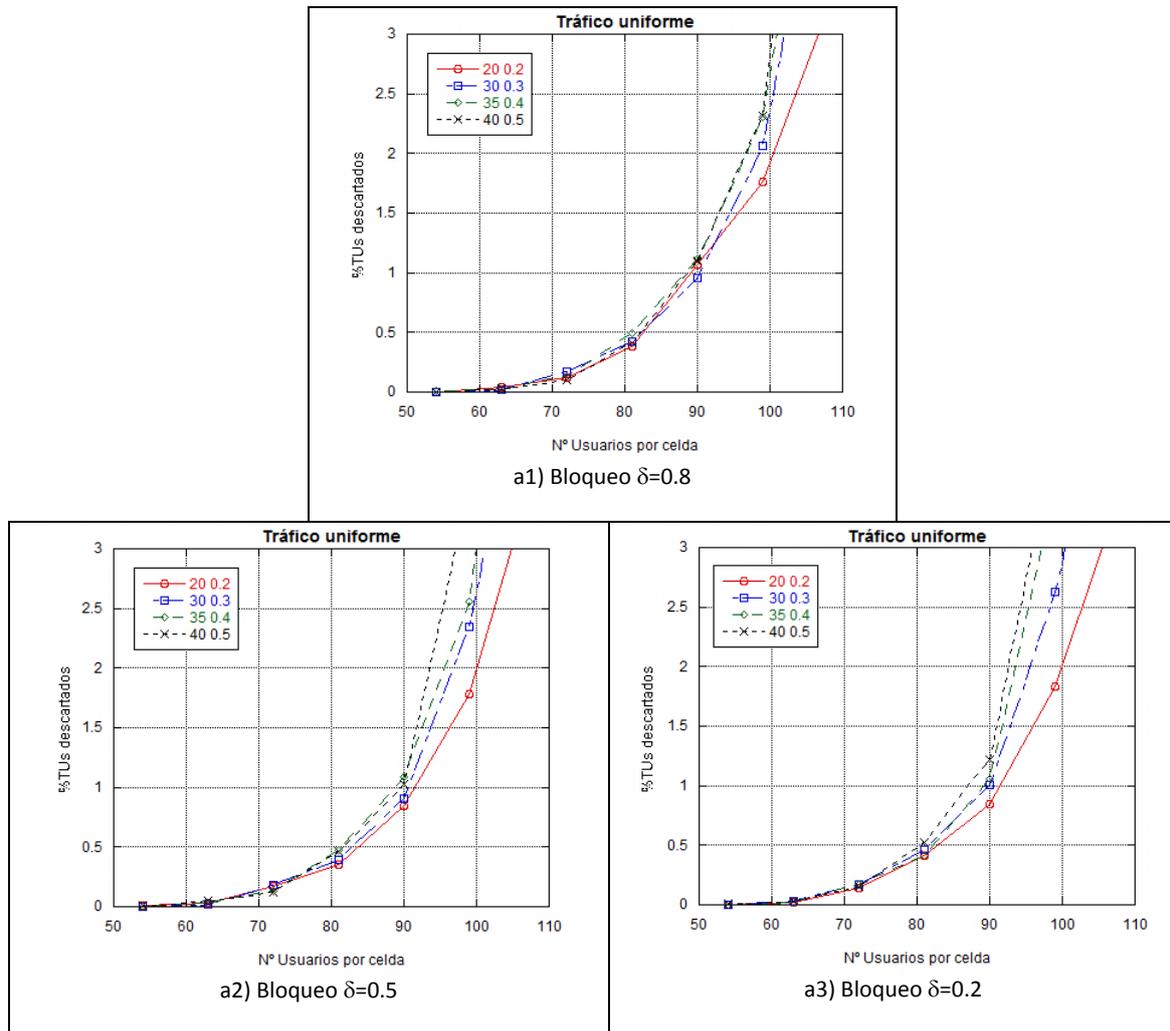


Figura B.21: Algoritmo 1-UEPS con tráfico uniforme (rA).

Al igual que sucedía con el bloqueo temporal en este escenario, la mejora resultante de aplicar bloqueo inicial sobre el algoritmo PF hace que las diferencias entre los distintos algoritmos sean menores que en el caso sin bloqueo, manteniéndose aún así el PF como el que peor funciona. Vemos que el Retardo es el que mejor funciona excepto en el caso de $\delta = 0.8$, valor con el que experimenta cierta reducción de su capacidad, lo que hace que sus prestaciones se vean superadas por las del algoritmo MLWDF.

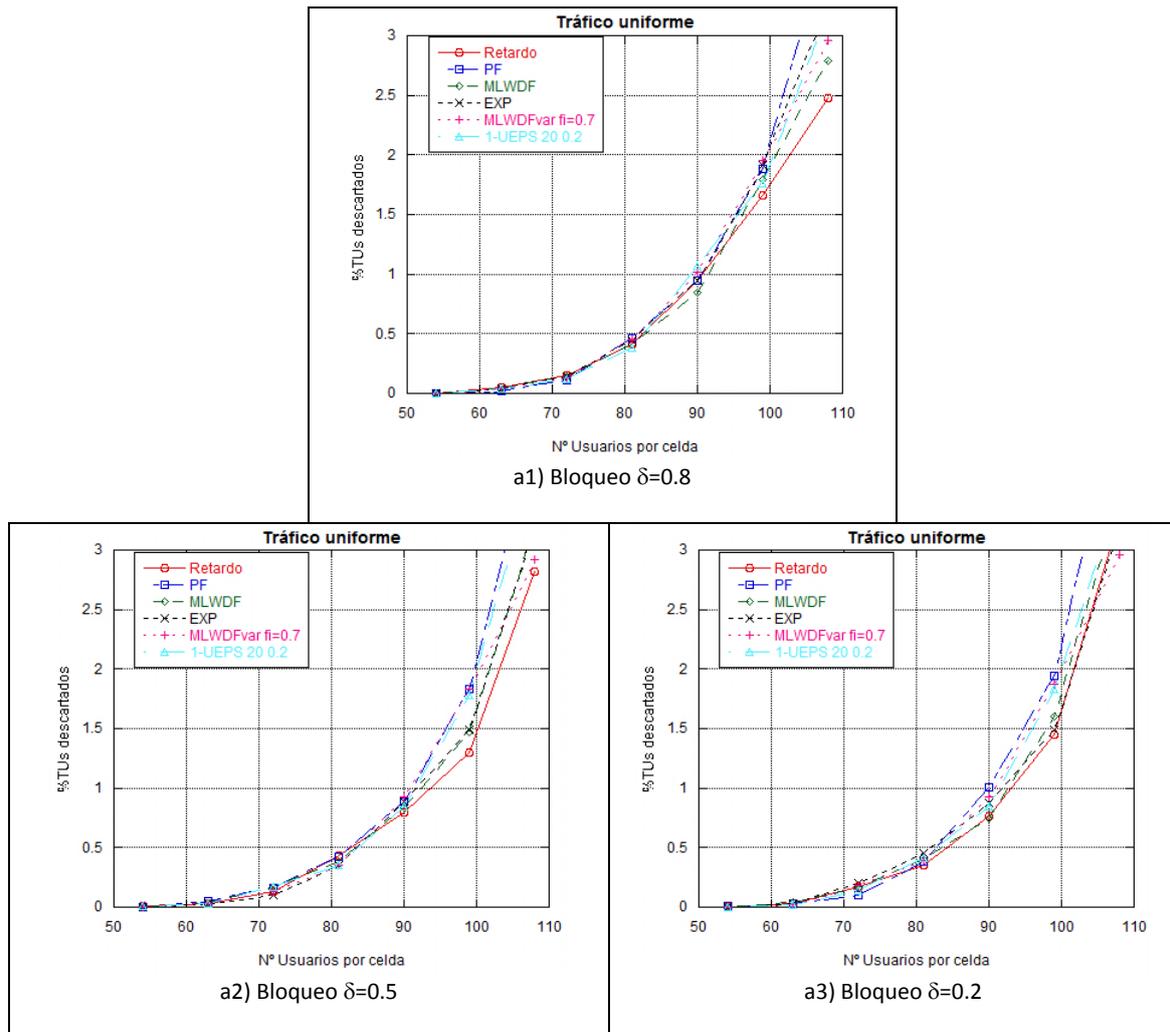


Figura B.22: Comparación bloqueos para los distintos algoritmos (rA). Tráfico uniforme.

B.2.2: Ráfagas A. Escenario multiservicio.

Evaluaremos en este apartado un escenario que mantiene las ráfagas del anterior pero con tráfico de los dos tipos ya conocidos: $d_{max1}=50ms$ y $d_{max2}=250ms$.

En la figura B.23 se observa que el bloqueo afecta de distinta manera a los algoritmos en función del valor de δ , siendo particularmente llamativo que todos excepto el PF empeoran sus prestaciones con $\delta=0.8$, por lo que no sería recomendable usar ese tipo de bloqueo en este escenario. Con $\delta=0.5$ se consiguen mejorar los resultados de algunos algoritmos, siendo muy interesante que con este valor se evita que se dispare el porcentaje de descartes cuando el sistema tiene un alto nivel de carga.

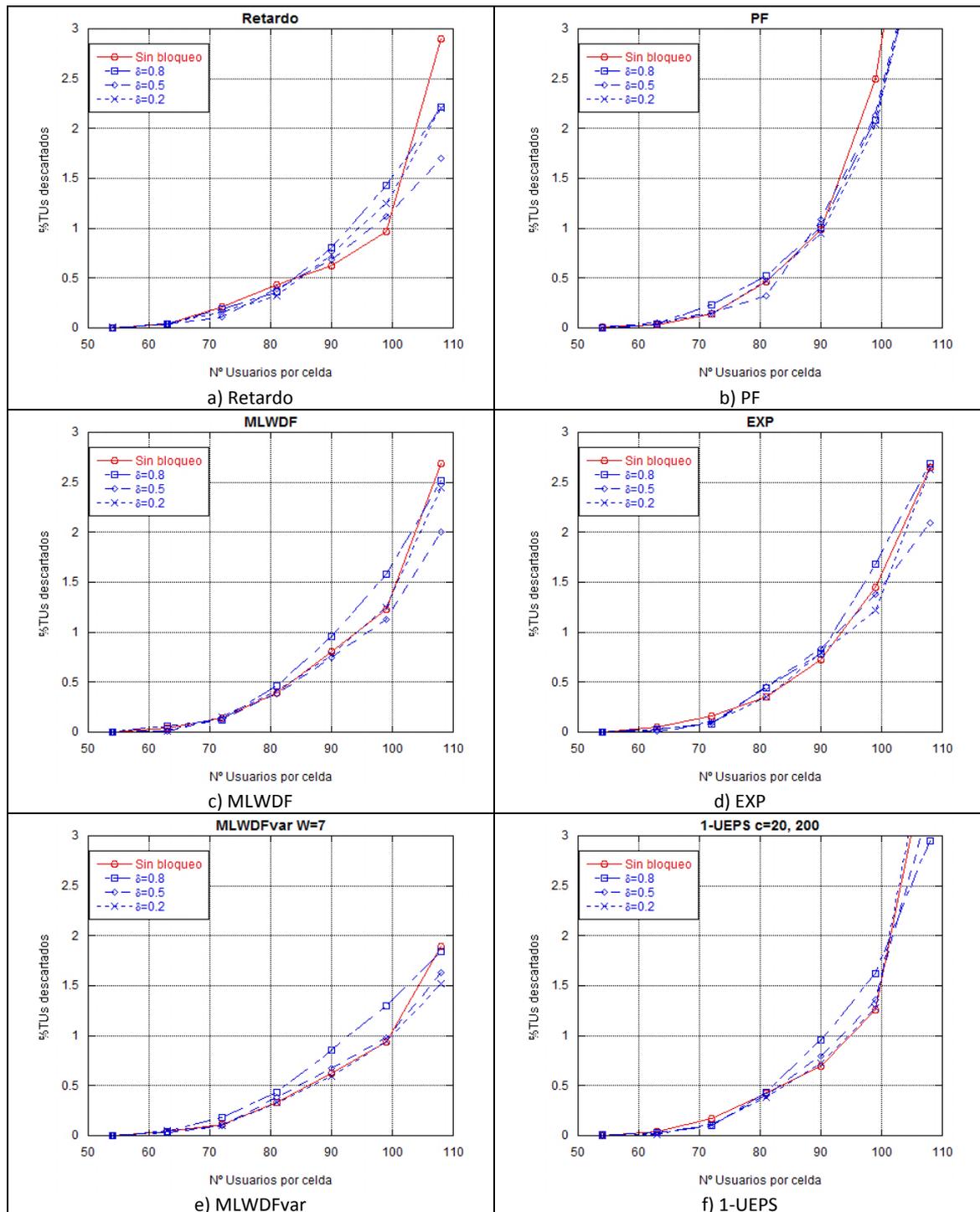


Figura B.23: Comparación entre algoritmos con/sin bloqueo en un escenario multiservicio (50% $d_{max}=50ms$ +50% $d_{max}=250ms$) (rA). Tráfico de cota 50ms

En la figura B.24 vemos que el efecto en el tráfico de 250ms, que empeora sus resultados para MLWDF, EXP y PF y los mejora ligeramente para Retardo, MLWDFvar y 1-UEPS. Vemos que para el algoritmo Retardo la respuesta al bloqueo de los dos tipos de tráfico es muy similar.

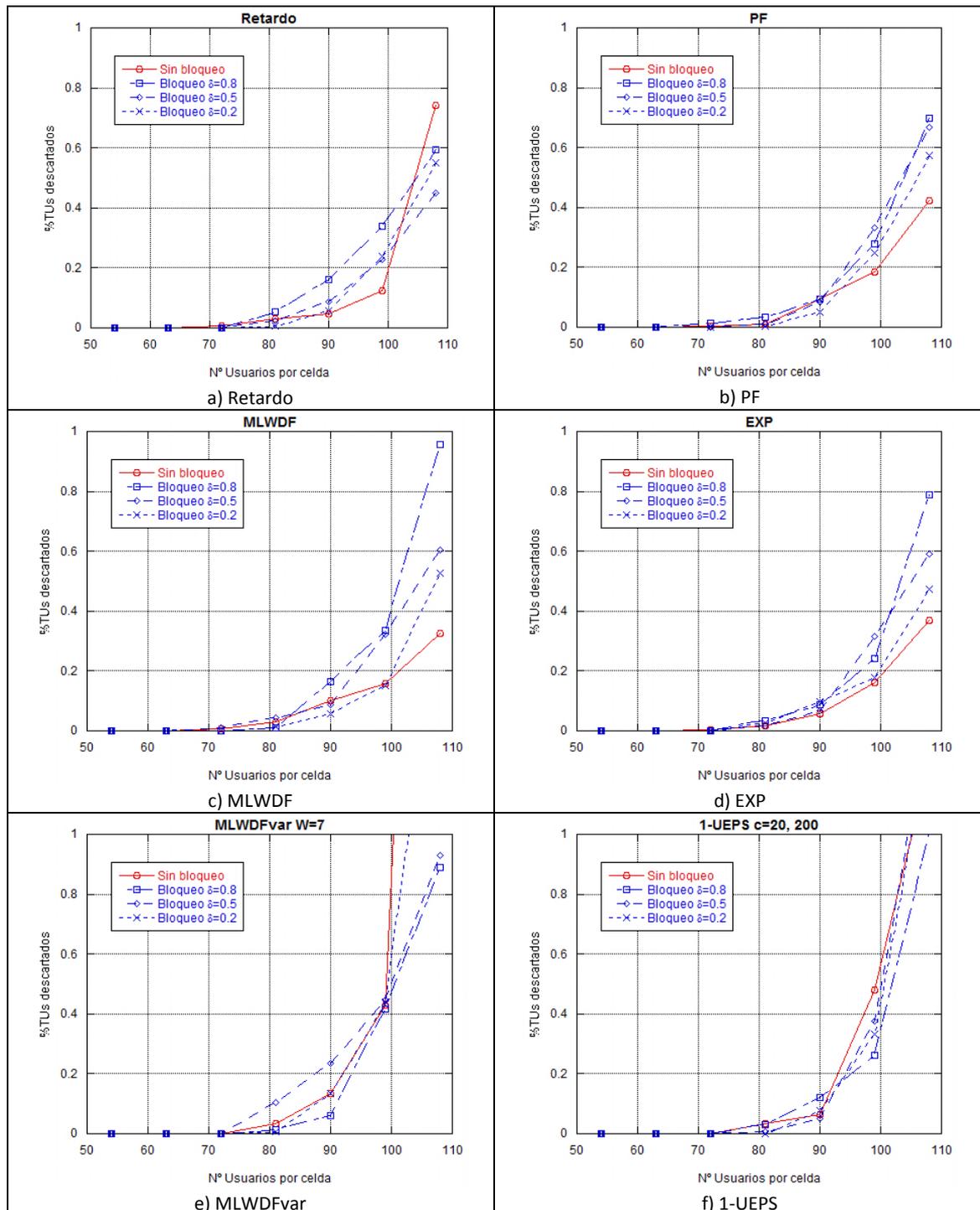


Figura B.24: Comparación entre algoritmos con/sin bloqueo en un escenario multiservicio ($50\% d_{max}=50ms + 50\% d_{max}=250ms$) (rA). Tráfico de cota 250ms

En la figura B.25 podemos ver como al aplicar bloqueo inicial por el estado de los buffers vemos que el algoritmo tiene una forma similar a la del caso sin bloqueo excepto en el caso de $\delta = 0.8$, en el que sobre todo los resultados de $W=7$, $W=10$ empeoran. Los resultados para el tráfico de 250ms mejoran sobre todo en casos de alta carga en el sistema.

Scheduling y provisión de QoS en redes móviles 4G OFDMA

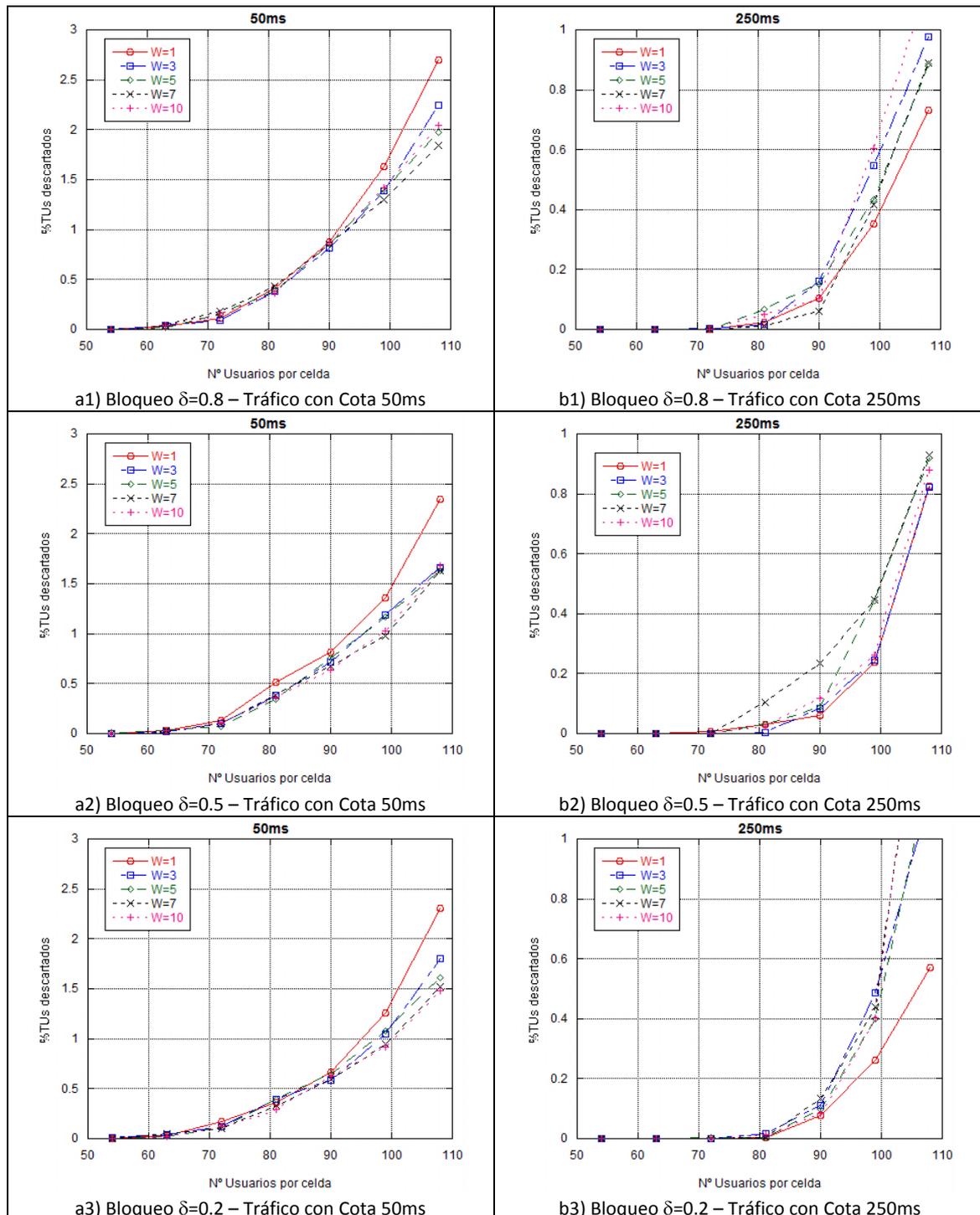


Figura B.25: Algoritmo MLWDFvar en el escenario multiservicio (rA).

En la figura B.26 podemos ver el algoritmo 1-UEPS, que cambia respecto al caso sin bloqueo en que ahora los resultados de los distintos parámetros son muy parejos hasta alcanzar el 1% de descartes.

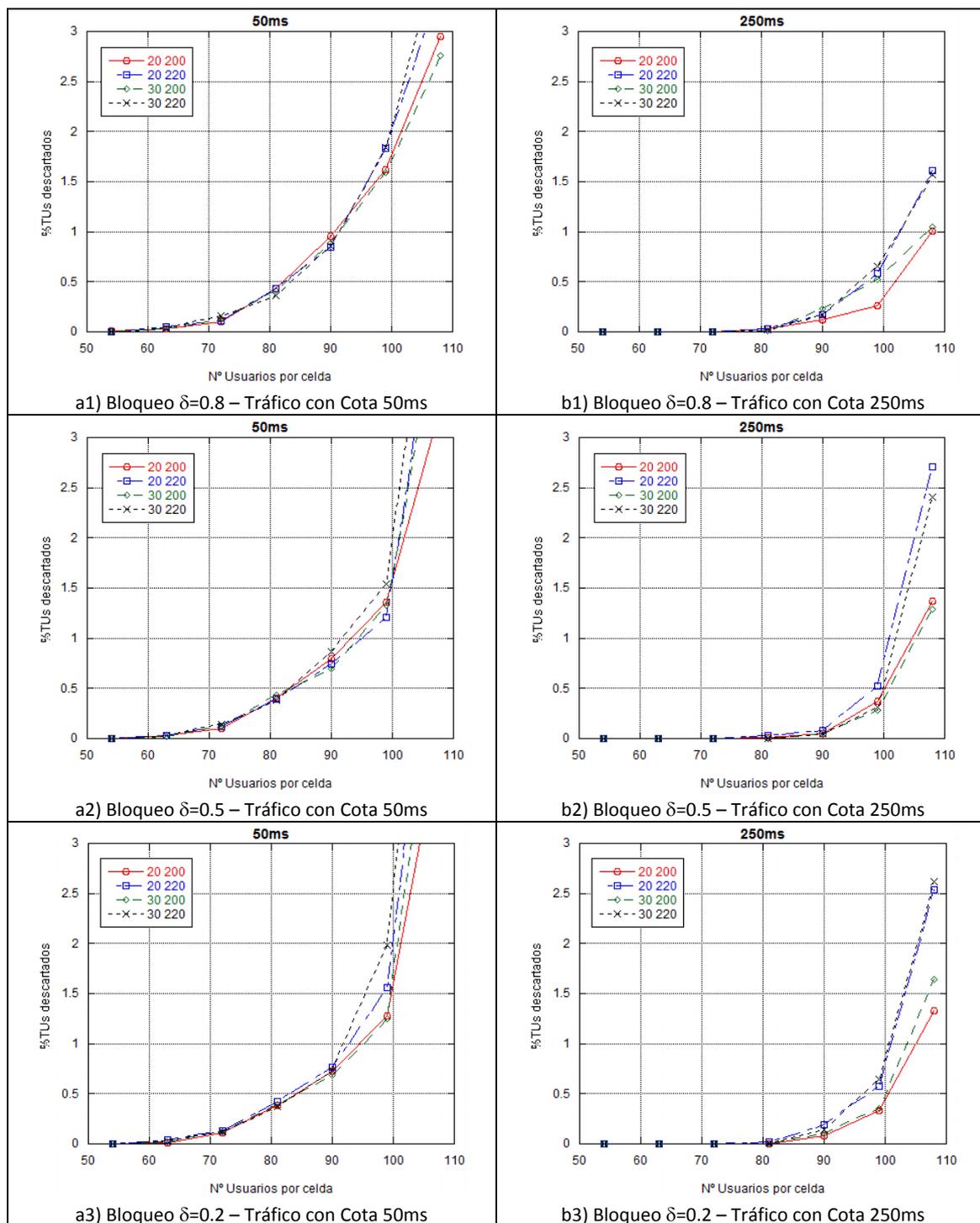


Figura B.26: Algoritmo 1-UEPS en el escenario multiservicio (rA).

En la figura B.27 tenemos las comparaciones de los resultados de todos los algoritmos. Como sucedía en el caso del bloqueo por retardo, no se mejoran los resultados respecto al caso sin bloqueo, pero con $\delta=0.2$ el algoritmo MLWDFvar mantiene sus prestaciones de capacidad y reduce el porcentaje de descartes en casos de elevada carga.

Scheduling y provisión de QoS en redes móviles 4G OFDMA

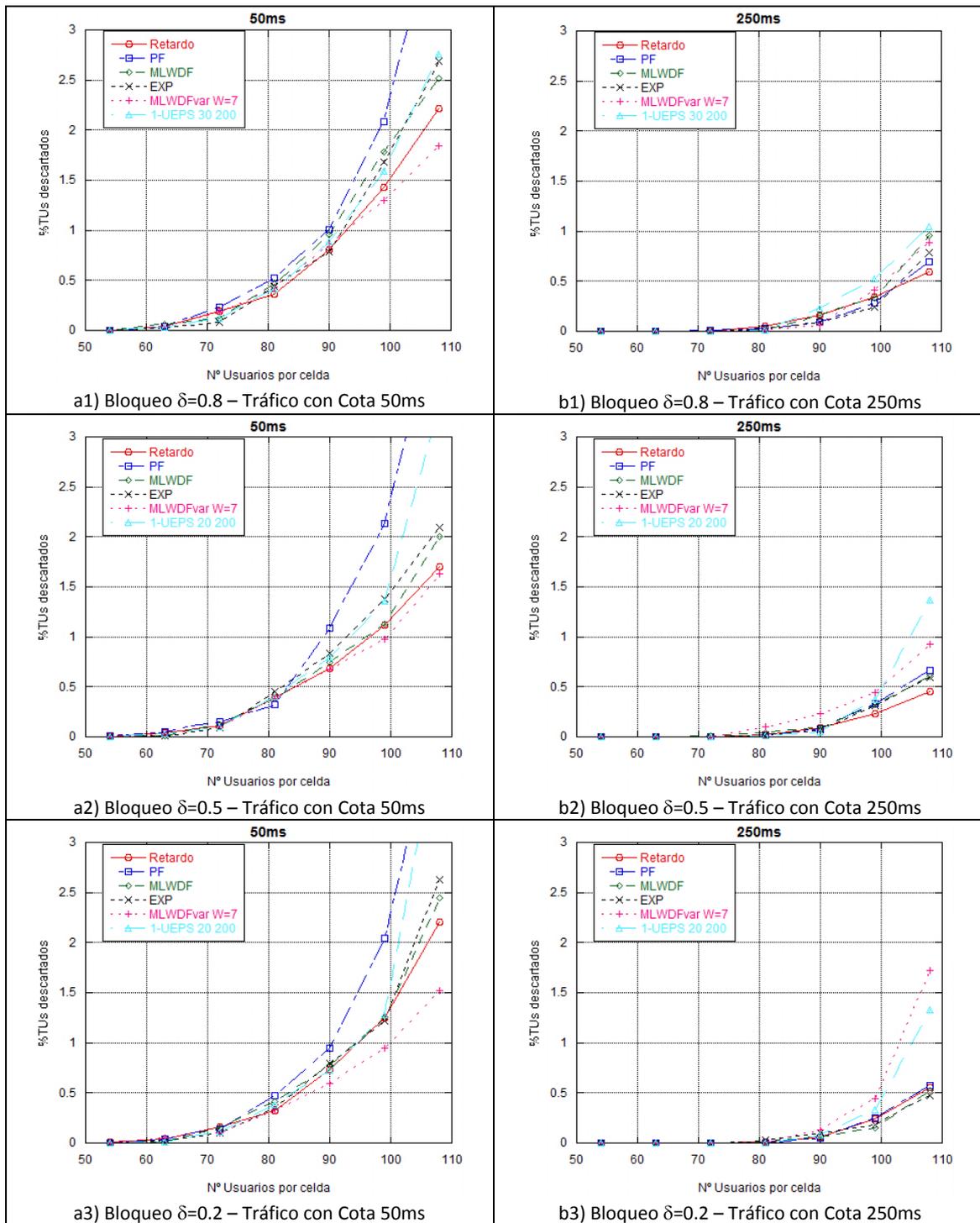


Figura B.27: Comparación bloqueos para los distintos algoritmos (rA). Escenario multiservicio.

B.2.3: Ráfagas B. Tráfico uniforme.

En este apartado se evalúa un escenario correspondiente al caso en el que todos los usuarios de la red generan un tráfico uniforme de cota máxima de retardo $d_{max}=50ms$, utilizando ráfagas con periodos de actividad y silencio de 300 y 900 milisegundos respectivamente.

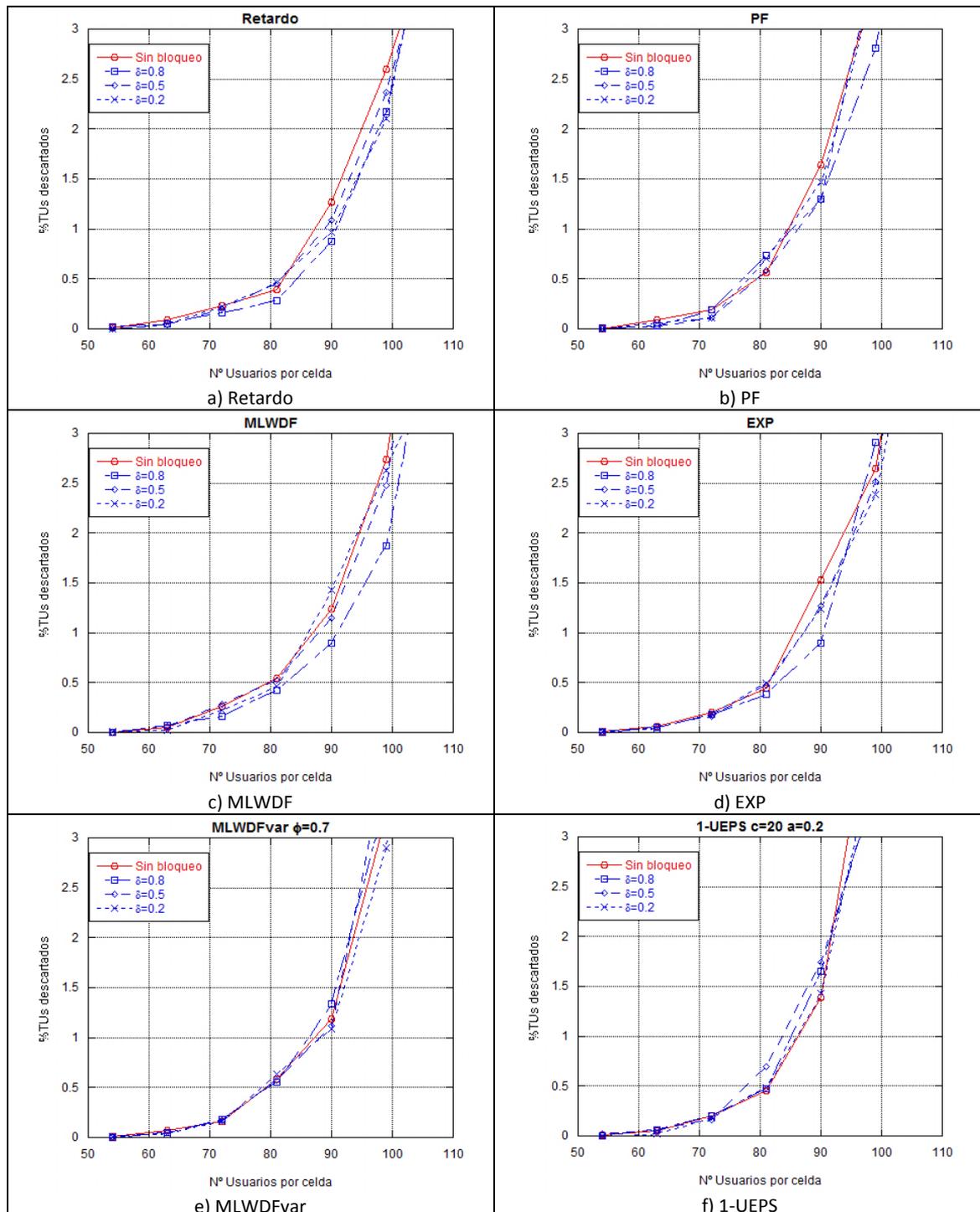


Figura B.28: Comparación entre algoritmos con/sin bloqueo con tráfico uniforme (rB).

Scheduling y provisión de QoS en redes móviles 4G OFDMA

La figura B.28 contiene los cambios al aplicar bloqueo por estado de los buffers en este escenario. Los algoritmos Retardo y PF mejoran ligeramente sus resultados, y 1-UEPS los empeora. El resto de algoritmos mantienen unas prestaciones parecidas al caso sin bloqueo, a excepción del caso $\delta=0.8$, con el que obtenemos sustanciosas mejoras para el MLWDF y el EXP.

En la figura B.29 vemos el algoritmo MLWDFvar. En ella se observa que las capacidades obtenidas con los valores de ϕ 0.5 y 0.7 son muy similares para todos los valores de δ , aunque para $\delta=0.8$, $\phi=0.5$ evoluciona mejor al aumentar la carga del sistema.

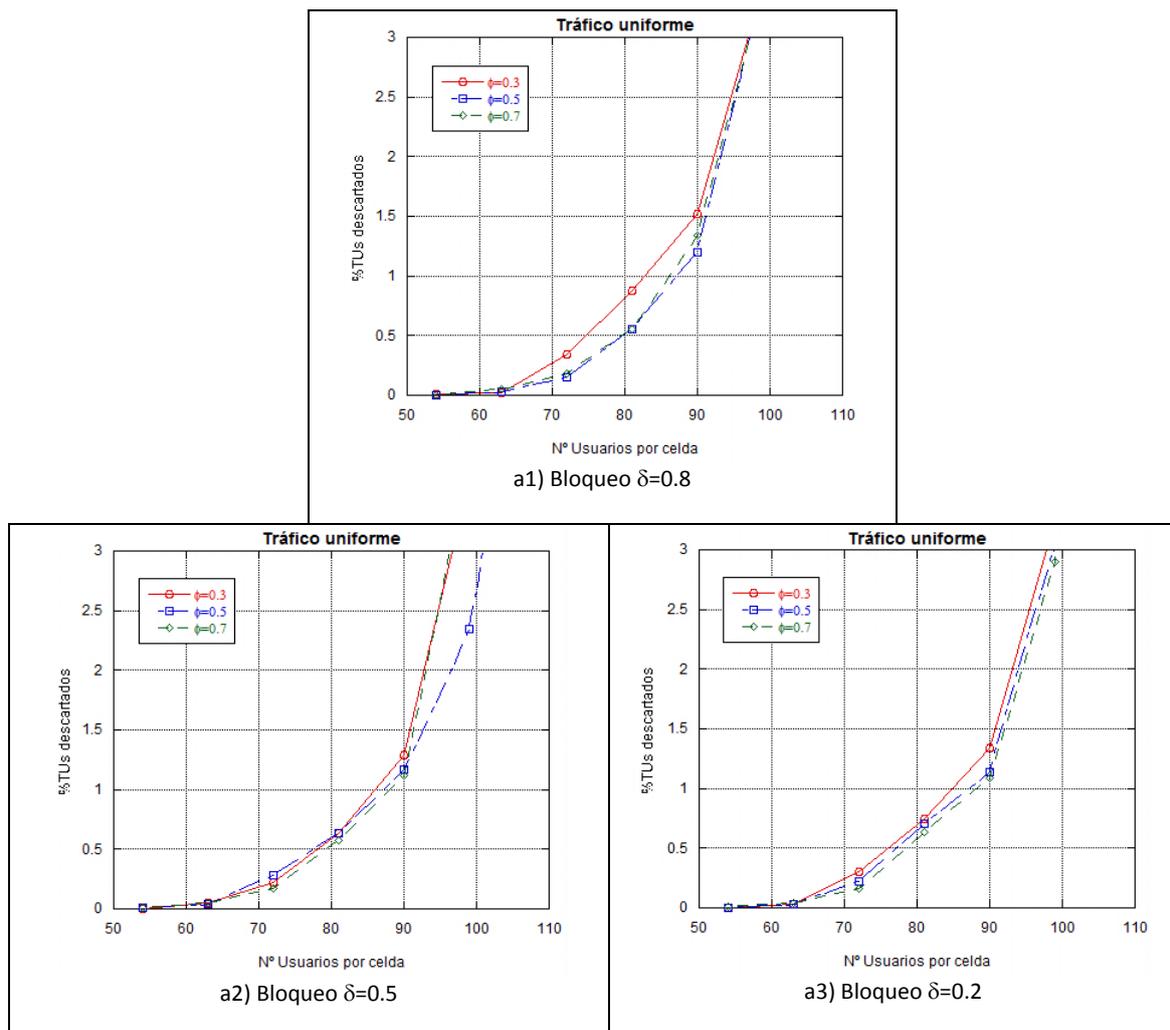


Figura B.29: Algoritmo MLWDFvar con tráfico uniforme (rB).

Los resultados del algoritmo 1-UEPS en este escenario se encuentran en la figura B.30. En este caso, con $c=20$, $a=0.2$ y $c=30$, $a=0.3$ se obtienen los mejores resultados, siendo muy pequeñas las diferencias entre los resultados con estos distintos parámetros. Con el resto de parámetros también tenemos unas prestaciones similares, excepto en el caso de $\delta=0.2$, con el que tienen un funcionamiento algo peor.

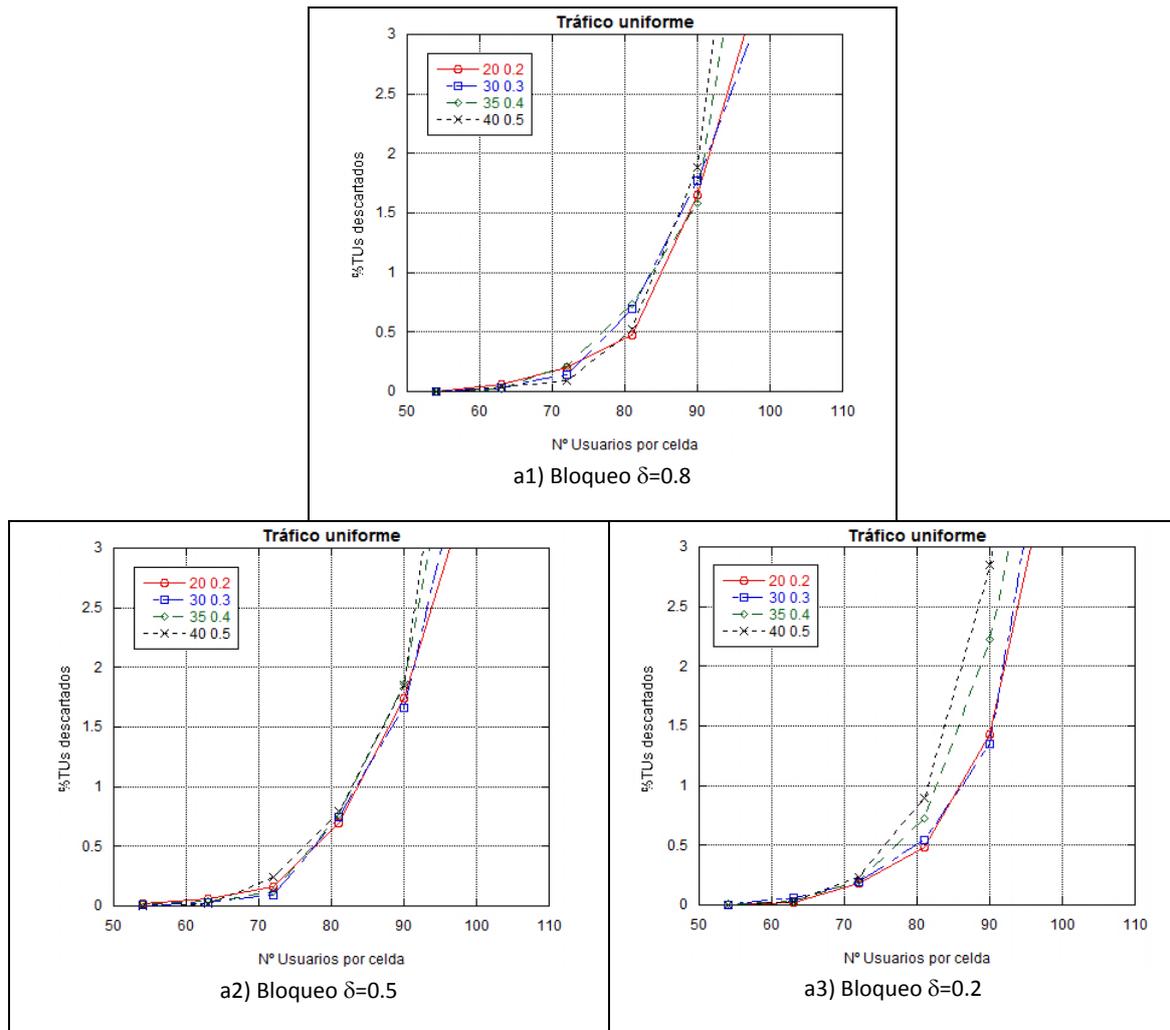


Figura B.30: Algoritmo 1-UEPS con tráfico uniforme (rB).

En la figura B.31 tenemos la comparación de los mejores resultados para los distintos algoritmos. Al igual que sucede en la gran mayoría de los casos, con tráfico uniforme el algoritmo Retardo es el que mejores prestaciones ofrece, excepto en el caso de $\delta=0.8$ donde, debido a la gran mejora experimentada por el algoritmo MLWDF que hace que se sitúe en esta ocasión como el que daría una mayor capacidad al sistema con una ligera ventaja sobre el Retardo, mejorándolo en mayor medida cuando la carga del sistema crece.

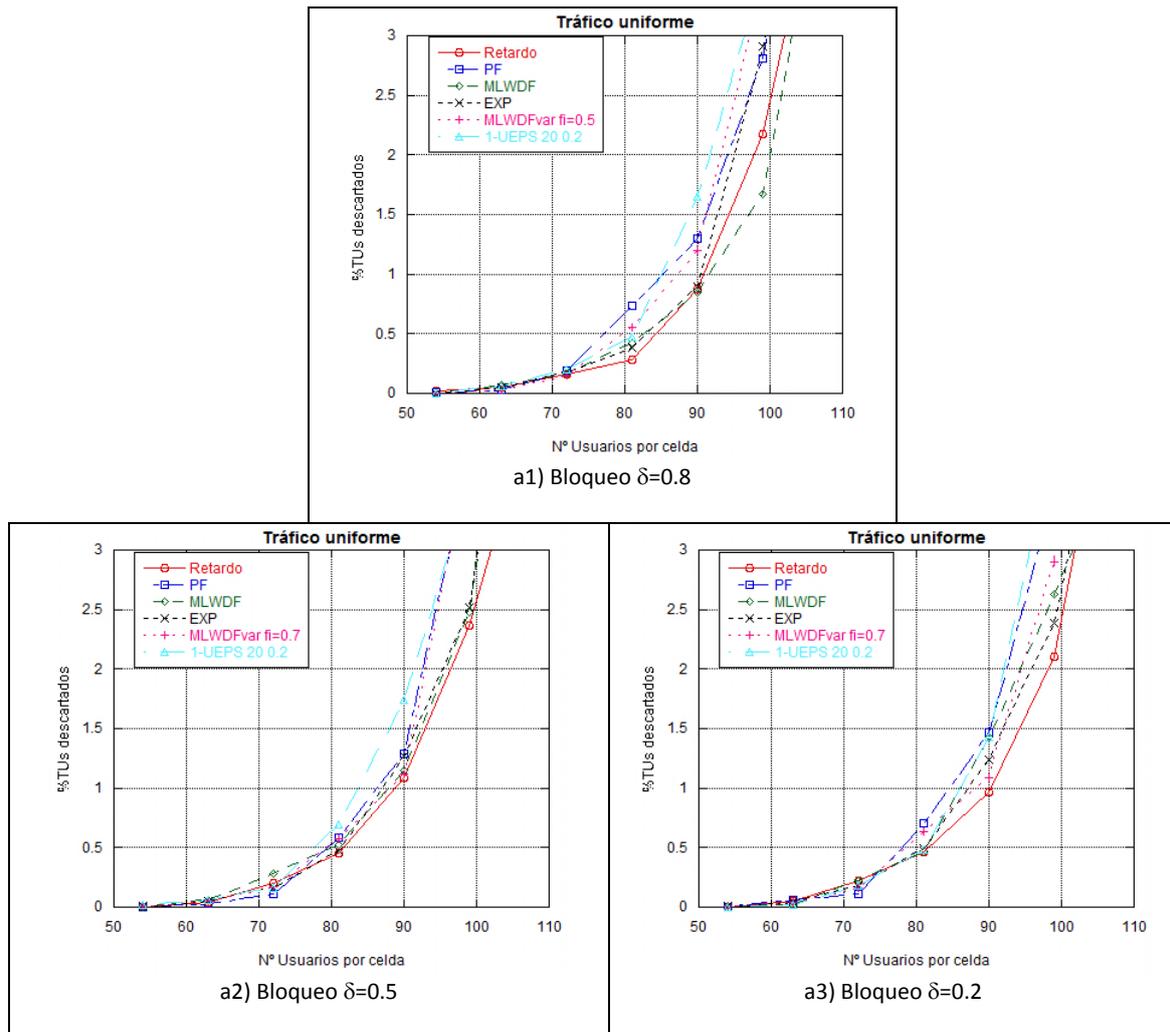


Figura B.31: Comparación bloqueos para los distintos algoritmos (rB). Tráfico uniforme.

B.2.4: Ráfagas B. Escenario multiservicio.

Por último se prueba un escenario que mantiene las ráfagas largas ($T_{ON}=300\text{ms}$, $T_{OFF}=900\text{ms}$) e incluye varios tipos de tráfico, con distintas cotas máximas de retardo (50ms y 250ms).

En la figura B.32 se observa que, en la mayoría de los casos, el bloqueo tiene un efecto positivo en este escenario. La mayoría de los algoritmos mejoran levemente sus prestaciones, siendo especialmente interesantes las mejoras para $\delta=0.8$ en el Retardo y MLWDFvar, ya que son los algoritmos que daban unas mejores prestaciones en el caso sin bloqueo.

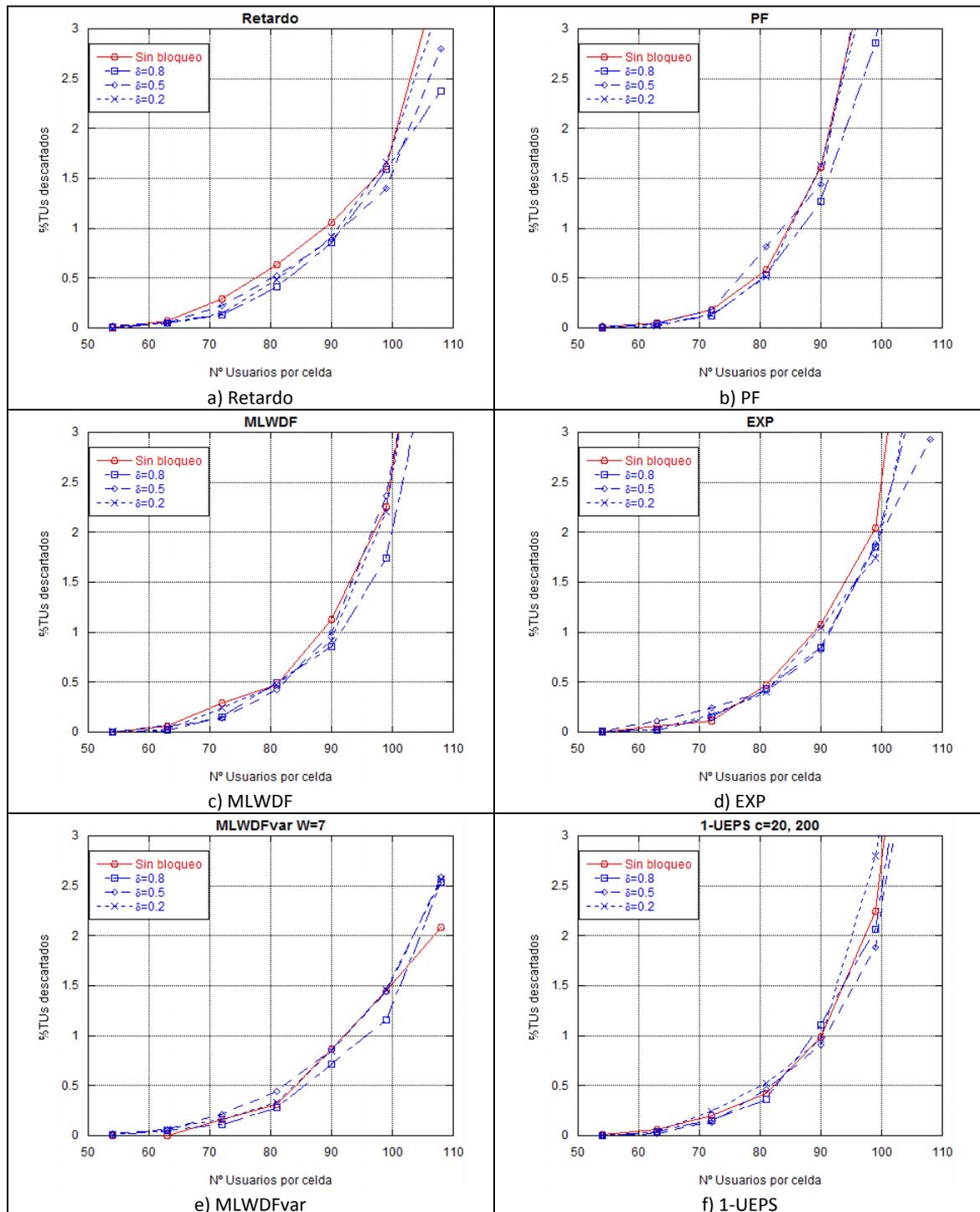


Figura B.32: Comparación entre algoritmos con/sin bloqueo en un escenario multiservicio (50% $d_{max}=50ms$ +50% $d_{max}=250ms$) (rB). Tráfico de cota 50ms

En la figura B.33 vemos que en el tráfico de 250ms tenemos un comportamiento opuesto al que se experimentaba en el caso de las ráfagas A, ya que mejoran todos los algoritmos excepto el MLWDFvar, que en aquel caso era el único que tenía cierta mejora. Habría que tener cuidado con los rápidos crecimientos del porcentaje de descartes para este algoritmo en los casos de carga alta en el sistema.

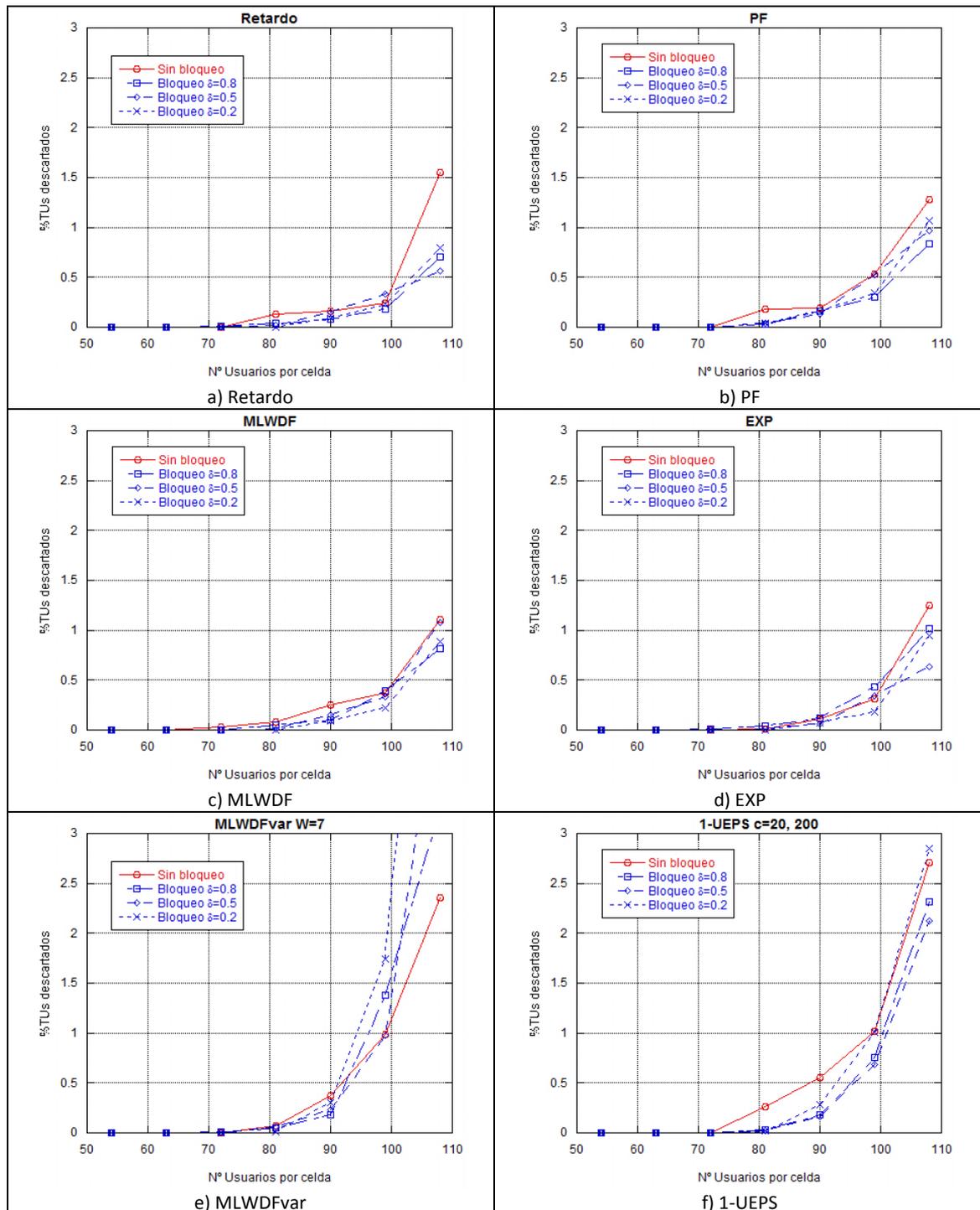


Figura B.33: Comparación entre algoritmos con/sin bloqueo en un escenario multiservicio ($50\% d_{max}=50ms + 50\% d_{max}=250ms$) (rB). Tráfico de cota 250ms

En la figura B.34 se aprecian las diferencias entre los resultados obtenidos para MLWDFvar con los distintos pesos W , y en este caso los valores que dan unas mejores prestaciones son $W=7$ y $W=10$. Se observa en ella que para el tráfico de 250ms, gran parte de los casos sufren un muy elevado crecimiento del descarte cuando la carga del sistema aumenta, como sucedía con el bloqueo por retardo.

Scheduling y provisión de QoS en redes móviles 4G OFDMA

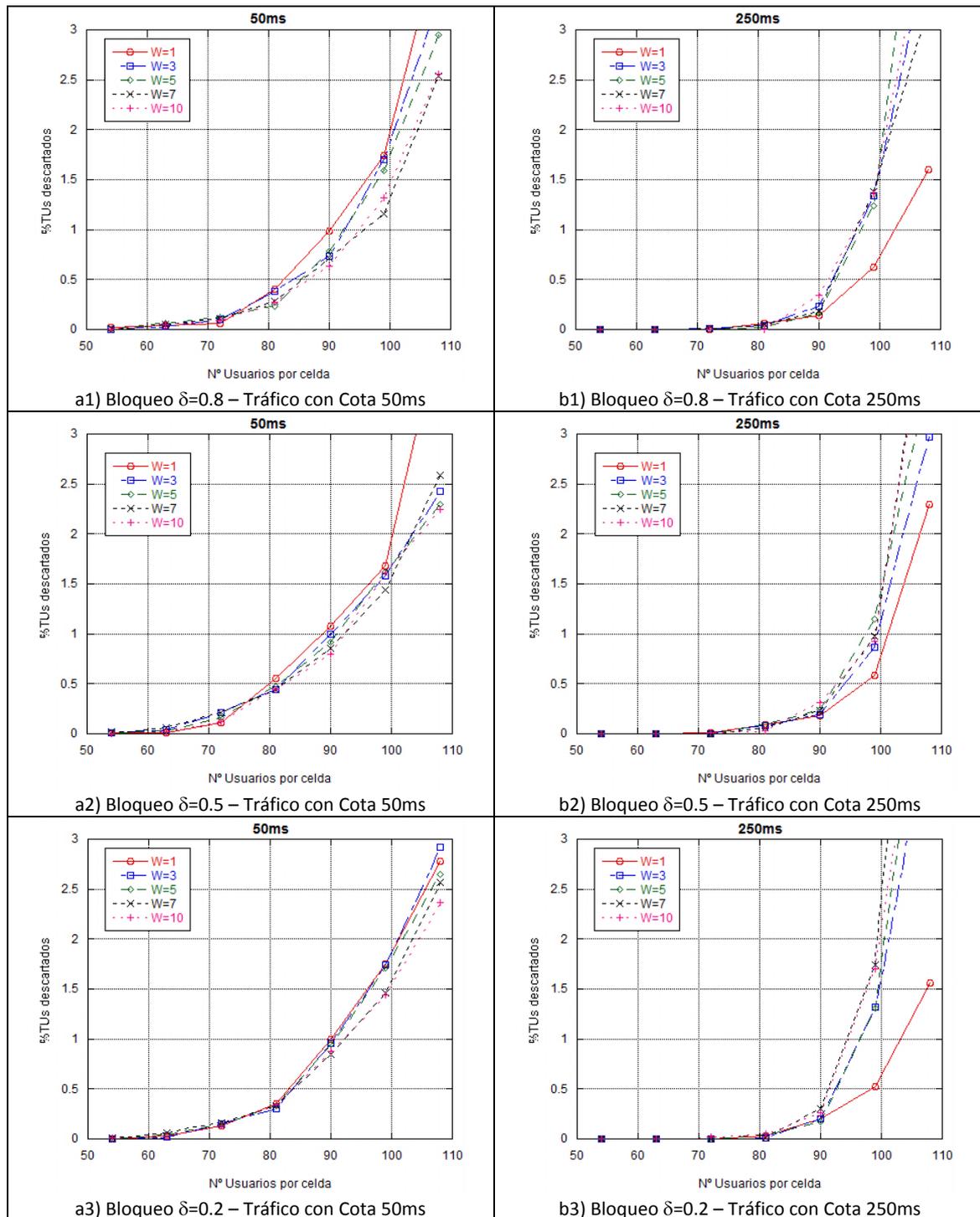


Figura B.34: Algoritmo MLWDFvar en el escenario multiservicio (rB).

Lo que sucede para el algoritmo 1-UEPS puede verse en la figura B.35, y excepto en el caso de $\delta=0.8$ los distintos valores dan unos resultados muy parecidos en los casos de baja carga con el tráfico de 50ms. Para el de 250ms vemos que el porcentaje de descartes depende en mayor medida de los parámetros utilizados, siendo los mejores resultados los obtenidos con $c=20, 200$.

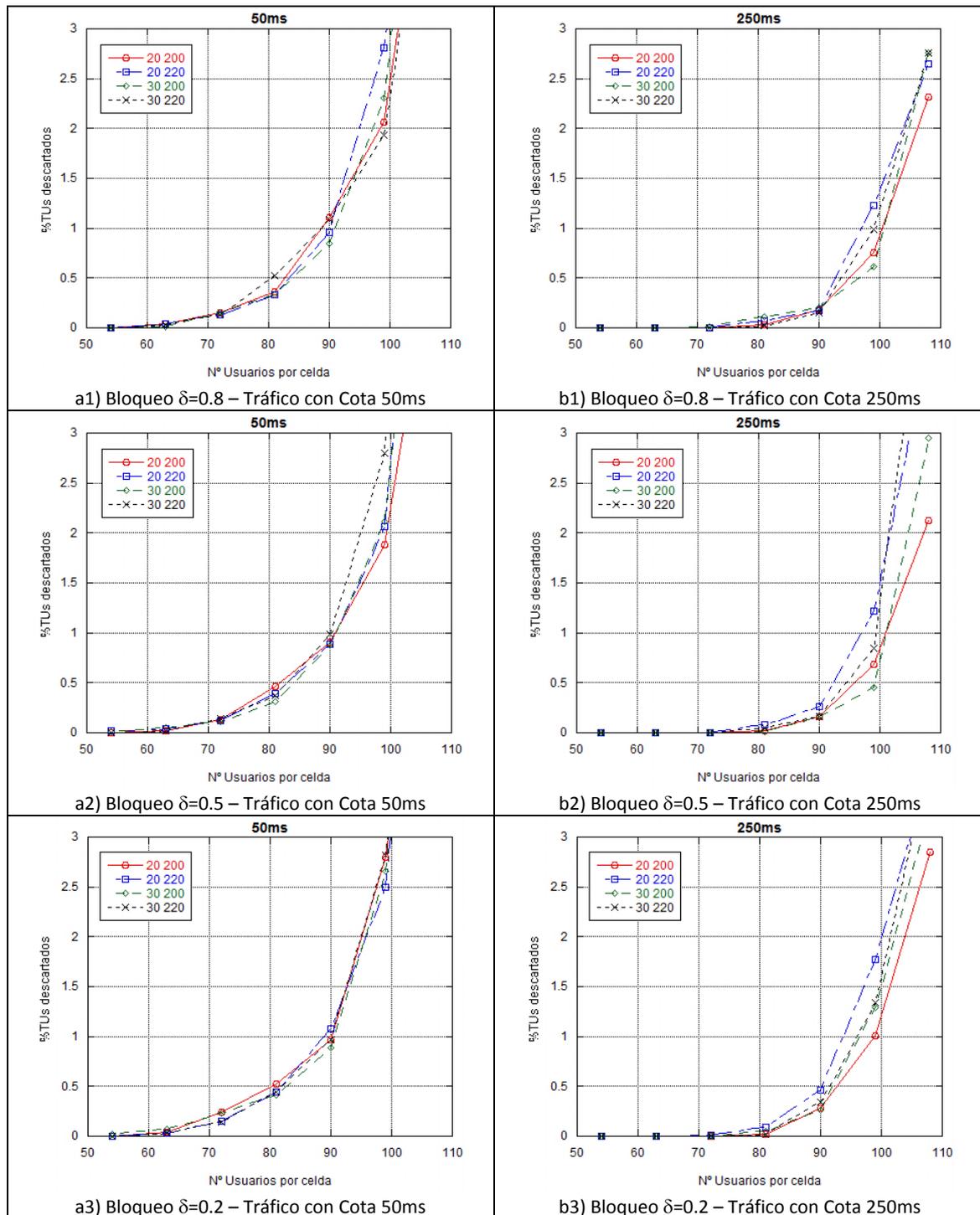


Figura B.35: Algoritmo 1-UEPS en el escenario multiservicio (rB).

Para acabar comparamos los resultados de todos los algoritmos evaluados en este escenario en la figura B.36. De nuevo se obtiene una mejora de las prestaciones y un aumento de la capacidad gracias a la aplicación de este tipo de bloqueo, sobre todo gracias a la mejora del MLWDFvar, que particularmente con $\delta=0.8$ mejora en unos 4 usuarios al Retardo, que es el siguiente algoritmo en cuanto a buenos resultados.

Scheduling y provisión de QoS en redes móviles 4G OFDMA

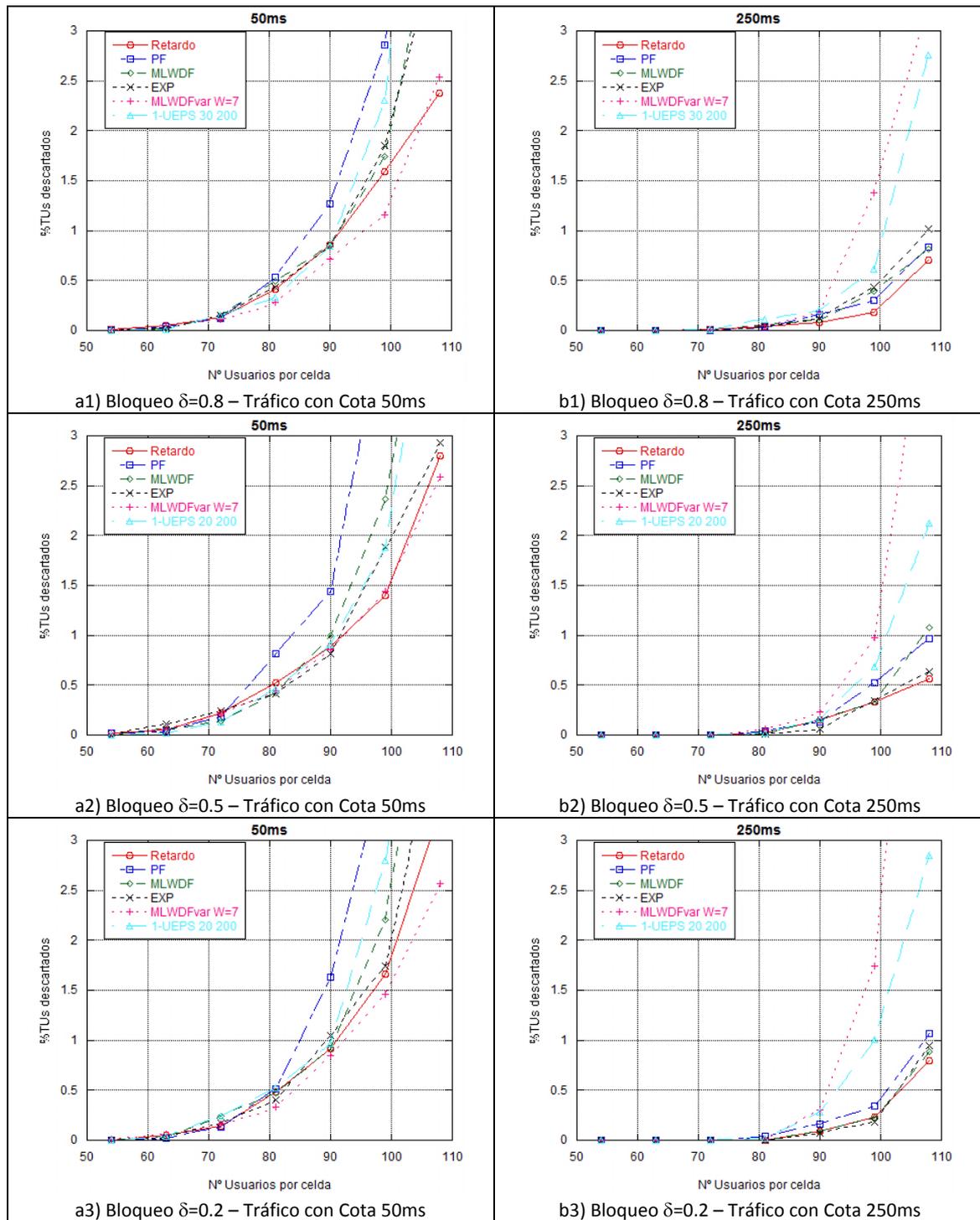


Figura B.36: Comparación bloqueos para los distintos algoritmos (rB). Escenario multiservicio.

Los cambios más relevantes resultantes de la aplicación de bloqueo están más extensamente comentados en el texto de la memoria.