Universidad de Málaga

Escuela Técnica Superior de Ingeniería de Telecomunicación

Tesis Doctoral

EVALUACIÓN DE TÉCNICAS DE GESTIÓN DE RECURSOS RADIO EN SISTEMAS CELULARES LTE-A

Autor

Isabel María Delgado Luque

Directores

Mª Carmen Aguayo Torres y **Gerardo Gómez Paredes**

Año 2019

Agradecimientos

Muchas gracias a mi familia por el apoyo que siempre me ha dado; a Jose, por estar siempre ahí y a mis directores de tesis, Gerardo y Mari Carmen, por animarme a continuar con este proyecto hasta el final.

Esta tesis ha sido parcialmente financiada por el Ministerio de Economía y Competitividad a través de los Proyectos TEC2013-44442-P y TEC2016-80090-C2-1-R.

Índice general

Índice de figuras

Índice de tablas

Lista de acrónimos

3GPP	3rd Generation Partnership Project
ABS	Almost Blank Subframe
AMC	Adaptive Modulation and Coding
AP	Access Point
ARQ	Automatic Repeat Request
AWGN	Additive White Gaussian Noise
BBU	Base Band Unit
BC	Best Channel
BER	Bit Error Rate
BLER	BLock Error Rate
BS	Base Station
CA	Carrier Aggregation
CC	Component Carrier
CD-EDD	Channel Dependent Earliest Deadline Due
CDF	Cumulative Distribution Function
CoMP	Coordinated Multipoint Transmission
CQI	Channel Quality Indicator
CRAN	Cloud Radio Access Network
CRC	Cyclic Redundancy Code
CRE	Cell Range Expansion
DL	Downlink
eICIC	enhanced Inter-Cell Interference Coordination
EDD	Earliest Deadline Due
eNB	enhanced NodeB

FDD	Frequency Division Duplex
FFR	Fractional Frequency Reuse
HARQ	Hybrid-Automatic Repeat Request
HetNets	Heterogeneous Networks
ICI	Inter-Cell Interference
ICIC	Inter-Cell Interference Coordination
ISI	Inter-Symbol Interference
JT	Joint Transmission
LA	Link Adaptation
LTE	Long Term Evolution
LTE-A	Long Term Evolution Advanced
M2M	Machine to Machine
MAC	Medium Access Control
MAP	Macro Access Point
MBS	Macro Base Station
MCS	Modulation and Coding Scheme
MIMO	Multiple Input Multiple Output
NR	Net Rate
OFDM	Orthogonal Frequency Domain Multiplexing
OLLA	Outer Loop Link Adaptation
PAP	Pico Access Point
PDF	Probability Density Function
PF	Proportional Fair
PFR	Partial Frequency Reuse
PRB	Physical Resource Block
QAM	Quadrature Amplitude Modulation
QCI	Quality Class Indicator
QoE	Quality of Experience
QoS	Quality of Service
RR	Round Robin
RRU	Remote Radio Unit
SBS	Small-cell Base Station

SFN	Single Frequency Network
SINR	Signal to Noise plus Interference Ratio
SISO	Soft Input Soft Ouput
SNR	Signal to Noise Ratio
SOVA	Soft Output Viterbi Algorithm
TB	Transport Block
TTI	Transmission Time Interval
UE	User Equipment
UL	Uplink
VoIP	Voice over IP

Resumen

Esta tesis doctoral está enfocada en el análisis de técnicas de gestión de recursos radio sobre redes celulares LTE-Advanced. En particular, se abordan dos grandes bloques de técnicas de gestión de recursos: en primer lugar, el análisis se centra en algoritmos de planificación de recursos radio; en segundo lugar, se han analizado aquellas técnicas orientadas a la gestión de interferencias sobre redes celulares.

Comenzando por los algoritmos de planificación de recursos radio, se han analizado varios algoritmos que tienen en cuenta requisitos de retardo. Además, se ha analizado la imparcialidad entre usuarios para un algoritmo Proportional Fair (PF) con criterios de Relación Señal a Ruido (SNR) frente al criterio de tasa de transmisión, obteniendo expresiones de forma cerrada para la distribución de la SNR por usuario y del sistema para este segundo caso. Se ha demostrado que existen notables diferencias en términos de distribución de probabilidad asociada con la SNR por usuario y por sistema. Por último, se ha abordado la gestión de recursos para una arquitectura de Red de Acceso Radio en la nube (Cloud-RAN). En particular, se ha analizado el impacto del retardo en el informe H-ARQ sobre el rendimiento del usuario.

En cuanto a las técnicas de gestión de interferencias, se ha comenzado por el análisis del rendimiento del sistema en un despliegue celular basado en Reutilización Fraccional de Frecuencias (FFR) cuando se utilizan esquemas de asignación de recursos diferentes. Se ha propuesto un método para determinar la tasa total de transmisión y el ancho de banda óptimo de partición de cada celda bajo cada estrategia de planificación de recursos. Además, se han analizado las técnicas coordinadas para la gestión de interferencias cuando se aplica desde varias puntos de acceso. En particular, se ha evaluado la técnica de Transmisión Conjunta (JT) sin precodifi-

ción y se ha comparado con la técnica de Reutilización Parcial de Frecuencias (PFR) para la transmisión de tráfico entre máquinas, mostrando que la técnica PFR presenta mejor resultado en términos de retardo y tasa media de transmisión para el tipo de tráfico considerado. Por último, se ha analizado la técnica de Expansión del Rango de la Celda (CRE), evaluando la influencia del valor del sesgo asociado a la técnica CRE sobre las prestaciones medias de una red LTE-A heterogénea considerando varias densidades de estaciones base.

Abstract

This dissertation is focused on the analysis of radio resource management techniques on LTE-Advanced cellular networks. In particular, two types of resource management techniques are addressed: first, the analysis focuses on radio resource scheduling algorithms; secondly, those techniques oriented to inter-cell interference management on cellular networks have been analyzed.

Starting with the radio resource scheduling, several algorithms that take into account delay requirements have been analyzed. In addition, the fairness among users for a Proportional Fair (PF) algorithm with signal-to-noise ratio (SNR) criterion was analyzed against the transmission rate criterion, obtaining closed-form expressions for the per user and system SNR distributions for the latter case. It has been shown that there are notable differences in terms of probability distribution associated with the per user and per system SNR. Finally, the resource management for a Cloud Radio Access Network (Cloud-RAN) has been addressed. In particular, the impact of the delay on the H-ARQ report on the user performance has been analyzed.

Regarding the interference management techniques, the analysis of the system performance in a cellular deployment based on Fractional Frequency Reuse (FFR) has been done when using different resource allocation schemes. A simple method has been proposed to determine the total transmission rate and the optimal partition bandwidth for each resource allocation strategy. In addition, coordinated techniques for interference management have been analyzed when applied from several access points simultaneously. In particular, Joint Transmission (JT) without precoding has been evaluated and compared with the Partial Frequency Reuse (PFR) technique for the case of Machine-to-Machine traffic type, showing that PFR technique outperforms JT in terms of delay and average transmission rate. Finally, the Cell Range

Expansion (CRE) technique has been studied, evaluating the influence of the value of the bias associated with the CRE technique on the average performance of a heterogeneous LTE-A network considering different base stations densities.

Capítulo 1

Introducción

1.1. Antecedentes y Motivación

1.1.1. Los sistemas celulares de Cuarta Generación

Las redes celulares están planificadas en forma de celdas (centradas alrededor de un punto de acceso) cuyo tamaño viene básicamente determinado por la potencia empleada en la comunicación y el entorno que le rodea. Tradicionalmente, los terminales móviles se conectan a la red estableciendo un enlace bidireccional que les conecta con el punto de acceso que les proporciona una mejor Relación Señal a Ruido más Interferencia (SINR), generalmente asociada al punto de acceso más cercano. En ocasiones, pueden existir varios puntos de acceso a los que sería posible conectarse, ya que la señal procedente de ambos se recibe con potencia similar (sobre todo en los bordes de las celdas). Este solape puede producir interferencias en el enlace establecido, que tradicionalmente se han reducido empleando bandas de frecuencia diferentes para celdas contiguas (reutilización relajada de frecuencias).

Las especificaciones de capa física de los últimos sistemas celulares (como LTE [1] y [2]) incluyen el uso de técnicas multiportadora basadas en multiplexación por división en frecuencias ortogonales (OFDM). Es conocido que el canal móvil presenta particularidades relativas a los desvanecimientos en espacio, tiempo y frecuencia, que es preciso manejar. La modulación multiportadora divide el ancho de banda disponible en bandas suficientemente estrechas para ser consideradas planas. De esta forma pueden emplearse técnicas de transmisión adaptativa que consideran las variaciones (en tiempo y frecuencia) de la SNR recibida para adaptarse a ellas mediante la modificación de la modulación, tasa de codificación y/o potencia instantánea en cada banda de frecuencias. OFDM también ha permitido planificar redes celulares en las que todos los puntos de acceso empleen una única banda de frecuencias. En estas redes de frecuencia única (SFN) la transmisión ocurre desde un conjunto de puntos de acceso sincronizados, lo que permite la combinación en el aire de las señales. El uso de redes de frecuencia única permite aumentar la capacidad del sistema, aunque requiere de técnicas complejas de cooperación entre puntos de acceso para minimizar las interferencias intercelda. Esta comunicación coordinada resulta especialmente beneficiosa para los terminales ubicados en el solape entre las áreas de cobertura de distintas celdas.

Además, los estándares desarrollados en los últimos años para redes celulares han ido incorporando el uso obligatorio de múltiples antenas en transmisión y recepción (MIMO) en sus especificaciones. El uso de múltiples antenas sobre redes celulares puede ser explotado tanto para reducir la tasa de error o para aumentar la velocidad de la transmisión [3].

Con el objetivo de seguir aumentando la velocidad de transmisión, recientemente se han planteado nuevas técnicas para explotar al máximo la diversidad espacial. Así, la conexión del terminal a la red se hace a través de más de un punto de acceso, en lo que se conoce como comunicaciones coordinadas. Por ejemplo, es posible hacer una transmisión sincronizada desde dos puntos de acceso que se combinan en el receptor o establecer dos transmisiones simultáneas hacia los dos puntos de acceso. Estas técnicas son especialmente útiles para terminales situados en los bordes de las celdas, donde la cobertura es peor [4].

Las técnicas de transmisión multipunto coordinada (CoMP) son una de las principales funciones definidas por el 3GPP para la tecnología celular LTE-Advanced con el objetivo de mejorar la cobertura, la eficiencia espectral y la velocidad de transmisión (throughput) en el borde de las celdas [5]. Para ello, los puntos de acceso necesitan estar sincronizados y coordinados, mediante el envío de señalización específica.

Este tipo de técnicas de transmisión coordinada es especialmente útil en redes heterogéneas (HetNets), comúnmente formadas por estaciones base que utilizan potencias de transmisión diferentes y/o múltiples tecnologías de acceso de radio [6]. Las redes heterogéneas constituyen un medio interesante para expandir la capacidad de la red móvil mediante la combinación de nodos de bajo consumo que dan servicios a celdas pequeñas (femtoceldas o picoceldas) con otros nodos de acceso que sirven a celdas grandes (macroceldas). La reutilización de las bandas de frecuencias en estas dos capas requiere una estrecha coordinación entre los nodos de acceso para lograr una adecuada gestión de las interferencias.

La transmisión multipunto coordinada añade diversidad espacial al disponer de varios canales de transmisión independientes hacia o desde el terminal. Pero es posible llevar más allá esta diversidad espacial cuando cada uno de los puntos de acceso está equipado con múltiples antenas. Este escenario puede verse como un sistema MIMO generalizado, compuesto por múltiples antenas distribuidas entre todos los terminales y nodos de acceso, donde la correlación entre caminos es variable en función de la distancia entre las antenas. La gestión de las interferencias en este sistema multiantena generalizado resulta verdaderamente compleja.

1.1.2. Planificación de recursos radio

Uno de los continuos retos en comunicaciones inalámbricas es mejorar el aprovechamiento de los recursos radioeléctricos compartiéndolos entre un mayor número de usuarios y ofreciéndoles un servicio de calidad adecuada. Actualmente, los sistemas inalámbricos ya permiten explotar la diversidad en espacio, tiempo y frecuencia para mejorar la capacidad y la cobertura mediante técnicas adaptativas de transmisión y multiplexación. Esto se ha conseguido mediante el uso de múltiples antenas, modulación adaptativa, técnicas multiportadora y el uso de un planificador de recursos radio adecuado.

Cuando se trabaja con canales móviles, la variabilidad del canal es el principal inconveniente con el que se ha de lidiar. No obstante, es posible aprovechar esta variabilidad para mejorar las prestaciones del sistema de comunicaciones. Esencialmente, el hecho de que el canal cambie permite esperar que, en diferentes condiciones, el canal pueda pasar de ser excepcionalmente malo a ser un canal apto para una buena comunicación. Y esto no sólo ocurrirá en la dimensión temporal, sino que se puede extender al dominio de la frecuencia y al dominio espacial.

El objetivo del planificador de recursos es conseguir el máximo aprovechamiento posible de los recursos radio tiempo-frecuencia, garantizando a su vez un acceso lo más equitativo posible a los recursos del sistema, independientemente de la posición del usuario dentro de la celda [7]. Es decir, se trata de evitar que los usuarios situados en el exterior de la celda perciban un servicio sensiblemente peor que el resto de usuarios.

Existen numerosos trabajos relacionados con la planificación de recursos (scheduling) para sistemas celulares basados en OFDM [8]. Otros trabajos se centran en la evaluación de algoritmos de planificación con múltiples antenas y OFDM, lo cual añade un grado más de libertad en la asignación de recursos [9]. También es responsabilidad del planificador garantizar ciertos requisitos de Calidad de Servicio (QoS) en términos de prioridad, velocidad de transmisión, retardo o probabilidad de pérdidas de paquetes. De esta manera se establece al mismo tiempo un reparto equitativo de los recursos del sistema [7][10].

El problema se complica para escenarios que utilicen transmisión multipunto coordinada. En este caso, la tarea de planificación de recursos (distribuida o centralizada) aún requiere un estudio pormenorizado dada la cantidad de variables en juego. Además, el análisis pormenorizado de la imparcialidad de los usuarios en algunos escenarios aún está pendiente en la literatura.

1.1.3. Gestión de interferencias

Las redes celulares de última generación han introducido otros métodos adicionales para combatir las interferencias en los bordes de las celdas, como es el uso de distintos patrones de frecuencia a lo largo de la celda o del ancho de banda de transmisión. Por ejemplo, la técnica de reutilización fraccional de frecuencias (FFR) se basa en la utilización de una reutilización 1 con baja potencia de transmisión a los usuarios más cercanos al nodo de acceso, mientras que se utiliza una reutilización mayor de 1 con mayor potencia a los usuarios que se encuentran en los bordes de las celdas [11].

Para las redes heterogéneas se está prestando especial interés a un conjunto de técnicas definidas en el 3GPP para LTE-Advanced para la coordinación de interferencias inter-celda mejorada (eICIC), que utiliza la potencia de transmisión en el dominio frecuencial y espacial para mitigar las interferencias [12] [13]. En esta línea se introduce el concepto de "subtramas casi vacías"(Almost blank subframe, ABS) [14], las cuales se envían con muy baja potencia al no transportar canales de datos. Esta funcionalidad permite reducir la cantidad de interferencias de las macroceldas sobre las picoceldas durante la subtramas ABS. Junto a la transmisión de subtramas ABS se emplea el concepto de Extensión de Rango de la Celda (CRE), que consiste en aumentar la potencia de transmisión de la picocelda de forma coordinada con la transmisión de subtramas ABS por parte de la macrocelda para aumentar el área de cobertura de la picocelda. En este campo existe aún numerosos aspectos por resolver, entre los que se encuentra la gran variación en el nivel de interferencia que un receptor recibe debido a dichas subtramas ABS o la sincronización entre macroceldas. Por tanto, la gestión de interferencias es un aspecto clave para mejorar el rendimiento de los sistemas inalámbricos coordinados y heterogéneos, y está siendo introducida en la mayor parte de los nuevos estándares de comunicaciones inalámbricas.

1.2. Objetivos de la Tesis

Esta tesis doctoral se centra en el estudio de técnicas relacionadas con la gestión eficiente de los recursos radio sobre sistemas inalámbricos adaptativos, coordinados y heterogéneos. Se analizarán distintas técnicas con el objetivo de mejorar la eficiencia espectral en los solapes entre las celdas y proporcionar una Calidad de Servicio (QoS) más justa entre usuarios. En todos los casos se definirán modelos refinados que incluyan imperfecciones y no idealidades de las técnicas estudiadas, con el fin de obtener unos resultados lo más realistas posible.

Este objetivo general puede concretarse en los siguientes objetivos específicos:

1. Planificación de recursos radio: se estudiarán mecanismos capaces de explotar la diversidad espacial y frecuencial disponible en redes inalámbricas multiusuario que emplean modulación y codificación adaptativa. En particular, se evaluarán distintas políticas adaptativas de asignación de recursos en escenarios en los que el tráfico tiene requisitos de retardo. Además, se analizará en detalle el rendimiento de algunos planificadores en cuanto a resultados de imparcialidad (*fairness*) entre usuarios. Por último, también se abordará la gestión de re-

cursos en escenarios en los que el planificador de plaquetes se encuentra en la nube, es decir, considerando una arquitectura de Red de Acceso Radio en la nube (Cloud-RAN). Se tendrán en consideración aspectos prácticos como restricciones de QoS o limitación en el canal de retorno.

2. Gestión de interferencias en redes heterogéneas: se analizarán las prestaciones de distintos esquemas de coordinación para redes heterogéneas (HetNets). Este análisis se enfocará en la coordinación entre los nodos de acceso para gestionar las interferencias producidas por la reutilización de frecuencias. Se analizarán las prestaciones de sistemas de comunicaciones móviles heterogéneos y coordinados al aplicar técnicas de gestión de interferencias entre macroceldas y pico/femptoceldas en el dominio del tiempo. Especial interés se prestará a la evaluación de técnicas denomidadas Coordinación de Interferencia Intercelda (eICIC) y, entre ellas, se analizarán las técnicas de reutilización fraccional de frecuencias (FFR), poniendo especial énfasis en la imparcialidad entre usuarios, así como en las técnicas de Expansión del Rango de la Celda (CRE). También se evaluarán esquemas de comunicación coordinada, en la que los terminales establecen enlaces con más de un punto de acceso. Se medirá la ganancia en capacidad/cobertura que estas técnicas permiten gracias a la gestión de las interferencias en los terminales móviles que se encuentran en las fronteras de las celdas. Se analizarán distintos esquemas de transmisión coordinada entre puntos de acceso para reducir las interferencias de los terminales que se encuentran en las fronteras de las celdas, y se buscarán métodos para reducir la información de señalización que las estaciones base deben compartir para poder cooperar. Además, se evaluará la eficiencia (en términos de cobertura, eficiencia espectral y velocidad de transmisión) de las técnicas propuestas en condiciones no ideales a partir de los mecanismos que ofrece el estándar y se determinarán sus limitaciones prácticas.

Todas estas técnicas serán evaluadas en un simulador a nivel de sistema para la tecnología LTE-Advanced, cuyo desarrollo también forma parte de esta tesis doctoral, con contribuciones contrastables.

1.3. Estructura de la Memoria

La presente memoria de tesis se ha presentado en la modalidad de "Tesis por compendio de publicaciones", cuya estructura de la memoria está definida por la Normativa de la Universidad de Málaga.

En el capítulo 1 se ha presentado la unidad temática de los trabajos presentados para conformar la tesis, incluyendo el estudio del estado de la cuestión y preliminares, así como los objetivos planteados.

El capítulo 2 contiene un resumen global de los resultados obtenidos y discusión de los mismos, incluyendo una descripción y resultados de las técnicas y funcionalidades evaluadas.

El capítulo 3 describe las principales conclusiones de este trabajo así como posibles líneas futuras de trabajo.

En el Apéndice A se incluye una copia de los trabajos que forman parte integrante de la tesis.

Capítulo 2

Resumen global de los resultados

2.1. Planificación de recursos radio

En esta tesis se han estudiado distintas técnicas de planificación dinámica de recursos (*scheduling*) orientadas a la calidad de servicio (QoS).

En primer lugar, se abordan escenarios en los que el retardo supone un requisito importante, como ocurre en las comunicaciones entre máquinas o en entornos celulares en los que la interactividad del servicio es alta. Se evaluará la ganancia de prestaciones (en términos de tasa de error, eficiencia espectral, velocidad de transmisión por usuario o retardo).

En segundo lugar, se analizará la justicia o imparcialidad (*fairness*) de algunos algoritmos de planificación de recursos entre los distintos usuarios del sistema. Se obtendrán métricas de rendimiento para evaluar el grado de parcialidad entre usuarios que experimentan diferentes calidades del canal o el grado de cumplimiento de la QoS.

En tercer y último lugar, se abordará un escenario de Red de Acceso Radio en la nube (Cloud-RAN), en el que el planificador de paquetes está ubicado en un elemento remoto. Se evaluará la pérdida de prestaciones debido al retardo en el interfaz entre dicho elemento y la estación base.

2.1.1. Planificadores de recursos radio con restricciones de retardo

Algunas estimaciones prevén que el mundo podría tener un billón de dispositivos de comunicaciones en la próxima década. Se espera que la mayoría de ellos sean inalámbricos y que un alto porcentaje no sea operado directamente por los humanos. Las comunicaciones entre máquinas tienen ciertas características que las hacen difíciles de acomodar en redes móviles. En particular, ciertos flujos de datos son muy sensibles al retardo.

Existen aplicaciones entre máquinas (*Machine-to-Machine*, M2M), como la detección y evitación de colisiones en vehículos, alarmas basadas en sensores y control remoto, etc. que requieren valores de latencia extremadamente bajos. Por ejemplo, este es el caso de aplicaciones de tipo videovigilancia, vehículos no tripulados o dispositivos que transportan cámaras de vídeo (generalmente robots) que son controlados de forma remota. En términos de latencia, el flujo de vídeo es más crítico que la señal de control.

Se están realizando esfuerzos para reducir dicho retardo en la Red de Acceso Radio (RAN) para las redes 5G. Sin embargo, hay poco trabajo identificado en la provisión de mejoras para aplicaciones M2M con restricciones de latencia. El uso de planificadores de recursos optimizados para la latencia es, por lo tanto, crucial para que estas aplicaciones con latencia restringida tengan asignados los recursos físicos necesarios para garantizar un funcionamiento correcto.

En un sistema multiusuario, una capa de Control de Acceso al Medio (MAC) optimizada debería asignar recursos radio a los usuarios de acuerdo con varios parámetros, incluidas las características de la fuente de tráfico, las necesidades de QoS [15] y la diversidad de frecuencia, tiempo y espacio del canal radio.

Cuando las fuentes de tráfico son de velocidad variable y sus requisitos de tasa de transmisión fluctúan de forma asíncrona para diferentes usuarios, la explotación de la diversidad multiusuario (ganancia estadística) permite que más usuarios se acomoden en el sistema. Un diseño de la capa MAC adaptable a las características cambiantes del tráfico y del canal, así como a los requisitos específicos de QoS, mejora el rendimiento del sistema al explotar los recursos radio de manera más eficiente [7].

Numerosos algoritmos de planificación radio sobre LTE se han propuesto en la literatura [16]. Entre ellos se encuentran enfoques heurísticos que tienen en cuenta el comportamiento de la fuente [17][18]. Las técnicas elegidas deben ser lo suficientemente flexibles como para acomodar las fuentes de tráfico tradicionales y la existencia de otras fuentes con diferentes requisitos de QoS.

En este trabajo, nos centramos en la evaluación de algoritmos de asignación de recursos para comunicaciones M2M sobre LTE. Los algoritmos analizados se han elegido teniendo en cuenta diferentes aspectos como: condiciones instantáneas del canal, requisitos de latencia o retransmisiones pendientes.

Descripción de los planificadores evaluados

Se han analizado 3 algoritmos de planificación radio que tienen en cuenta requisitos de retardo: *Opportunistic hard priority* [12][13], *Channel Dependent Earliest Deadline Due* (CD-EDD) [14] y *CD-EDD with postponed EDD term* [15].

a) *Opportunistic Hard Priority*

Este algoritmo aplica una prioridad a las transmisiones de flujos sensibles al retardo si se excede un cierto retardo máximo. El algoritmo establece la misma prioridad para todos los paquetes siempre que el retardo del paquete esté por debajo del umbral y establece una prioridad alta para los paquetes que excedan el umbral. Se asigna un umbral de retardo máximo D_t y un umbral de retardo D_b para cada flujo sensible al retardo. LTE define un retardo máximo para el interfaz radio de 80 ms para tráfico de VoIP [19]. El umbral de retardo del flujo debe elegirse para que sea más bajo que este retardo máximo, con un margen suficiente para que el planificador radio pueda servir paquetes que excedan este umbral antes de violar el retardo máximo. Los paquetes que excedan dicho retardo máximo se descartan. Inicialmente, los usuarios se ordenan cíclicamente según el tiempo de llegada a las colas de transmisión.

El funcionamiento del algoritmo es el siguiente:

1. Configurar los parámetros de retardo: por flujo, D_b, y máximo, D_t.
2. Configurar las prioridades para cada flujo de datos:
   - IF $(D_b - Retardo_{paquete}) < 0$ THEN se descarta el paquete

- ELSE IF $(D_b - Retardo_{paquete}) < D_t$ THEN prioridad = 1
- ELSE prioridad = 0

3. Ordenar la lista de flujos en base a su prioridad.

- Asignar un conjunto de Bloques de Recursos Físicos (PRBs) al flujo con la prioridad más alta. En caso de empate, se aplica un criterio Round Robin (RR).
- Quitar el flujo servido de la lista.
- Volver a ejecutar el punto 3 si hay más flujos esperando en la lista.

b) *Channel Dependent Earliest Deadline Due* (CD EDD)

La prioridad asignada por este esquema depende de dos componentes: el componente sensible al retardo (EDD) y el componente sensible al canal, que sigue un criterio de *Proportional Fair* (PF). El término PF sigue la expresión $\frac{T_k[n]}{R_k[n]}$, donde $T_k[n]$ es la tasa de transmisión del usuario k en el TTI n, $R_k[n]$ es la tasa media de transmisión del usuario k en el TTI n. El término EDD funciona de tal manera que los usuarios reciben prioridad a medida que el retardo del paquete a ser servido se acerca al retardo máximo. El término PF favorece a los terminales con buenas condiciones instantáneas del canal. La prioridad asignada al usuario k se calcula, por lo tanto, de acuerdo con la siguiente fórmula:

$$\frac{T_k[n]}{R_k[n]} \cdot \frac{W_k[n]}{D_k^b - W_k[n]} \tag{2.1}$$

donde $W_k[n]$ es el tiempo de espera en cola del paquete del usuario k (en TTIs), D_k^b es el retardo máximo admisible para el usuario k (en TTIs). A medida que el retardo del paquete aumenta, el término EDD ($\frac{W_k[n]}{D_k^b - W_k[n]}$) domina rápidamente la prioridad.

c) *CD-EDD with postponed EDD term*

Este esquema es una modificación del algoritmo anterior. El término PF ahora dominará la decisión siempre que el retardo del paquete esté lejos de superar el retardo máximo. Dicho enfoque se basa en la siguiente función de utilidad:

$$\frac{T_k[n]}{R_k[n]} \cdot \left(\frac{\max(0, \mathrm{W_k[n]} - \mathrm{D_k^t})}{D_k^b - W_k[n]} + 1 \right) \tag{2.2}$$

donde D_k^t es el umbral de retardo associado al usuario k. El término asociado al retardo proporcionará prioridad a los usuarios cuyos retardos de paquete sean mayores que este umbral (D_k^t). Si un paquete que espera en la cola ha excedido el retardo máximo, se descarta.

Resultados de simulación

Se ha elegido un algoritmo PF como algoritmo de referencia para evaluar la reducción de latencia lograda con otros algoritmos. Un resumen de los parámetros de simulación se muestra en la Tabla 2.1. La fuente de tráfico M2M utilizada corresponde a la generada por una cámara de videovigilancia IP [11] que transmite un flujo de video de control remoto sensible al retardo. Este tipo de tráfico se puede asignar al Identificador de Clase de Calidad (QCI) 7 en LTE [19]. Para este QCI, el máximo retardo de paquete permitido es de 100 ms. Los parámetros de retardo asociados a cada algoritmo se muestran en la Tabla 2.2.

En la Fig. 2.1 se muestran los resultados de simulación para los algoritmos propuestos: retardo medio, percentil 95 del retardo, tasa de pérdida de paquetes y tasa de transmisión por usuario.

Cuando el retardo del paquete excede el retardo máximo, los tres algoritmos estudiados descartan dicho paquete en lugar de incrementar el retardo de los paquetes restantes a ser atendidos, lo que podría ser más perjudicial para la QoS. El descarte de paquetes es más probable que ocurra cuando la SNR media es baja (0-5 dB). En estas condiciones, la tasa de transmisión por usuario es muy baja debido a la alta probabilidad de interrupción (*outage*) y la necesidad de utilizar esquemas de codificación robustos que aseguren la BLER objetivo. Cuando aumenta la SNR media, la tasa de pérdida de paquetes disminuye hasta hacerse nula en torno a los 15 dB.

Para los tres algoritmos propuestos, la tasa de transmisión es un poco menor que para el algoritmo de referencia (*baseline*), ya que estos algoritmos favorecen a los usuarios que experimentan altos retardos. En general, estos usuarios tienen peores

Tabla 2.1: Parámetros de simulación por defecto

LTE Parameter	**Value/Mode**
Channel model	Extended pedestrian A (TS 36.803)
Mobile Speed	4 km/h (pedestrian)
Channel Bandwidth	20 MHz
OFDM symbols per TTI	14
PRB size	12 subcarriers
Carrier Frequency	2.5 GHz
Modulation schemes	QPSK, 16QAM and 64QAM
Target BLER	10%
ACK feedback delay	8 ms
CQI delay	2 ms
Nº of CQI bits	4
Max. Number of HARQ retransmissions	8
Number of parallel HARQ processes	8
MIMO mode	2x2 1 layer spatial multiplexing (Beamforming)
Codebook	TS 36.211
Channel Estimation	Non-ideal Zhao
MIMO detection	ZF
Noise power estimation	Error based
Signalling overhead	2/21
Number of users	10
Simulation length	20s
Type of traffic	M2M

Tabla 2.2: Parámetros de los algoritmos de planificación radio

Opportunistic Hard Priority		**CD-EDD**		**CD-EDD with postponed EDD term**	
D^b	100 ms	D^b	100ms	D^b	100 ms
D^t	50 ms			D^t	50 ms

condiciones de canal, por lo que su rendimiento es, en consecuencia, menor.

Para el algoritmo *Opportunistic Hard Priority*, el rendimiento cuando aumenta la SNR media es similar a un algoritmo RR, es decir, hay un número menor de

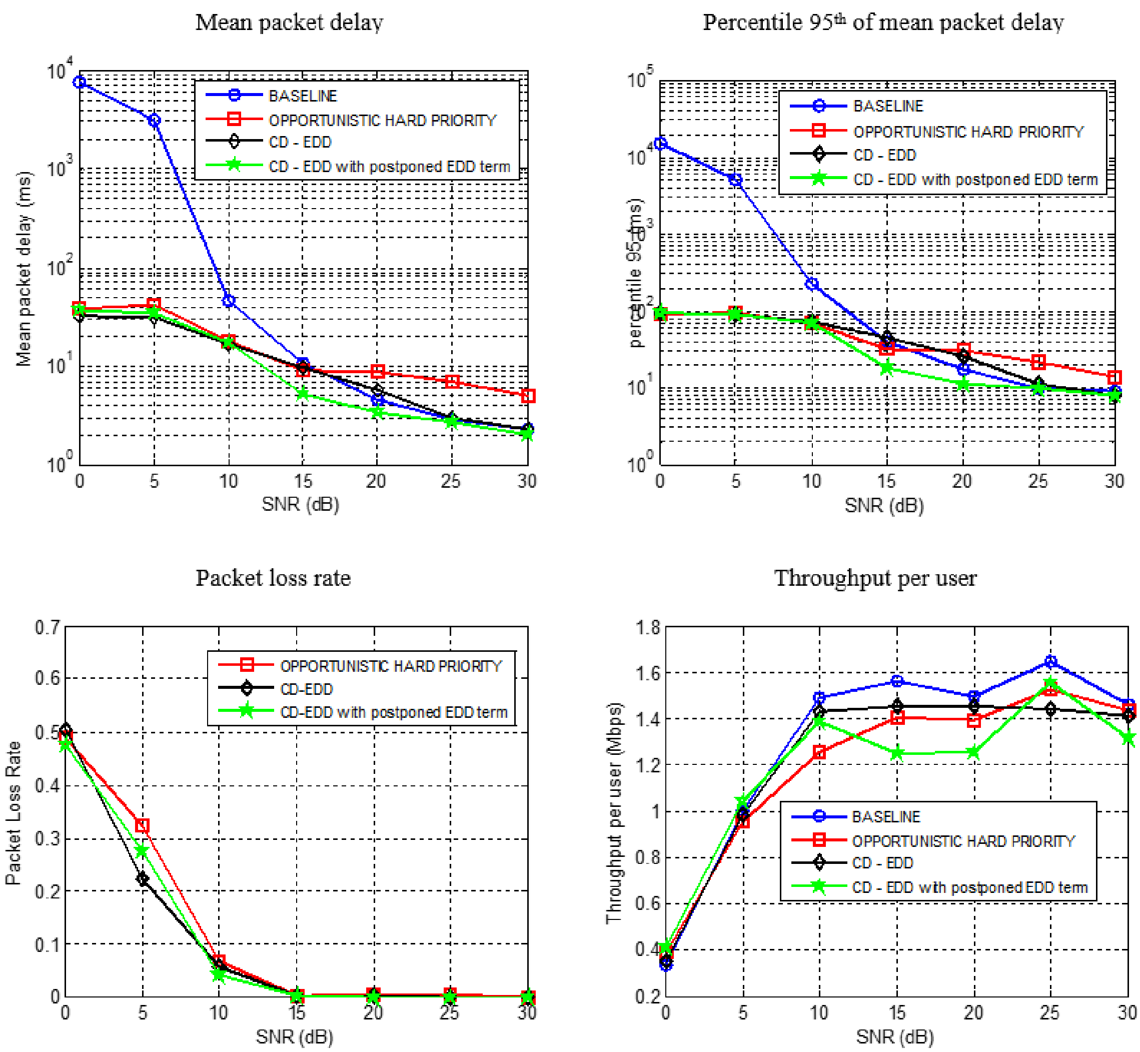


Figura 2.1: Retardo medio, percentil 95 del retardo, tasa de pérdida de paquetes y tasa de transmisión por usuario para los algoritmos evaluados

paquetes cuyo tiempo de espera excede el umbral de retardo; por lo tanto, el algoritmo establecerá la misma prioridad para casi todos los flujos de datos, que se asignarán en un orden cíclico. Esta es la razón por la que el algoritmo de referencia (PF) logra un mejor rendimiento para valores altos de SNR.

Para el algoritmo *Channel Dependent Earliest Deadline Due (CD EDD)*, el retardo medio para niveles altos de SNR obtiene valores ligeramente más altos que para el algoritmo de referencia. Esto se debe a que los retardos bajos en tal escenario y el término EDD otorga una prioridad bastante baja, lo que degrada el rendimiento en comparación con el caso en el que el término PF domina la decisión. Sin embargo, el algoritmo CD-EDD mejora al *Opportunistic Hard Priority*.

Por último, el algoritmo *CD-EDD with postponed EDD term* mejora los resultados de retardo para todo el rango de valores de SNR. A medida que aumenta el valor de la SNR, los retardos disminuyen, por lo que la probabilidad de exceder el valor D_t también disminuye. En esta situación, el algoritmo CD-EDD funciona como un algoritmo PF. Cuando el retardo está por encima de D_t, el término EDD otorga mayor prioridad a ese usuario, lo que disminuye los resultados de retardo medio de paquete. Por lo tanto, los valores de SNR no conducen a prioridades más bajas, como ocurre con el CD-EDD.

2.1.2. Análisis de la imparcialidad entre usuarios

La capacidad instantánea sobre canales inalámbricos cambia aleatoriamente con el tiempo debido a los desvanecimientos de la señal recibida. En ese sentido, es ya conocido que el algoritmo de Mejor Canal (*Best Channel*, BC) alcanza la máxima capacidad del sistema a costa de degradar la imparcialidad entre usuarios [7].

Con el objetivo de mejorar la imparcialidad entre usuarios, usualmente se utiliza una modificación de la estrategia anterior denominada Proporcional Fair (PF), la cual asigna recursos al usuario con las mejores condiciones de canal relativas a su propia media. La función de utilidad a maximizar se definió inicialmente utilizando un criterio de tasa de transmisión [20][21][22][23], donde los recursos se asignan al usuario con la mayor velocidad media de transmisión.

Sin embargo, muchos otros trabajos utilizan un criterio de SNR, mediante el cual los recursos se asignan al usuario con la SNR instantánea ponderada más grande, ya que es más sencillo de tratar matemáticamente, lo que hace posible evaluar la tasa de transmisión por usuario [24] y a nivel de sistema [25]. Por ejemplo, en [26] se presenta un modelo analítico para la distribución de SNR de los usuarios con una estrategia PF cuyo criterio de asignación está basado en SNR. En [27] se analiza la estrategia PF con información de realimentación parcial asumiendo un criterio basado en SNR. El mismo enfoque se ha asumido en muchos otros trabajos, como en [28] [29] para la estrategia PF con múltiples antenas, o en [30], que propone un nuevo algoritmo PF híbrido que incluye un nuevo método de agrupación de usuarios. En [31] se muestra un enfoque semi-analítico para modelar la interferencia del enlace ascendente considerando un algoritmo PF con criterio basado en SNR.

Puesto que ambos criterios (SNR y tasa de transmisión) utilizan la misma idea [25], se puede inferir que la versión del algoritmo PF basada en SNR proporciona resultados similares a los proporcionados por el que utiliza la tasa de transmisión. Sin embargo, hasta donde sabemos, el rendimiento del criterio basado en la tasa de transmisión aún no se ha analizado en detalle. En un trabajo reciente, [32] presenta expresiones cerradas para el rendimiento del algoritmo PF en escenarios con interferencia considerando la SNR como criterio, aunque señala que, cuando la tasa de transmisión se utiliza como métrica, se debería tener en cuenta la correspondencia entre la SINR y la velocidad del sistema.

En este apartado se analiza en detalle el rendimiento del criterio PF basado en la tasa de transmisión definida en [7]. Las principales contribuciones de este apartado se resumen a continuación:

- Se ha obtenido la función densidad de probabilidad (pdf) de la SNR del sistema y de la SNR por usuario considerando un canal de Rayleigh ergódico.

- Los resultados se han comparado con los obtenidos en [24] y [25] para el criterio basado en SNR con objeto de evaluar si es apropiado o no asumir un comportamiento equivalente en ambos esquemas. Se ha demostrado que el tiempo de acceso al canal para diferentes usuarios es similar para el criterio basado en la SNR, mientras que difiere para el criterio basado en la tasa de transmisión.

- También se ha comprobado analíticamente que para el criterio PF basado en SNR, la pdf de la SNR por usuario solo depende de la SNR promedio de un usuario en particular, mientras que en el criterio basado en tasa de transmisión, no solo depende de la SNR promedio del usuario, sino también de la SNR promedio del resto de usuarios en el sistema.

En el análisis se ha considerado un sistema con una única celda en la que L usuarios (UEs) se comunican con una estación base (BS). Se asume que tanto la

BS como los UEs disponen de una única antena. Se considera la existencia de ruido aditivo blanco gaussiano (AWGN) complejo de media cero y variaza σ_i^2, y modelo de desvanecimiento Rayleigh para la propagación multicamino.

A continuación se muestran las expresiones analíticas resultantes para ambos algoritmos, cuyo desarrollo pormenorizado puede encontrarse en el artículo [33] adjuntado en el Anexo A.

PF con criterio basado en SNR

Considerando γ_i la SNR instantánea, la función de utilidad en un instante t para el algoritmo PF con criterio basado en SNR viene dada por:

$$u_i^{SNR}(t) = \gamma_i(t)/\overline{\gamma}_i \tag{2.3}$$

Las pdfs resultantes para la SNR por usuario y del sistema vienen dadas por las siguientes expresiones [24] [25], respectivamente:

$$\begin{aligned} f_{\gamma_i^*}^{SNR}(\gamma) = Pr(u_i^{SNR} < u_{-i}^{SNR})\delta(\gamma) + \\ \frac{1}{\overline{\gamma}_i}\sum_{k=0}^{L-1}(-1)^k \begin{pmatrix} L-1 \\ k \end{pmatrix} \exp\left(-(1+k)\frac{\gamma}{\overline{\gamma}_i}\right) \end{aligned} \tag{2.4}$$

$$f_{\gamma_s}^{SNR}(\gamma) = \sum_{i=1}^{L}\frac{1}{\overline{\gamma}_i} \cdot \sum_{k=0}^{L-1}(-1)^k \begin{pmatrix} L-1 \\ k \end{pmatrix} \exp\left(-(1+k)\frac{\gamma}{\overline{\gamma}_i}\right) \tag{2.5}$$

donde $\delta(\gamma)$ es la función Delta de Dirac.

PF con criterio basado en tasa de transmisión

En este caso, la función de utilidad para el usuario i en el instante t viene dada por:

$$u_i^{RATE}(t) = r_i(t)/\overline{r}_i \tag{2.6}$$

donde $r_i(t)$ y $\overline{r}_i$ representan la tasa de transmisión potencial en el instante t y la capacidad ergódica del usuario i, respectivamente.

Las pdfs resultantes para la SNR por usuario y del sistema vienen dadas por las siguientes expresiones [33], respectivamente:

$$\begin{aligned} f_{\gamma_i^*}^{RATE}(r) = \; & Pr(u_i^{RATE} < u_{-i}^{RATE})\delta(\gamma) + \\ & \tfrac{1}{\overline{\gamma}_i} \exp\left(-\tfrac{\gamma}{\overline{\gamma}_i}\right) \cdot \prod_{\substack{k=1 \\ k\neq i}}^{L} \left(1 - \exp\left(-\frac{\left((1+\gamma)^{\frac{\overline{r}_k}{\overline{r}_i}} - 1\right)}{\overline{\gamma}_k}\right)\right) \end{aligned} \tag{2.7}$$

$$f_{\gamma_s}^{RATE}(\gamma) = \sum_{i=1}^{L} \frac{1}{\overline{\gamma}_i} \exp\left(-\frac{\gamma}{\overline{\gamma}_i}\right) \cdot \prod_{\substack{k=1 \\ k\neq i}}^{L} \left(1 - \exp\left(-\frac{\left((1+\gamma)^{\frac{\overline{r}_k}{\overline{r}_i}} - 1\right)}{\overline{\gamma}_k}\right)\right) \tag{2.8}$$

La primera diferencia importante entre ambos criterios se refleja atendiendo a la definición de su función de utilidad. Cuando se asume la estrategia basada en SNR, todos los usuarios son idénticos en un sistema con desvanecimiento de Rayleigh, por lo que se asigna el mismo porcentaje de tiempo de acceso al canal para cada usuario. Debe tenerse en cuenta que no significa que todos los usuarios alcancen la misma velocidad de transmisión, ya que la utilización del canal depende de la SNR media asociada con cada usuario. Por el contrario, con el criterio basado en tasa de transmisión no ocurre esto, por lo que se asigna un porcentaje diferente de tiempo de acceso de canal a cada uno.

Además, en un criterio PF basado en SNR, la pdf de la SNR por usuario (ecuación (2.4)) solo depende de la SNR promedio del usuario y del número de usuarios. Sin embargo, cuando se usa un criterio de PF basado en tasa de transmisión, esta distribución (ecuación (2.7)) no solo depende de la SNR promedio del usuario individual (y la capacidad ergódica) sino también de la SNR promedio del resto de usuarios del sistema. Al analizar las ecuaciones (2.4) y (2.7), se observa que ambos criterios son solo equivalentes cuando todos los usuarios experimentan el mismo valor de SNR promedio y, en este caso, ambos se comportan como la mejor estrategia de canal.

A continuación se muestran una comparativa entre ambos algoritmos, considerando un escenario con 5 usuarios, cuya SNR media, $\overline{\gamma}_i$, se ha obtenido mediante una distribución lognormal de media Γ (que representa la SNR media de la celda) y desviación estándar σ de 4 dB (valor típico para microceldas [34]). En particular, los usuario {1, 2, 3, 4, 5} se corresponden con el percentil {10th, 30th, 50th, 70th, 90th}. De esta forma, el usuario 1 está asociado con la SNR media más baja.

La Fig. 2.2 muestra la tasa de transmisión alcanzable por usuario, R_i, como una función de la SNR media de la celda, Γ, en un escenario de 5 usuarios.

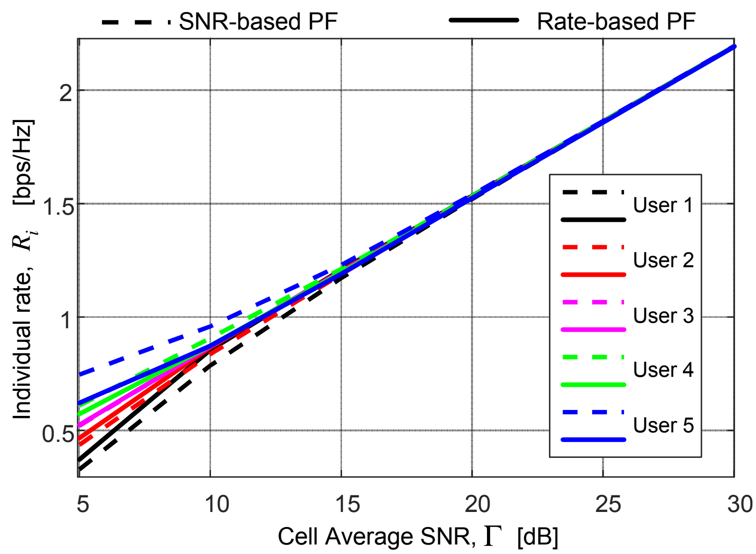


Figura 2.2: Tasa media de transmisión por usuario

Se puede observar que la imparcialidad (en términos de tasa de transmisión) es más justa para valores bajos de Γ. Si se usa el criterio PF basado en SNR, el rendimiento por usuario para el conjunto de usuarios es siempre el mismo para cualquier valor de Γ: una SNR media $\overline{\gamma}_i$ más alta representa tasas asignadas más altas. Sin embargo, cuando se aplica el criterio de PF basado en tasa de transmisión, el usuario con la SNR media más alta puede no ser el usuario con las tasas asignadas más altas, pero depende del valor de la SNR promedio de la celda. Esta diferencia se debe al hecho de que el criterio PF basado en SNR asigna el mismo porcentaje de tiempo de acceso al canal a todos los usuarios [24] mientras que este porcentaje depende del escenario específico si se usa el criterio PF basado en la velocidad. Al analizar la expresión de utilidad, es fácil ver que el valor de utilidad se incrementa para los usuarios con tasas medias potenciales más bajas (es decir, SNR media más baja). Por lo tanto, se les asigna un mayor porcentaje de tiempo de acceso al canal.

Fig. 2.3 representa la evolución en el tiempo de acceso al canal por usuario (medido en %) en un escenario de 5 usuarios.

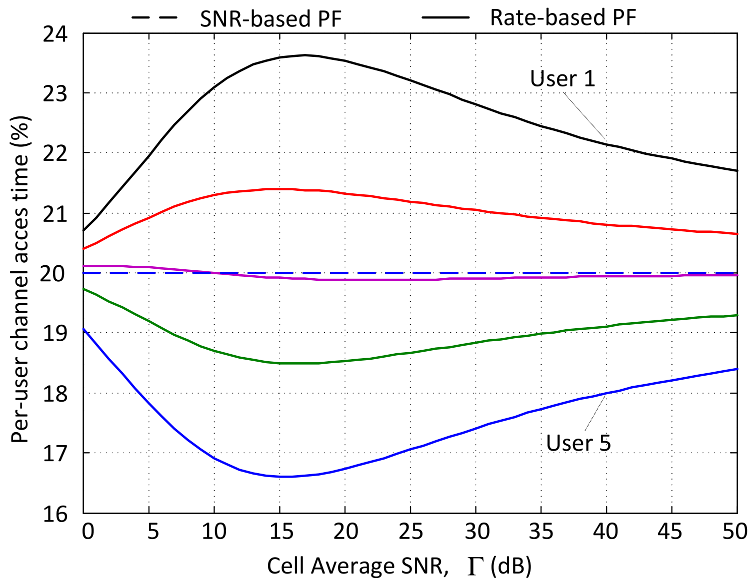


Figura 2.3: Tiempo de acceso al canal por usuario (%)

Se puede ver que el esquema PF basado en SNR es estríctamente justo en términos de tiempo de acceso al canal. De hecho, todos los usuarios reciben el mismo porcentaje de tiempo de acceso al canal (20 % en este caso), por lo que todas las curvas están superpuestas. Sin embargo, el porcentaje de tiempo de acceso al canal asociado a cada usuario es muy diferente cuando se utiliza un esquema PF basado en la tasa. Como se mencionó anteriormente, el tiempo de acceso al canal asignado a cada usuario depende de su SNR media, es decir, cuanto mayor sea la SNR media, menor será el porcentaje del tiempo de acceso del canal asignado. Tenga en cuenta que los resultados de tasa de transmisión individual que se muestran en la figura anterior ya incluye el impacto del tiempo de acceso de canal asignado a cada usuario. Cabe señalar que todas las curvas tienden a un valor constante a medida que aumenta la SNR media de la celda. La razón es que, a medida que aumenta la SNR media de la celda, la diferencia relativa entre el valor de R_i asociado a cada usuario es menor, ya que su desviación σ se mantiene constante. Como resultado, el porcentaje de tiempo de acceso al canal es similar para todos los usuarios. Sin embargo, debe notarse que si se usa un cálculo más realista de la tasa de transmisión potencial (por ejemplo, incluyendo el impacto de un conjunto limitado de esquemas de modulación), este efecto ocurre a valores más bajos de SNR media.

La Fig. 2.4 muestra la diferencia en la tasa de transmisión por usuario (en bps/Hz) entre el mejor y el peor usuario en función de la SNR media para diferentes números de usuarios en la celda (de 3 a 81 usuarios). Es decir, se representa el grado de imparcialidad entre usuarios, medido como: $\max(R_i) - \min(R_i)$. Se ha seleccionado el valor de la SNR media asociada con cada usuario, $\overline{\gamma}_i$, de modo que cada uno de ellos represente un percentil predefinido de la distribución de SNR media de la celda; en particular, los percentiles se establecen de acuerdo con los siguientes criterios: [10:40:90]th para 3 usuarios, [10:20:90]th para 5 usuarios, [10:10:90]th para 9 usuarios, [10:5:90]th para 17 usuarios y [10:1:90]th para 81 usuarios. Los resultados muestran que el criterio de PF basado en la tasa de transmisión siempre es más justo que el basado en SNR. Además, observamos diferencias mucho más altas entre el mejor y el peor usuario en valores bajos de SNR media. Finalmente, vale la pena señalar que la imparcialidad mejora a medida que aumenta el número de usuarios en la celda.

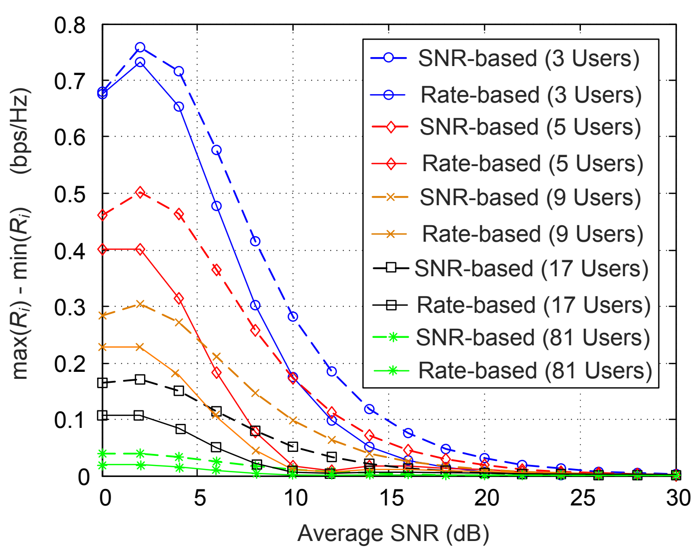


Figura 2.4: Diferencia en la tasa de transmisión por usuario (en bps/Hz) entre el mejor y el peor usuario en función de la SNR media

Se debe tener en cuenta que, aunque el criterio basado en la tasa de transmisión proporciona una mayor imparcialidad que el basado en la SNR, esto se logra a costa de degradar la tasa de transmisión total del sistema. Esto se confirma en la Fig. 2.5, que muestra la tasa de transmisión total (en bps/Hz) en función de la SNR media para un número diferente de usuarios en la celda (de 3 a 81 usuarios). Por un lado, los resultados muestran que un mayor número de usuarios proporciona una ganancia por diversidad multi-usuario. Por otro lado, se observa que los algoritmos proporcionan resultados similares solo para valores altos de SNR media; para valores bajos de SNR media, el criterio basado en tasa de transmisión alcanza una tasa total más baja (especialmente cuando el número de usuarios es alto) para mejorar la imparcialidad.

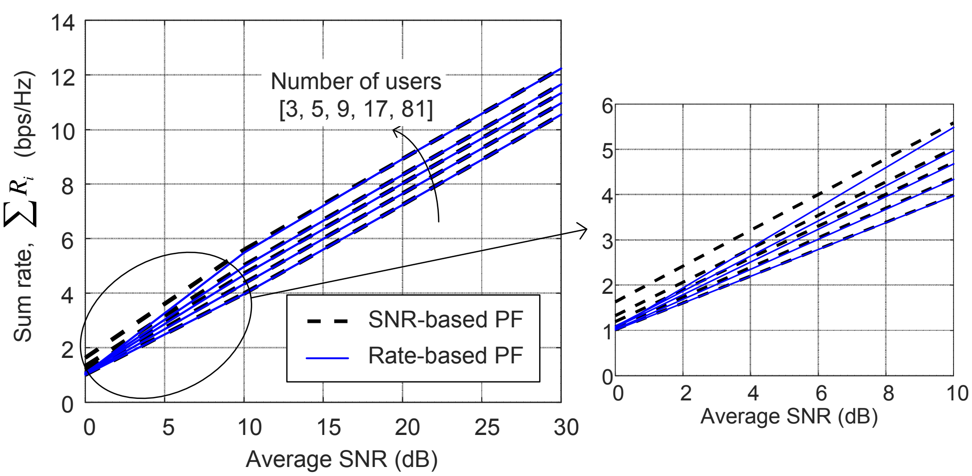


Figura 2.5: Tasa de transmisión total (en bps/Hz)

2.1.3. Gestión de recursos en la nube (CRAN)

Las redes móviles están evolucionando hacia despliegues ultra densos donde coexisten diferentes redes de acceso radio, como se muestra en la Fig. 2.6.

El diseño de la Red de Acceso de Radio (RAN) centralizada en 5G surge como una solución para gestionar eficazmente los recursos, lo que hace que algunas funciones se implementen en un hardware compartido. La recopilación de los recursos de procesamiento en centros de datos compartidos aumenta la reducción de los costes

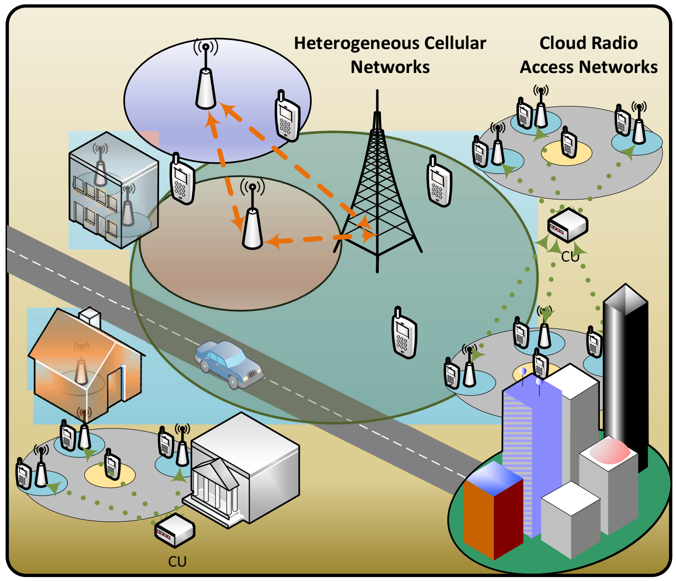


Figura 2.6: Ejemplo de red Ultra-Densa

de implementación y administración. Si esos recursos se ejecutan en infraestructuras virtualizadas ayudadas por las tecnologías de red definida por software (SDN) [35] y virtualización de la función de red (NFV) [36], la RAN centralizada se convierte en una RAN en la nube (*Cloud RAN*, CRAN) [37].

En el contexto de la CRAN, pueden coexistir varios métodos de centralización entre NodeBs evolucionados (eNBs). El 3GPP recomienda diferentes niveles de división funcional (ver Fig. 2.7) en los que una parte de la lógica está ubicada en la Unidad Radio Remota (RRU) mientras que el resto está ubicado en la Unidad de Banda Base centralizada (BBU) [38]. Esta centralización implica una nueva arquitectura de red en la que todo el procesamiento de banda base se realiza en centros de datos centralizados (BBU) y las señales radio se intercambian con RRUs a través de conexiones de fibra óptica de baja latencia (Fig. 2.8). Existen distintas opciones para realizar dicha división funcional. Por ejemplo, la opción 6 [38] centraliza las funcionalidades hasta la capa de control de acceso al medio (MAC) en el lado BBU [39], dejando solo la capa física en las RRU. En términos de latencia, los requisitos de la conexión de ida y vuelta son críticos [40]. Dado que el despliegue de fibra óptica es caro, se han considerado otras soluciones para facilitar el acceso a la RAN centralizada.

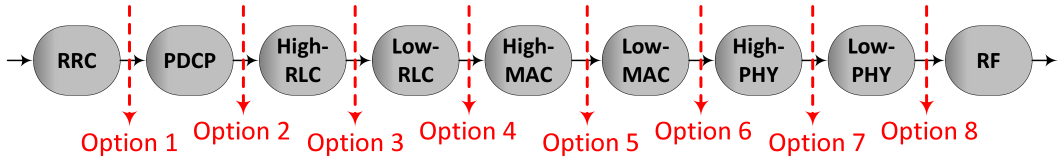


Figura 2.7: Propuesta de división funcional (3GPP)

En este apartado se analiza el impacto del retardo debido a conexiones no ideales entre la BBU y RRU sobre el rendimiento del usuario, teniendo especial interés en los procesos HARQ de la capa física.

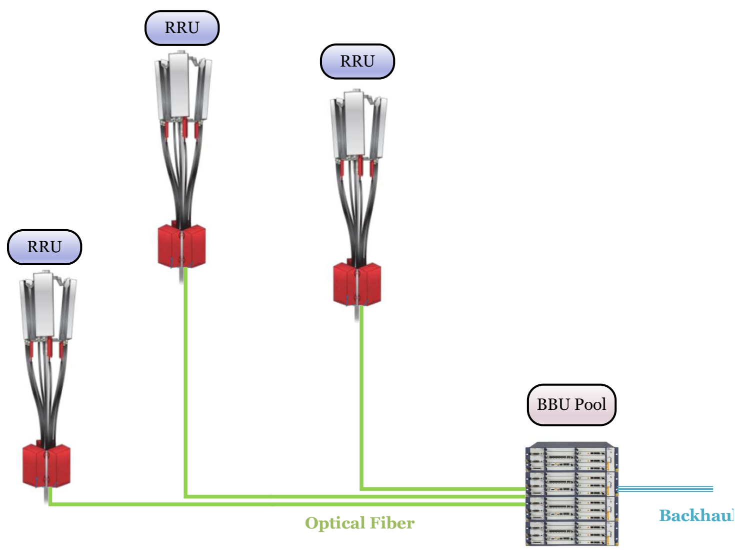


Figura 2.8: Ejemplo de RAN centralizada

Modelo de sistema

Se ha considerado la opción 6 de división funcional [38], en la que la capa MAC está centralizada, mientras que las capas inferiores están ubicadas en la RRU. Los datos del usuario se encapsulan y codifican en Bloques de Transporte (TB) cuyos tamaños dependen del Esquema de Modulación y Codificación (MCS) y del número de Bloques de Recursos Físicos (PRB) asignados por el planificador de recursos radio. Además, la modulación y codificación adaptativa (AMC) junto con la planificación de PRBs hacen que el MCS cambie con el tiempo. Los usuarios informan periódicamente sobre su indicador de calidad de canal (CQI) para realizar un seguimiento de la SINR recibida. El tamaño de TB seleccionado depende del índice MCS finalmente asignado (es decir, la calidad del canal) y del número de PRBs asignados. El tamaño promedio de TB para un determinado usuario se denomina $\overline{\mathrm{TB}_{\mathrm{tamaño}}}$.

Además, el tiempo medio entre dos asignaciones de recursos consecutivas a un usuario determinado se denomina t_{TB}. El planificador asigna recursos a un usuario con una periodicidad máxima de un Intervalo de Tiempo de Transmisión (TTI), es decir, $t_{\mathrm{TB}} \geq 1ms$. Se debe tener en cuenta que, desde el punto de vista de la asignación de recursos radio, el planificador solo tendrá en cuenta a un usuario si hay datos para transmitir. Por lo tanto, el intercambio de recursos no solo depende de los recursos disponibles sino también del patrón de tráfico.

La funcionalidad HARQ [41] utiliza un protocolo de parada y espera: cuando se envía un TB, la entidad transmisora espera hasta que se recibe un ACK o NACK. La recepción de un TB erróneo conduciría a la retransmisión de dicho TB hasta que el paquete se reciba correctamente o se alcance el número máximo de retransmisiones. Para optimizar la velocidad de transmisión, generalmente se emplean varios procesos HARQ paralelos; cada proceso HARQ envía un TB cuando los recursos se asignan al usuario y espera su confirmación asociada. Este procedimiento se repite hasta que se alcanza el número máximo de procesos HARQ, W, después de $W \cdot t_{\mathrm{TB}}$ segundos. En el caso de LTE FDD, hay un máximo de $W = 8$ procesos HARQ paralelos [42].

Tasa de transmisión neta (*Net Rate*)

Definimos el concepto de *Net Rate* (NR) como la tasa de transmisión que un usuario determinado puede lograr teniendo en cuenta las condiciones reales de la transmisión, definido como:

$$\mathrm{NR} = \frac{\overline{\mathrm{TB}_{\mathrm{size}}} \cdot W}{t_W}, \tag{2.9}$$

donde W es la ventana de transmisión (es decir, el número de procesos HARQ) y t_W es el tiempo requerido para reutilizar un proceso HARQ, dado por el máximo entre el tiempo necesario para enviar de vuelta el ACK y el tiempo requerido para la ejecución de todos los procesos HARQ:

$$t_W = \max\left(\mathrm{t_{ACK}}, \mathrm{W} \cdot \mathrm{t_{TB}}\right). \tag{2.10}$$

Se debe tener en cuenta que si $t_{\mathrm{ACK}} > W \cdot t_{\mathrm{TB}}$, después de $W \cdot t_{\mathrm{TB}}$, el proceso HARQ estará esperando su llegada de ACK/NACK. Como consecuencia, los procesos HARQ asociados al usuario se bloquearán, lo que reducirá el *Net Rate*. Por otro lado, si $t_{\mathrm{ACK}} < W \cdot t_{\mathrm{TB}}$, el proceso HARQ estará listo para transmitir (o retransmitir) un TB en el próximo TTI. Se alcanzaría así el rendimiento máximo: $\overline{\mathrm{TB}_{\mathrm{size}}}/t_{\mathrm{TB}}$.

La Fig. 2.9 muestra este concepto gráficamente para dos ejemplos diferentes de t_{ACK}. El primero, $t_{\mathrm{ACK\text{-}1}}$ representa un ACK que se recibe antes de que se gestionen todos los procesos HARQ y, por lo tanto, la entidad HARQ puede transmitir sin más demora. Sin embargo, en el segundo caso, $t_{\mathrm{ACK\text{-}2}}$ muestra un ACK que se recibe después de $W \cdot t_{\mathrm{TB}}$, lo que provoca el bloqueo del proceso HARQ. Tenga en cuenta que el mínimo de t_{TB} definido para un sistema LTE es 1ms y, por lo tanto, el mínimo de t_W es 8 ms si se utilizan 8 procesos HARQ. Por este motivo, un $t_{\mathrm{ACK}} < 8$ ms no afectará al *Net Rate* para $W = 8$. De lo contrario, debe analizarse el umbral a partir del cual disminuye la velocidad de transmisión.

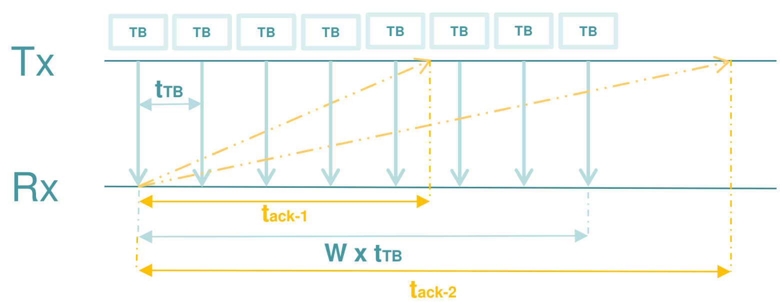


Figura 2.9: Transmisión HARQ con diferentes tiempos de recepción de ACKs

Una vez que se haya definido el *Net Rate*, el efecto de un mayor retardo en el informe podría evaluarse como un incremento en t_{ACK}. Sin embargo, el impacto de dicho retraso en el rendimiento del usuario también depende de la tasa de transmisión de la fuente (*Source Rate*, SR), es decir, la velocidad de tráfico de datos transmitida al usuario. Mientras que el *Net Rate* sea mayor que la velocidad de la fuente transmitida a un usuario determinado, su rendimiento no experimenta ninguna degradación. En conclusión, el retardo de HARQ-ACK solo tendrá un impacto negativo en la experiencia del usuario si se cumplen dos condiciones: $t_{\mathrm{ACK}} > W t_{\mathrm{TB}}$ y $\mathrm{NR} < \mathrm{SR}$. El umbral de retardo ($t_{\mathrm{W,th}}$) a partir del cual se degrada el rendimiento se puede evaluar como:

$$t_{W,th} = \frac{\overline{\mathrm{TB}_{\mathrm{size}}} W}{SR} \tag{2.11}$$

Mientras el retardo del ACK t_W sea inferior al umbral $t_{\mathrm{W,th}}$, el rendimiento del usuario no se verá afectado. De lo contrario, el rendimiento del usuario disminuirá por debajo de SR.

Resultados de simulación

Se ha evaluado el enlace descendente de un eNB con un ancho de banda de 10 MHz (equivalente a 50 PRB) y 6 usuarios en la celda, numerados como 1 - 6 en distancias crecientes a la RRU. Los valores medios de SINR para esos usuarios son 27.8 dB, 22.5 dB, 18 dB, 13.3 dB, 9.7 dB y 4 dB. El promedio de los índices CQI reportados es de 15, 13, 11, 9, 7 y 5, respectivamente. Se ha considerado el modelo de canal Macro Urbano de la UIT (UMa) [43] con una velocidad de usuario de 4 km/h. Hemos considerado un algoritmo de adaptación de enlace de bucle externo (OLLA) [44] para mejorar la selección de MCS en cada TTI. Por lo tanto, el índice MCS instantáneo de cada usuario, así como su tamaño de bloque de transporte, varía a lo largo de la simulación de acuerdo con la SINR instantánea. Los tamaños de bloque de transporte se promedian para cada usuario de forma independiente para obtener su $\overline{\mathrm{TB}_{\mathrm{size}}}$, que es más grande para usuarios con mejor calidad del canal. Las simulaciones se llevan a cabo a lo largo de 20000 subtramas. Se ha asumido un planificador de paquetes Round Robin (RR) [45], configurado con un tamaño de asignación por usuario y subtrama de 15 PRB; es decir, se pueden programar 3 usuarios por subtrama, cada usuario planificado después de dos subtramas, $t_{\ rmTB} = 2ms$, si todos los buffers contienen datos.

Se han simulado un conjunto de retardos para el HARQ-ACK: 1, 4, 8, 16 y 24 ms. Además, se ha estudiado la influencia de la tasa de fuente. Primero, asumimos una velocidad de fuente de transmisión de 6 Mbps, caracterizada por paquetes de 1400 bytes cada 1.9 ms. También se han simulado tasas de fuente más bajas de 300 kbps, 950 kbps y 2.6 Mbps (enviando ráfagas de datos cada 2 segundos). Los principales parámetros de simulación se resumen en la Tabla 2.3.

En primer lugar se evalúa el *Net Rate* para cada usuario a partir de la Ecuación (2.9). El efecto de la ventana de transmisión, es decir, el número de procesos HARQ,

Tabla 2.3: Parámetros de simulación por defecto

Parameter	Value
Carrier Frequency	2 GHz
Bandwidth	10 MHz
Transmission Power	43 dBm
Simulation Time	2×10^4 ms
Transmission Mode	SISO
Channel Model	ITU UMa
Mobile speed	4 km/h
Number of users	6
Set of avg. SINRs per user	27,8 dB, 22,5 dB, 18 dB, 13,3 dB, 9,7 dB, 4 dB
Set of avg. CQIs per user	$15, 13, 11, 9, 7, 5$
Scheduling Policy	Round Robin
Max. allocation per user	15 PRBs
Source rate	6 Mbps, 2,6 Mbps, 950 kbps, 300 kbps
HARQ processes	8
No. of retransmissions	3
Decoding algorithm	SOVA
Link adaptation method	OLLA

se ilustra en la Fig. 2.10 y Fig. 2.11 para $t_{\mathrm{TB}} = 1$ ms y $t_{\mathrm{TB}} = 2$ ms, respectivamente. Se pueden observar dos regímenes operativos: 1) un régimen donde la *Net Rate* aumenta con W; y 2) un régimen donde la *Net Rate* es independiente de W. El primer régimen se produce cuando el retardo del ACK, t_{ACK}, es mayor que el tiempo entre dos intervalos de transmisión consecutivos, es decir, $t_{\mathrm{ACK}} > W \cdot t_{\mathrm{TB}}$. En este caso, si asumimos que no hay errores en la transmisión de datos, cada proceso HARQ se bloquea esperando su mensaje ACK durante $t_{\mathrm{ACK}} - W \cdot t_{\mathrm{TB}}$ segundos, lo que aumenta el tamaño de la ventana (es decir, la cantidad de procesos HARQ paralelos), mejorando el *Net Rate*. El segundo régimen operativo aparece cuando $t_{\mathrm{ACK}} < W \cdot t_{\mathrm{TB}}$. En este caso, el retardo del ACK es lo suficientemente pequeño como para evitar el bloqueo en los procesos HARQ. Por lo tanto, el *Net Rate* es máximo para cada usuario, y su valor es el cociente entre el tamaño promedio de TB (que depende de la calidad del canal del usuario) y el tiempo entre dos asignaciones de recursos consecutivas, t_{TB}.

Curiosamente, observamos que el valor de W que divide estas dos regiones tam-

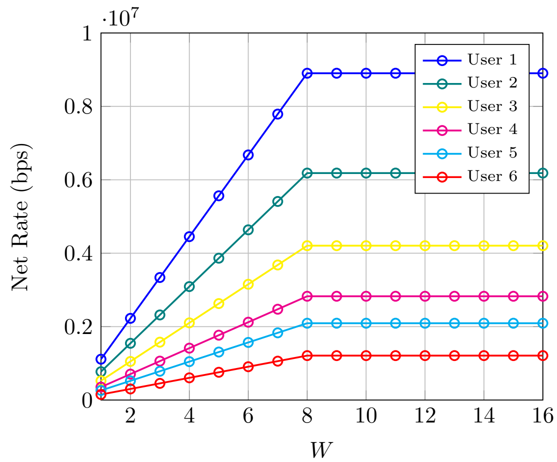


Figura 2.10: *Net Rate* (bps) para $t_{\text{TB}} = 1$ms, $t_{\text{ACK}} = 8$ ms

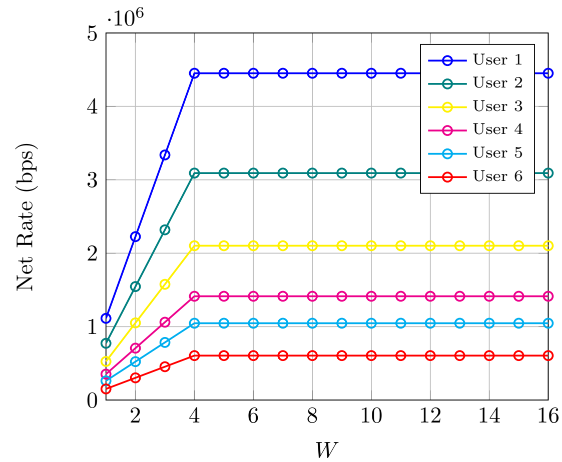


Figura 2.11: *Net Rate* (bps) para $t_{\text{TB}} = 2$ms, $t_{\text{ACK}} = 8$ ms

bién depende de t_{TB}. Recordemos que un valor mayor de t_{TB} significa que la transmisión hacia un usuario determinado ocurre con menos frecuencia, lo que implica que el régimen sin procesos HARQ bloqueados se alcanza con una ventana de transmisión más pequeña, W. Se observa que la región sin procesos bloqueados se alcanza con $W = 8$ para $t_{\text{TB}} = 1$ ms, mientras que se alcanza con solo $W = 4$ para $t_{\text{TB}} = 2$ ms.

Los umbrales de retardo teóricos $t_{W,th}$ según lo evaluado por la Ecuación (2.11) se dan en la Tabla 2.4 para diferentes tasas de fuente y $W = 8$. Esta tabla recopila los $t_{W,th}$ necesarios para lograr una *Net Rate* igual a la tasa de fuente, de modo que un $t_W > t_{W,th}$ causa un degradación del rendimiento del usuario. Bajo la configuración de simulación descrita anteriormente, a los usuarios con mejor calidad de canal se les asignan bloques de transporte más grandes, por lo que pueden resistir mayores retardos. Como se esperaba de la Ecuación (2.11), los requisitos de velocidad de fuente menos estrictos también permiten mayores retardos.

La Fig. 2.12 y 2.13 muestran el *Net Rate* en función del retardo de ACK para dos valores de t_{TB} diferentes. En el primer caso, $t_{\text{TB}} = 1ms$ y, en consecuencia, $W \cdot t_{\text{TB}} = 8ms$, por lo que t_{ACK} no afecta a los usuarios si $t_{\text{ACK}} \leq 8ms$. A partir del retardo de 8 ms, el *Net Rate* disminuye junto con t_{ACK}, siendo más notables para valores de transmisión más altos. Del mismo modo, los cálculos para $t_{\text{TB}} = 2ms$ ($W \cdot t_{\text{TB}} = 16ms$) se ilustran en la Fig. 2.13. Por un lado, el *Net Rate* inicial es la

Tabla 2.4: Umbrales de retardo para usuarios con diferentes tasa de fuente, de acuerdo a la Ecuación (2.11) con $W = 8$ and $t_{\mathrm{TB}} = 2ms$

User	6 Mbps	300 kbps	2.6 Mbps	950 kbps
1 (best)	11.6 ms	246.1 ms	26.8 ms	75.5 ms
2	8.1 ms	175.6 ms	18.7 ms	50.7 ms
3	5.6 ms	110.9 ms	12.8 ms	36.1 ms
4	3.8 ms	75.2 ms	8.8 ms	23.5 ms
5	2.7 ms	56.5 ms	6.4 ms	17.8 ms
6 (worst)	1.6 ms	33.0 ms	3.7 ms	9.9 ms

mitad que el mostrado en la Fig. 2.12 debido a que el tiempo entre la asignación de dos recursos se ha duplicado ($t_{\mathrm{TB}} = 2ms$). Por otro lado, el retardo en el informe requerido para tener un impacto negativo en el *Net Rate* también se duplica.

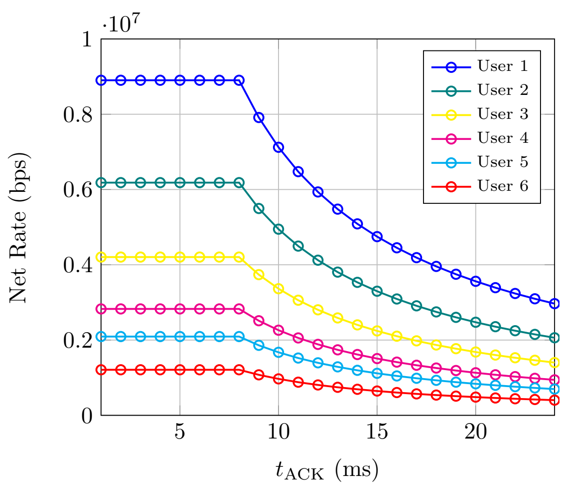


Figura 2.12: *Net Rate* (bps) para $t_{TB} = 1$ms, $W = 8$

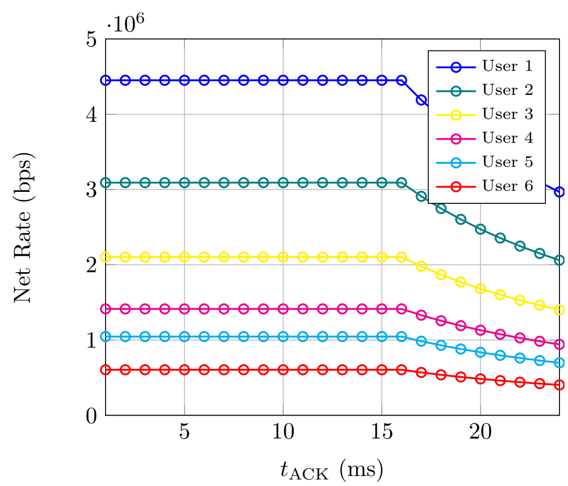


Figura 2.13: *Net Rate* (bps) para $t_{TB} = 2$ms, $W = 8$

Una vez que se han obtenido los resultados teóricos, las siguientes figuras analizan los resultados de simulación para las fuentes descritas anteriormente. Las figuras muestran el rendimiento por usuario a medida que aumenta el retardo en el recorrido de ida ($\mathrm{t_{ACK}}$) debido a una conexión no ideal. Los resultados son diferentes de los anteriores, ya que en este caso el planificador solo asigna recursos a aquellos usuarios que tienen datos en sus colas. Los usuarios capaces de transmitir con bloques de transporte más grandes vacían sus colas antes. Los recursos libres son empleados posteriormente por aquellos usuarios con peor calidad de canal, lo que reduce el tiempo entre asignaciones de recursos t_{TB}.

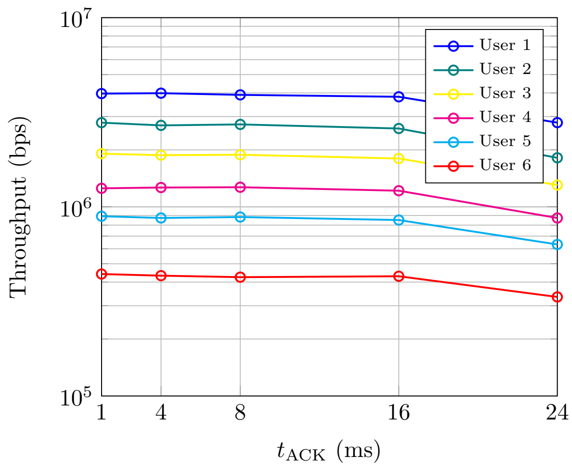


Figura 2.14: Tasa de transmisión por usuario, SR=6 Mbps

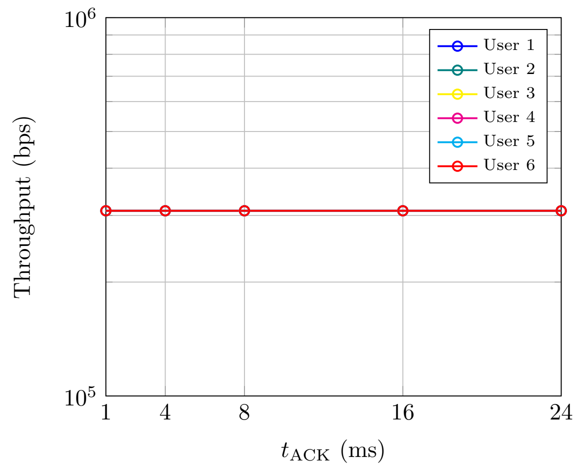


Figura 2.15: Tasa de transmisión por usuario, SR=300 kbps

La Fig. 2.14 muestra los resultados para la tasa de fuente más alta, SR = 6 Mbps. De la ecuación (2.10) es evidente que $t_W \approx \max(\mathrm{t_{ACK}}, 16\mathrm{ms})$. Los umbrales de retardo, según lo indicado por la Ecuación (2.11) (que se muestra en la Tabla 2.4), son inferiores a 16 ms para todos los usuarios. Por lo tanto, ningún usuario podrá alcanzar su tasa de fuente (es decir, NR $<$ SR para todos los usuarios) y, por lo tanto, el rendimiento experimentará una degradación significativa. Los usuarios 1 y 2 podrían mejorar su rendimiento si el planificador pudiera asignarles recursos a una tasa de subtrama ($t_{\mathrm{TB}} = 1ms$). Sin embargo, los usuarios 3-6 no pueden alcanzar la tasa de fuente debido al hecho de que $t_{W,th}$ está por debajo del mínimo de 8 ms impuesto por el sistema. Para ellos, incluso si t_{ACK} no influye en el rendimiento, la tasa obtenida es más baja que la tasa de fuente, ya que el canal no puede transportar todo el flujo de información de 6 Mbps debido a: 1) la tasa máxima alcanzable; 2) el hecho de que los recursos de transmisión se comparten entre los 6 usuarios; y 3) el efecto de las retransmisiones. Por lo tanto, se obtendrían los mismos resultados (en términos de rendimiento) si se utiliza una tasa de fuente más alta ya que el sistema está trabajando en condiciones de saturación.

Contrariamente al caso anterior, una tasa de fuente de 300 kbps es lo suficientemente lenta para garantizar que todos los usuarios cumplan que NR > SR, como se observa en la Fig. 2.15. Los valores calculados para los umbrales de retardo (en la Tabla 2.4) muestran que t_{ACK} debe ser demasiado alto para percibir sus efectos sobre el rendimiento del usuario (33 ms para el UE más débil en términos de calidad radio mientras que se necesitan 246.1 ms para los UEs que reciben una señal de buena calidad).

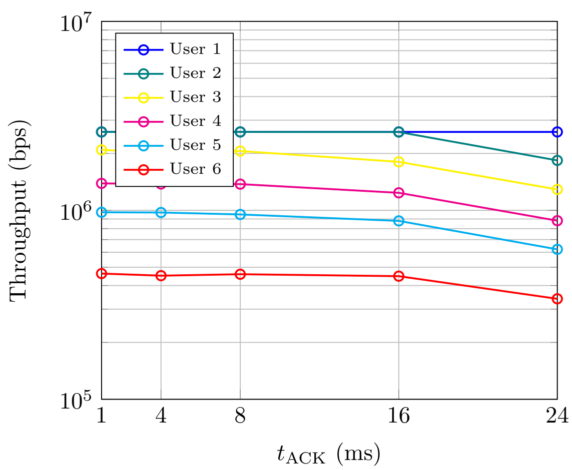


Figura 2.16: Tasa de transmisión por usuario, SR=2.6 Mbps

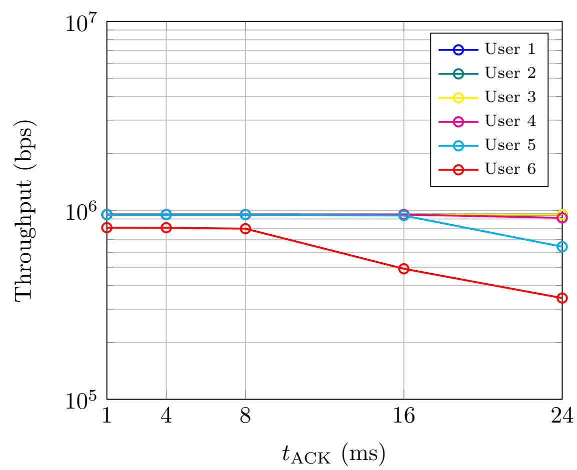


Figura 2.17: Tasa de transmisión por usuario, SR=950 kbps

Los resultados para la velocidad de fuente de 2.6 Mbps se muestran en la Fig. 2.16. Esta tasa de fuente es lo suficientemente alta para que todos los buffers de usuario tengan datos pendientes de transmisión. A este respecto, dado que el número máximo de usuarios programados en el mismo TTI es 3, a cada usuario se le asignan recursos cada 2 ms, lo que lleva a un aumento de los períodos de espera para cada paquete. Los usuarios 1 y 2 alcanzan la velocidad de fuente de 2.6 Mbps; sin embargo, el rendimiento de los usuarios de 3 a 6 es inferior a la tasa de fuente debido a que $NR < SR$. Este es un resultado que se puede contrastar con la Tabla 2.4, donde $t_{W,th}$ es inferior a 16 ms para dichos usuarios (12.8 ms, 8.8 ms, 6.4 ms y 3.7 ms, respectivamente). Los usuarios 4 y 5 podrían mejorar su rendimiento si el planificador pudiera asignarles recursos a ritmo de subtrama (por ejemplo, $t_{\mathrm{TB}} = 1$ ms); sin embargo, el valor de $t_{W,th}$ para los usuarios 4 y 5 está por debajo de 8 ms (6.8 ms y 3.7 ms), por lo que nunca alcanzarán un tasa de transmisión de 2.6 Mbps.

El usuario 1 no se ve afectado por retardos inferiores a 26.8 ms (Tabla 2.4), que está fuera del rango de la figura. Los usuarios 2 a 6 experimentan una degradación de su rendimiento cuando $t_{\text{ACK}} > 16$ ms, como se indica en la Tabla 2.10.

La Fig. 2.17 muestra los resultados para una tasa de fuente intermedia de 950 kbps por usuario. Los resultados permiten clasificar a los usuarios en tres grupos. Los tres mejores usuarios (etiquetados como 1-3) experimentan una calidad del canal lo suficientemente buena como para mantener su *Net Rate* por encima de la velocidad de la fuente para todos los retardos evaluados, es decir, NR > SR. En tal caso, $W \cdot t_{\text{TB}} \approx 16$ ms. Por lo tanto, t_{ACK} no tiene un impacto en su rendimiento hasta que alcanza 16 ms. A este valor, cualquier incremento de t_{ACK} reducirá su *Net Rate*. Como se muestra en la Tabla 2.4, el rendimiento de los usuarios 1-3 no se degrada hasta que sus retardos alcanzan los 75.5, 50.7 y 36.1 ms, respectivamente. En contraste, los usuarios 4 y 5 tienen una *Net Rate* más alta que la tasa de fuente NR > SR. Sin embargo, el rendimiento de estos usuarios disminuye a medida que t_{ACK} crece más de 23.5 y 17.8 ms, respectivamente. Esto se debe a que el valor de $t_{\text{ACK}} > W \cdot t_{\text{TB}}$ es lo suficientemente alto como para que NR < SR. Para el peor usuario, el tamaño promedio de bloque de transporte es tan bajo que cualquier paquete de capa superior se segmenta en un conjunto de TBs. Por lo tanto, cuando otros usuarios han finalizado su transmisión, este usuario todavía está enviando TBs. Es decir, su t_{TB} es aproximadamente 1 ms porque la mayor parte del tiempo se transmite solo en la celda, por lo que $W \cdot t_{\text{TB}} \approx 8$ ms. En consecuencia, su NR disminuye si $t_{\text{ACK}} > 8ms$. De acuerdo con la Tabla 2.4, su rendimiento disminuye cuando $t_{\text{ACK}} > 9{,}9ms$, puesto que, en ese caso, $NR < SR$.

2.2. Gestión de interferencias

Tradicionalmente, la gestión de interferencias se ha realizado en base a una planificación estática de frecuencias en cada una de las celdas desplegadas, siguiendo algún método que maximice la distancia entre celdas co-canal. Aunque este tipo de técnicas se ha ido empleando desde los mismos inicios de las redes celulares, sigue teniendo mucho potencial debido a que las técnicas coordinadas y las redes heterogéneas ofrecen nuevas posibilidades a este tipo de soluciones. Así, la combinación de técnicas de planificación estática con técnicas coordinadas puede relajar el número de celdas cooperantes. Al reducir la interferencia también se puede reducir la cantidad de información a intercambiar obteniendo las mismas prestaciones.

Tradicionalmente, la planificación estática ha considerado patrones regulares en la distribución espacial de los puntos de acceso. Si además se considera el caso de las redes heterogéneas en la que hay distintos niveles de puntos de acceso con distintas densidades de puntos y distintas potencias de transmisión, aparecen nuevos grados de libertad a la hora diseñar algoritmos de planificación que pueden ser explotados para aumentar las prestaciones.

En esta sección se estudiará el problema de la gestión de interferencias sobre redes heterogéneas, con especial énfasis en el análisis de técnicas de control de potencia en femtoceldas o picoceldas que comparten los recursos con macroceldas superpuestas (cuyos puntos de acceso suelen utilizar mayor potencia de transmisión). Se evaluarán distintas técnicas orientadas a la reducción de las interferencias provocadas por la celda agresora sobre los terminales conectados a la celda víctima.

Concretamente, el estudio comenzará por el análisis de técnicas de gestión coordinada de interferencias entre celdas (enhanced Inter-Cell Interference Coordination, eICIC) en el dominio del tiempo.

En primer lugar, se analizarán las técnicas de reutilización fraccional de frecuencias (FFR), poniendo especial énfasis en la imparcialidad entre usuarios.

En segundo lugar se analizarán distintos esquemas de transmisión coordinada entre puntos de acceso para reducir las interferencias de los terminales que se encuentran en las fronteras de las celdas.

En tercer y último lugar, se evaluará la técnica de Expansión del Rango de la Celda (CRE) en los que se juega con las potencias de transmisión de las celdas pequeñas para captar a usuarios del borde.

2.2.1. Reutilización Fraccional de Frecuencias (FFR)

Las redes celulares están diseñadas para soportar patrones agresivos de reutilización de frecuencias con el fin de maximizar la eficiencia espectral. Sin embargo, este tipo de planificación de frecuencia conduce a un aumento considerable de la interferencia inter-celda (ICI), lo que puede causar una importante degradación del

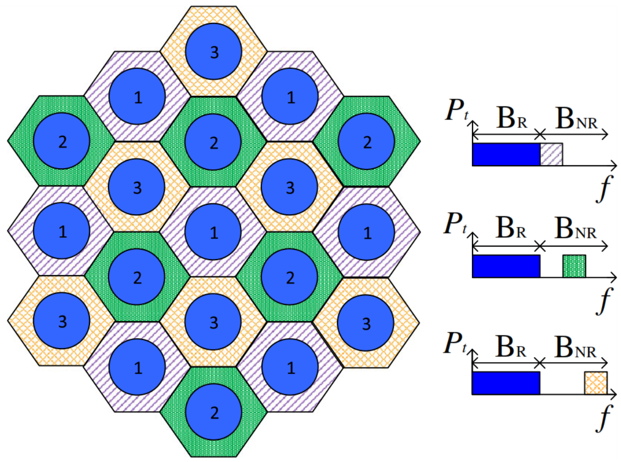


Figura 2.18: Ejemplo de topología celular con FFR

rendimiento, especialmente para los usuarios ubicados en el borde de la celda.

Las técnicas de Coordinación de Interferencias entre Celdas (ICIC) [46] ha sido ampliamente investigada como una tecnología clave para aliviar el impacto de la interferencia. Entre las técnicas existentes, este apartado se centra en la reutilización fraccional de frecuencias (FFR). Estos esquemas apuntan a reducir la interferencia en el borde de la celda aplicando un patrón de reutilización de frecuencia más flexible a los usuarios del borde de la celda, reduciendo así la interferencia para dichos usuarios [47]. Sin embargo, la alta eficiencia se mantiene mediante un patrón agresivo de reutilización 1 para aquellos usuarios con una relación alta de señal a interferencia más ruido (SINR), que generalmente se encuentra cerca de la estación base (BS).

En LTE-A, el Acceso Múltiple por División de Frecuencia Ortogonal (OFDMA) facilita el uso de esquemas FFR ya que toda la banda se puede dividir de manera bastante sencilla en dos patrones de reutilización diferentes. La Fig. 2.18 muestra un ejemplo de escenario FFR: una subbanda de frecuencia común, B_R, se emplea en todas las celdas (es decir, con una reutilización de frecuencia 1) mientras que el uso de la banda restante, B_{NR}, se coordina entre las celdas vecinas para crear subbandas con una reutilización de frecuencia superior a 1 (comúnmente 3) [47].

Uno de los objetivos de FFR es brindar un tratamiento más justo a los usuarios ubicados en el borde de las celdas. Pocos son los trabajos sobre FFR que abordan el problema de la imparcialidad entre usuarios. Cierto interés se muestra en [48], donde se analiza el rendimiento del peor usuario para un algoritmo FFR. El rendimiento de una planificación oportunista sobre un esquema FFR se evalúa en [49], considerando los requisitos de imparcialidad de los usuarios. En [47], se evalúan todas las posibles asignaciones de ancho de banda para la reutilización 1 y 3. Sin embargo, estos trabajos no permiten una comparación fácil entre las técnicas de asignación.

Para evaluar la imparcialidad en términos de tasa de tranmsisión por usuario, se pueden considerar varias métricas. El conocido índice de Gini [50], originalmente propuesto para estudiar la desigualdad en economía, es capaz de captar la imparcialidad general de la distribución y también se propuso para estudiar la imparcialidad de los planificadores en Internet [51].

En este trabajo consideramos un sistema con un número arbitrario P de celdas. Una de las celdas será la celda observada, mientras que las $P-1$ celdas restantes se tratarán como celdas interferentes. Se utiliza un método FFR como el de la Fig. 2.18. El ancho de banda total disponible del sistema B se divide en cuatro bandas de frecuencia ortogonales no superpuestas indicadas por B_R, B_{NR1}, B_{NR2} y B_{NR3}, cunmpliendo que $B = B_R + B_{NR}$ y $B_{NR} = B_{NR1} + B_{NR2} + B_{NR3}$. B_R representa el ancho de banda disponible para la región 1 de reutilización y B_{NRi} (donde $i \in [1,2,3]$) son tres conjuntos de subbandas de igual tamaño que representa el ancho de banda disponible para la región 3 de reutilización; cada una de estas subbandas se asignará a las celdas de una manera que ninguna otra celda circundante está utilizando la misma subbanda, como en una planificación clásica de frecuencias.

En la celda observada consideramos N usuarios, cada uno de ellos con una ubicación asociada s_k, $k \in [1..N]$. La SINR media asociada al usuario k^{th} depende de la región de la celda específica (reutilización 1 o 3) a la que el usuario ha sido asociado. De manera genérica, la SINR media asociada al usuario k^{th}, $SINR_k$, será igual a $SINR^{R1}(s_k)$ si se encuentra en la región de reutilización 1 o será $SINR^{R3}(s_k)$ si se encuentra en la región de reutilización 3. $SINR^{R1}$ y $SINR^{R3}$ representan la SINR media asociada con las regiones de reutilización 1 y 3, respectivamente.

Definimos la tasa de transmisión potencial, r_{0k}, asociada al usuario k^{th} como la tasa de transmisión alcanzable por el usuario k si todos los recursos disponibles estuviesen asignados a él. Dado que r_{0k} es una función de la SINR media asociada al usuario, r_{0k} puede tener diferentes valores dependiendo de la región de la celda a la que está asociado el usuario. Específicamente, si el usuario k está asociado a la región de reutilización 1, r_{0k} será igual a $c\left(SINR^{R1}\left(s_k\right)\right)$, mientras que si está en la región de reutilización 3, r_{0k} será igual a $\frac{1}{3}c\left(SINR^{R3}\left(s_k\right)\right)$.

El objetivo es encontrar una clasificación óptima que conduzca a una tasa de transmisión máxima para cada usuario. Esto es equivalente a encontrar un umbral de SINR óptimo, ya que r_{0k} depende de la SINR media asociada al usuario. Por lo tanto, la siguiente expresión se aplicará para clasificar a los usuarios:

$$\text{If } c\left(SINR^{R1}\left(s_k\right)\right) < \frac{1}{3}\left(SINR^{R3}\left(s_k\right)\right), k \in U^{R3} \Rightarrow r_{0k} = \frac{1}{3}c\left(SINR^{R3}\left(s_k\right)\right) \tag{2.12}$$

donde U^{R1} y U^{R3} representan el conjunto de usuarios asociados a las regiones de reutilización 1 y 3, respectivamente.

Planificación de recursos radio

Sea ρ_k la fracción de los recursos (ancho de banda y tiempo) asignados al usuario k^{th}; se debe cumplir la siguiente restricción si todos los recursos disponibles se asignan a los usuarios: $\sum_{k=1}^{N} \rho_k = 1$. Además, si consideramos una implementación basada en FFR, se deben tener en cuenta las dos siguientes restricciones adicionales para dimensionar adecuadamente las bandas de frecuencia de reutilización 1: $B_R/B = \sum_{k \in U^{R1}} \rho_k$, y reutilización 3: $B_{NR}/B = \sum_{k \in U^{R3}} \rho_k$, según la distribución de usuarios dentro de cada región.

La cantidad real de recursos asignados a un usuario depende de la política de planificación seleccionada, es decir, el valor de ρ_k. Por lo tanto, la tasa de transmisión asignada al usuario k se puede calcular como $r_k = \rho_k \cdot r_{0k}$. Teniendo en cuenta la definición anterior, la tasa agregada de la celda, R_N, está determinada por la política de planificación concreta y se puede calcular como $R_N = \sum_{k=1}^{N} r_k = \sum_{k=1}^{N} \rho_k \cdot r_{0k}$.

Con respecto a la política de planificación específica que determina el valor de ρ_k, se pueden aplicar muchos esquemas diferentes, algunos de ellos se describen a continuación:

- *Equal Resource Sharing (ERS)*: Los recursos tiempo/frecuencia se comparten de forma equitativa entre los usuarios independientemente de su tasa de transmisión potencial, es decir, $\rho_k = 1/N$. Por lo tanto, la tasa agregada de transmisión es $R_N^{ERS} = 1/N \sum_{k=1}^{N} r_{0k}$, debiendo satisfacer que $B_R/B = N^{R1}/N$ y $B_{NR}/B = N^{R3}/N$ donde N^{R1} y N^{R3} es el número de usuarios asignados a las bandas de reutilización 1 y 3, respectivamente.

- *Equal Transmission Rate (ETR)*: La asignación de recursos es inversamente proporcional a las tasas potenciales de transmisión de los usuarios (es decir, se asignan más recursos a los usuarios con SINR más baja), de modo que las tasas de transmisión reales se distribuyen por igual entre los usuarios, es decir, $\rho_k = K_1/r_{0k}$, donde $K_1 = \left(\sum_{k=1}^{N} 1/r_{0k}\right)^{-1}$ para satisfacer la restricción de asignación de recursos. En este caso, la tasa agregada se puede calcular como $R_N^{ETR} = N \cdot K_1$ y la proporción entre las bandas de frecuencia 1 y 3 deberán satisfacer $B_R/B = 1/N \cdot \sum_{k \in U^{R1}} R_N^{ETR}/r_{0k}$ y $B_{NR}/B = 1/N \cdot \sum_{k \in U^{R3}} R_N^{ETR}/r_{0k}$, respectivamente.

- *Truncated-ETR*: Esta es una versión modificada de la política anterior que considera un valor máximo de ρ_k para evitar asignar demasiados recursos a los usuarios que reciben con una SINR muy baja, es decir: $\rho_k = K_2 \cdot \text{mín}\{K_1/r_{0k}, \varepsilon/N\}$, donde $K_2 = \left(\sum_{k=1}^{N} \text{mín}\, K_1/r_{0k}, \varepsilon/N\right)^{-1}$ y $\varepsilon \in [0, \infty]$ es un factor de truncamiento con respecto al esquema ERS. ε/N significa que los usuarios con una SINR pobre obtendrán un máximo de ε veces los recursos que se les asignaría si se utilizara un esquema ERS. Al variar el valor de ε, este esquema puede comportarse de manera muy diferente. De hecho, la tasa agregada $R_N^{\varepsilon} = K_2 \sum_{k=1}^{N} \text{mín}\, R_N^{ETR}/(N \cdot r_{0k}), \varepsilon/N$ estará dentro de los límites de algoritmos anteriores: $R_N^{ETR} = \lim_{\varepsilon \to \infty} R_N^{\varepsilon} \leq R_N^{\varepsilon} \leq \lim_{\varepsilon \to 0} R_N^{\varepsilon} = R_N^{ERS}$.

Indicador de imparcialidad entre usuarios

El análisis de la imparcialidad implica el estudio de la distribución de la cantidad de recursos (ancho de banda y tiempo) asignados a cada usuario. Sin embargo, como las diferentes SINR conducen a una tasa potencial distinta, en este apartado se analiza la tasa de transmisión asignada (en lugar de los recursos asignados) mediante la curva de Lorenz y el índice de Gini [50].

Sea $\langle r_k, k = 1 \cdots N \rangle$ el conjunto ordenado de tasas asignadas a los usuarios bajo cierta política de planificación. La tasa acumulada por los primeros n de esos usuarios viene dada por $R_n = \sum_{k=1}^{n} r_k$.

La curva de Lorenz (ver Fig. 2.19 a)) muestra, en el eje x, el porcentaje del número acumulado de usuarios normalizado al número total de usuarios, n/N, mientras que en el eje y muestra la tasa acumulada de esos n usuarios normalizado a la tasa agregada, R_n/R_N. La recta a 45° representa la igualdad perfecta en el coeficiente de Gini de tasa asignada. El índice de Gini se puede considerar como la proporción del área que se encuentra entre la línea de igualdad y la curva de Lorenz sobre el área total debajo de la línea de igualdad. Este coeficiente puede variar de 0 (igualdad perfecta) a 1 (desigualdad máxima).

La curva de Lorenz se puede generalizar utilizando el valor no normalizado de R_n en el eje y, dando como resultado la Fig. 2.19 b). Como se verá en la siguiente sección, esta representación puede mostrar el intercambio de la tasa asignada entre usuarios aventajados y desfavorecidos para ciertos valores de ε.

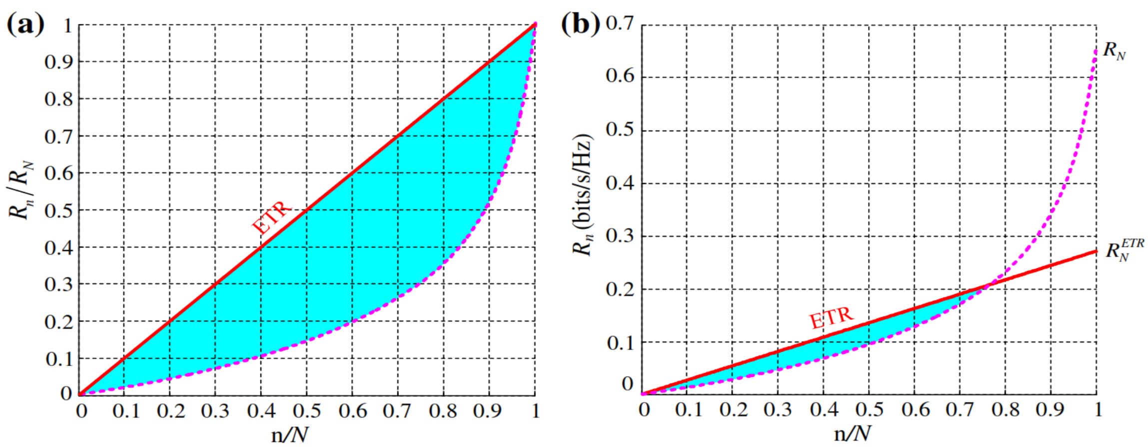


Figura 2.19: a) Curva de Lorenz, b) Curva de Lorenz generalizada

Resultados

Se considera el enlace descendente de una red celular LTE-A con 61 celdas hexagonales de radio R=1 km. Suponemos que las BS utilizan antenas omnidireccionales ubicadas en los centros de las celdas y con la misma potencia de transmisión, $P_t = 49$ dBm. El escenario asumido es un área urbana densa servida por macroceldas. Hay una cantidad de usuarios distribuidos aleatoriamente en la celda observada cuyas posiciones son variables aleatorias distribuidas uniformemente en la celda.

La señal recibida por los usuarios de la celda observada sufre pérdidas de propagación según la siguiente expresión:

$$\begin{array}{rc} 98{,}42 + 20log_{10}\,(d) & \text{if d} < 0{,}5 \text{ km} \\ 140{,}35 + 38log_{10}\,(d) & \text{if } 0{,}5 \leq d < 1 \text{ km} \\ 140{,}36 + 35{,}04log_{10}\,(d) & \text{if d} \geq 1 \text{ km} \end{array} \tag{2.13}$$

Se considera la existencia de desvanecimientos a pequeña escala que siguen un modelo de Rayleigh. Además, para facilitar la evaluación numérica, hemos considerado que la interferencia recibida del resto de las celdas en el sistema sigue una distribución gaussiana [49] y su intensidad de señal solo se ve afectada por las pérdidas de propagación. Por simplicidad, suponemos que el sistema está limitado por interferencia.

Para determinar la tasa de transmisión potencial, hemos tenido en cuenta las restricciones establecidas en un sistema LTE-A. En tales sistemas, AMC se emplea para maximizar el rendimiento mientras se mantiene la tasa de error de bloque (BLER) bajo un objetivo predefinido ($BLER_t$). Este objetivo se puede lograr seleccionando el MCS óptimo sujeto a la $BLER_t$, que se asume 0.1 en este trabajo, entre un conjunto de 16 posibles esquemas. Por lo tanto, la tasa de transmisión puede calcularse considerando la probabilidad de usar cada canal MCS dada una SINR media. Suponiendo que el MCS i se asocia con una tasa de modulación de b_i bits/símbolo y una tasa de codificación c_i, la tasa de datos efectiva cuando se usa MCS i se puede calcular como $r_i = b_i \cdot c_i$. En un canal con desvanecimiento Rayleigh, la probabilidad de seleccionar un MCS i en particular viene dada por:

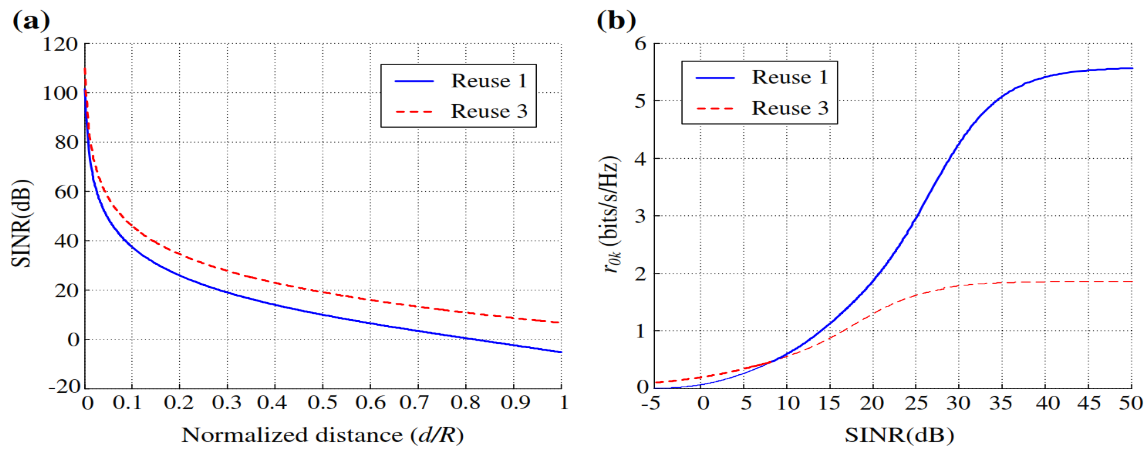


Figura 2.20: a) SINR, b) Tasa de transmisión potencial

$$p_i = \exp\left(-\frac{\gamma_i}{\bar{\gamma}}\right) - \exp\left(-\frac{\gamma_{i+1}}{\bar{\gamma}}\right) \quad . \tag{2.14}$$

donde γ_i y γ_{i+1} son los umbrales de SINR inferior y superior asociados al MCS i, respectivamente; y $\bar{\gamma}$ es la SINR media. A continuación, se puede obtener r_{0k} promediando las tasas de transmisión individuales para cada MCS: $r_{0k} = \sum_{i=0}^{15} r_i \cdot p_i$.

La Fig. 2.20 a) representa la evolución de la SINR asociada a cada región en función de la distancia normalizada entre un usuario y la BS en la celda observada. Se puede ver que $SINR^{R3}$ siempre es mayor que $SINR^{R1}$. En la Fig. 2.20 b) se puede ver cómo se logra la reducción del nivel de interferencia a costa de reducir la tasa alcanzable en esta región (línea discontinua). Como se indica en (2.12), se puede ver que la forma de maximizar la tasa potencial es seleccionar el valor más alto de ambas expresiones en cada SINR. Debe observarse que el umbral de SINR que conduce a la máxima tasa potencial para cada usuario es el mismo independientemente de la política de planificación seleccionada y se puede ver en la Fig. 2.20 b) que tiene un valor de 9.3 dB. En este caso, dado que no se considera el efecto del *shadowing*, el umbral de SINR es equivalente al umbral de distancia, X_f, con un valor de $0.515R$. Esto significa que las regiones de reutilización 1 y 3 pueden ser renombradas como región interna y externa, respectivamente [47].

Una vez que se haya establecido el umbral óptimo de SINR, vamos a analizar el rendimiento del sistema bajo las tres políticas consideradas. La Fig. 2.21 muestra la

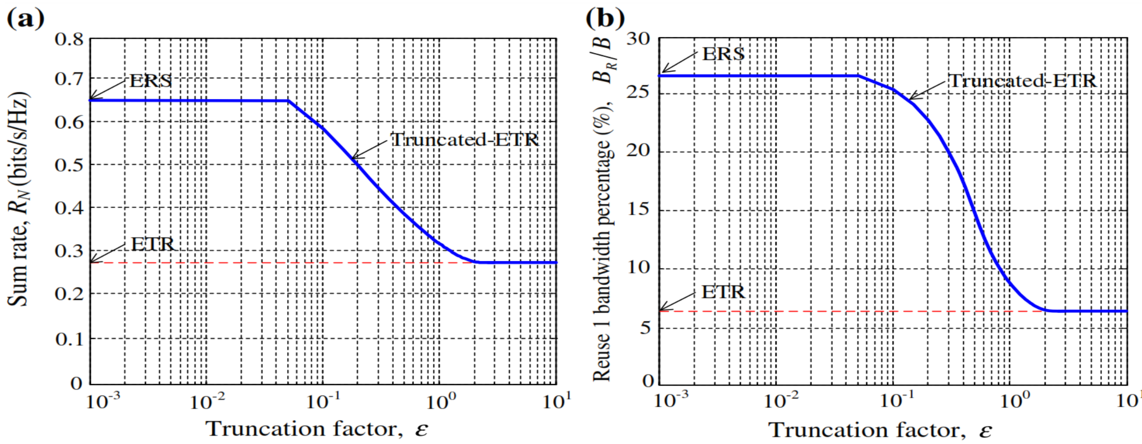


Figura 2.21: a) Tasa de transmisión alcanzable por el sistema, b) Partición del ancho de banda del sistema

tasa de transmisión alcanzable (2.21 a)) y la división del ancho de banda del sistema (2.21 b)) en función del parámetro ε . Se puede ver en la Fig. 2.21 a) que el valor de la tasa agregada para la estrategia ERS es mayor que la asociada a ETR. Además, en la Fig. 2.21 b), se observa que la política ETR asigna casi el 95 % del ancho de banda para usuarios de la banda de reutilización 3, mientras que la estrategia ERS asigna más del 25 % del ancho de banda para los usuarios de la banda de reutilización 1.

El índice de Gini asociado a la política ETR truncada en función del parámetro ε y en función de la tasa agregada se muestra en la Fig. 2.22. Se puede ver en la Fig. 2.22 a) cómo aumenta el índice de Gini a medida que disminuye el valor de ε. Esto se debe a que, a medida que disminuye ε, ciertos recursos se transfieren de usuarios con menores tasas potenciales r_{0k} a otros cuyo uso es más eficiente, lo que aumenta las desigualdades entre ellos. Finalmente, cuando $\varepsilon \to 0$, la política ETR truncada asigna a todos los usuarios la misma cantidad de recursos, pero aquellos con SINR más alta lo explotan de una manera más eficiente. Por lo tanto, la política ERS logra una mayor tasa agregada R_{ERS} a costa de lograr una mayor injusticia entre los usuarios.

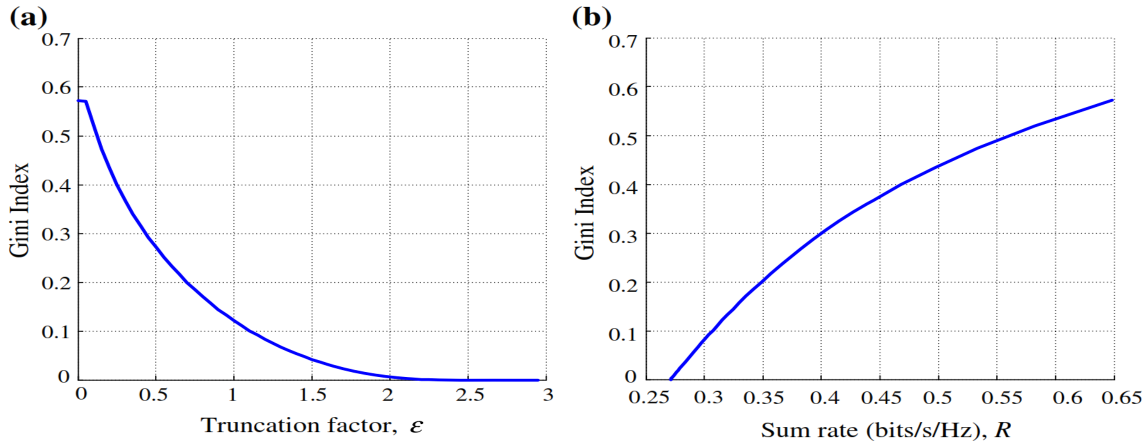


Figura 2.22: Evolución del índice de Gini (ETR truncado) vs. factor de truncamiento, b) Tasa total de transmisión

2.2.2. Técnicas de transmisión coordinadas multipunto (CoMP)

Las técnicas de transmisión coordinadas multipunto (CoMP) suponen una de las propuestas más novedosas para la tecnología LTE-Advanced. CoMP puede considerarse como un sistema de múltiples antenas (MIMO) distribuido, en el que los nodos distribuidos geográficamente forman múltiples antenas y cooperan para transmitir y/o recibir información. El objetivo es aumentar el rendimiento del sistema, especialmente en el borde de la celda, donde la interferencia inter-celda (ICI) suele ser alta.

Esta sección se centra en dos técnicas de transmisión coordinada que se utilizan para aumentar el rendimiento del sistema, con el objetivo de evaluar su rendimiento para las comunicaciones M2M a través del enlace descendente de un sistema LTE-A.

La primera técnica considerada es la transmisión conjunta (JT) [12] sin precodificación, donde un NodeB evolucionado (eNB) y una unidad radio remota (RRU) colaboran para transmitir datos de forma conjunta a uno o más UEs.

La segunda técnica estudiada es la Reutilización de Frecuencia Parcial (PFR), donde las células adyacentes deben coordinarse para dividir todo el ancho de banda en dos patrones de reutilización (reutilización 1 y 3) [13-14].

PFR mejora el proceso de recepción de los usuarios al evitar que interfieran. La Fig. 2.23 (izquierda) muestra el escenario PFR considerado: el ancho de banda

B total del sistema disponible se divide en dos bandas de frecuencia ortogonales no superpuestas indicadas por B_{center} y B_{edge}. B_{center} representa el ancho de banda disponible para los usuarios cerca del eNB y B_{edge} se divide en tres conjuntos de subbandas de igual tamaño que representan el ancho de banda disponible para los usuarios de borde de celda; cada una de estas subbandas se asignará a las celdas de manera que ninguna otra celda circundante esté utilizando la misma subbanda. De esa manera, cada celda utiliza el espectro B_{center} para atender a los usuarios del centro de la celda y un tercio del espectro de B_{edge} para atender a los usuarios de la celda. Esta asignación de frecuencia garantiza que los usuarios de borde de celda no recibirán interferencia de las celdas vecinas más cercanas porque a estas celdas se les asignan diferentes sub-bandas de frecuencia para su transmisión en esa región. Esta reducción en el nivel de interferencia se logra a costa de reducir la cantidad de ancho de banda asignable por celda (solo se utiliza un tercio de B_{edge} en cada celda).

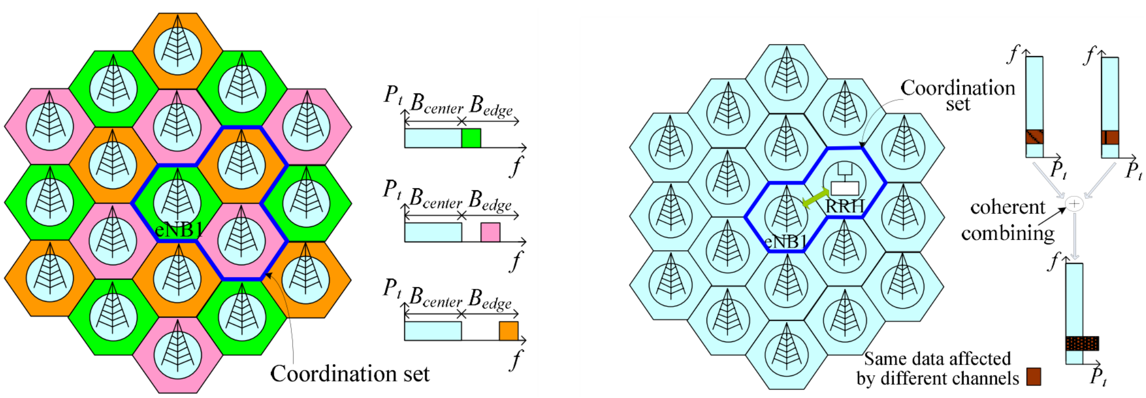


Figura 2.23: Topología celular de un esquema PFR (izquierda) y Transmisión Conjunta (JT) en transmisión (derecha)

A diferencia de la técnica anterior, la técnica de transmisión conjunta (JT) sin precodificación intenta explotar el hecho de que un usuario recibe señales de celdas adyacentes. De hecho, esta técnica consiste en transmitir la misma información simultáneamente desde diferentes eNBs para mejorar el proceso de recepción. Esta información debe combinarse coherentemente en el receptor móvil. La Fig. 2.23 (derecha) muestra el proceso realizado cuando se aplica esta técnica. Como puede verse, los mismos recursos de frecuencia de la celda adyacente están dedicados a la transmisión de la misma información a los usuarios de borde de celda. Esto significa que la capacidad de estos recursos de frecuencia se divide por el número de entidades involucradas en el proceso de JT. Por lo tanto, es importante decidir qué usuarios se consideran en la región del borde de la celda para evitar el uso de esta técnica por parte de usuarios cuyo proceso de recepción no está limitado por interferencias. Esta decisión se puede tomar en base a la potencia media recibida de la celda que le da servicio y de las otras celdas. Si ambos valores son similares, significa que el usuario está en el borde de la celda. De lo contrario, el usuario está cerca del eNB y las interferencias no afectan significativamente al proceso de recepción.

La eficiencia de esta técnica depende de si las señales provenientes de diferentes celdas se suman en fase o en contrafase (ya que no se supone una precodificación). En el peor de los casos, la adición puede ser destructiva y la señal recibida será nula. De todos modos, la recepción mejorará a medida que aumente el número de celdas transmisoras.

Cabe señalar que las celdas vecinas deben coordinarse en ambas técnicas, pero el grado de coordinación es muy diferente en cada caso. Por un lado, en la técnica PFR, las celdas adyacentes deben coordinarse para definir las bandas de frecuencia no superpuestas, B_{center} y B_{edge}, y para decidir qué subbanda pueden usar cada una. Sin embargo, este intercambio de información se realiza típicamente en un tiempo de gran escala (es decir, varios días). Por otro lado, debe existir una coordinación perfecta entre las celdas adyacentes cuando se aplica la técnica JT, ya que las entidades coordinadas deben conocer, no solo los recursos de frecuencia asignados a cada usuario de borde de celda, sino también los datos que deben transmitirse.

Resultados de simulación

Se considera el enlace descendente de una red celular LTE con 61 celdas hexagonales de radio 0.5 km. Todos los eNBs utilizan la misma potencia de transmisión. Un resumen de los parámetros de simulación se muestra en la Tabla 2.5.

Se han simulado 10 usuarios distribuidos aleatoriamente en la celda observada. Cada uno de ellos recibe tráfico generado mediante un modelo de fuente de cámara de video IP [52], que corresponde a una aplicación M2M de control remoto sensible al retardo. Se considera que un usuario está en el área del borde de la celda siempre que

Tabla 2.5: Parámetros de simulación

LTE Parameter	Value/Mode
Number of cells	61
Cell radius	500 m
Number of antennas	Single antenna at eNBs and UEs
Transmit power of eNBs	49 dBm
Path loss model	Hata model [13]
Multipath model	Rayleigh fading
Mobile Speed	4km/h (pedestrian)
Channel bandwidth	20 MHz
OFDM symbols per TTI	14
PRB size	12 subcarriers
Carrier frequency	2.5 GHz
Modulation schemes	QPSK, 16QAM and 64QAM
Target BLER	10%
Number of users	10
Simulation length	20s
Type of traffic	M2M

la diferencia entre la potencia recibida de la celda observada y de las celdas adyacentes sea inferior a 5 dB. Este valor se ha elegido para garantizar que la cooperación solo se aplique a los usuarios que tienen condiciones de canal muy malas. En la técnica PFR, el tamaño de B_{edge} se determina proporcionalmente al número de usuarios de borde. En la técnica JT, este proceso no es necesario ya que no se realiza ninguna división de ancho de banda.

Los resultados se asocian a tres casos de estudio: 1) no se aplica ninguna técnica, es decir, se trata de una red celular con un factor de reutilización 1 (denominado *baseline*), 2) técnica JT sin precodificación y 3) técnica PFR.

La Fig. 2.24 muestra la SINR media asociada a usuarios de borde (líneas rojas) y usuarios de centro (líneas azules) para los tres casos de estudio. Los niveles de SINR se presentan para diferentes valores de incremento de potencia de transmisión en el eNB, definidos como:

$$\Delta P_{Tx} = P_{Tx} - P_n - Lp_{\max}[\text{dB}] \tag{2.15}$$

donde $Lp_{\max}$ corresponde a la pérdida de propagación en el borde de la celda, P_n es la potencia de ruido y P_{Tx} es la potencia de transmisión. Se puede ver que para valores bajos de potencia de transmisión, la SINR está dominada por la potencia de ruido. A medida que ΔP_{Tx} aumenta (por encima de 20 dB), la SINR está dominada por la interferencia. Finalmente, los valores medios de SINR se saturan porque el efecto de aumentar la potencia de señal deseada compensa el aumento de potencia de la señal interferente.

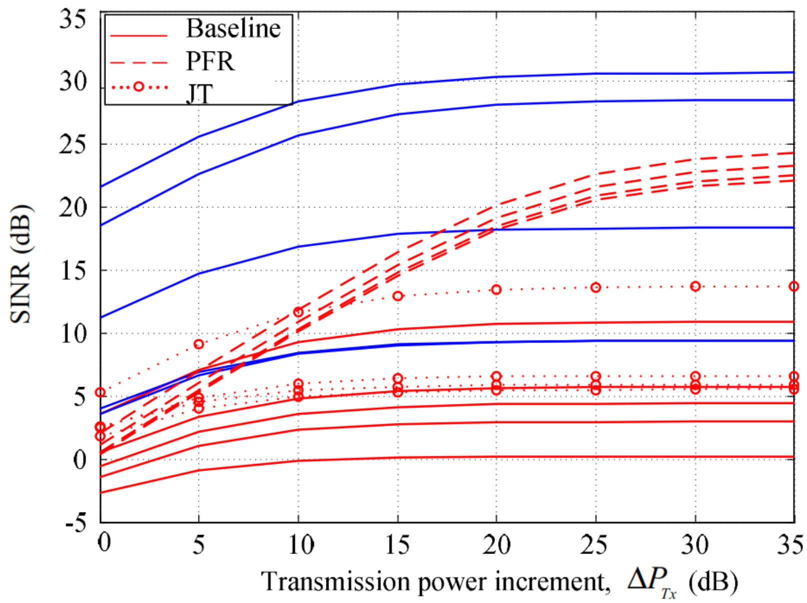


Figura 2.24: SNR media por usuario (usuario del borde de la celda en color rojo)

La Fig. 2.25 muestra los resultados de retardo medio de paquete en la celda observada (izquierda) y el retardo medio de paquete para los usuarios del borde de la celda (derecha). Se puede observar que el retardo medio de la celda es esencialmente el mismo tanto para la configuración de referencia como para la técnica JT, mientras que el retardo para los usuarios de borde se reduce ligeramente. Por otro lado, PFR mejora los resultados de retardo para la mayoría de los valores de ΔP_{Tx}. Solo para los valores más bajos de ΔP_{Tx}, los resultados son bastante similares porque en este punto la SINR de los usuarios en el borde de la celda es bastante similar.

La Fig. 2.26 presenta una comparación de la tasa media de transmisión del usuario en la celda completa (izquierda) y en el borde (derecha). Vale la pena notar que dicha tasa es ligeramente peor para la técnica JT que para la configuración de referencia. Esto es así porque la celda adyacente tiene menos recursos para compartir debido al

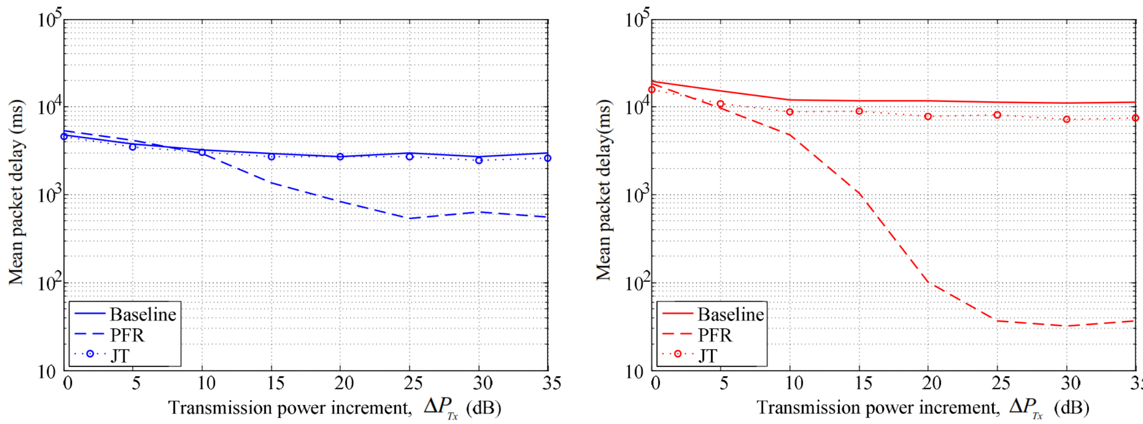


Figura 2.25: a) Retardo medio de paquete en toda la celda, b) Retardo medio de paquete para los usuarios del borde de la celda

uso de JT. Solo es aconsejable el uso de JT cuando la tasa de transmisión del usuario aumenta más del doble. Este es en realidad un criterio de decisión más exigente que un umbral de potencia de 5 dB. Sin embargo, el interés de esta técnica proviene de la mejora del rendimiento de los usuarios del borde, como se puede ver en la figura de la derecha. En particular, la ganancia a 20 dB es del 64 %, mientras que la pérdida relativa en el rendimiento medio de los usuarios es del 5 %. Además, se observa que los resultados para la técnica PFR supera a los de JT y la referencia para valores medios y altos de ΔP_{Tx}. Cabe señalar que el uso de PFR tiene un equilibrio entre la imparcialidad entre usuarios y el rendimiento del sistema.

2.2.3. Expansión del Rango de la Celda (CRE)

A la hora de realizar la asociación entre usuarios (UEs) y estaciones base (BSs) se pueden tener en cuenta distintos criterios. Típicamente se consideran dos posibles criterios: 1) pérdidas de propagación mínima; 2) mayor potencia media recibida. Si se elige este último cuando se tiene un escenario heterogéneo, es posible aplicar un incremento (sesgo) sobre la potencia de transmisión de las celdas pequeñas (*Small-cell Base Stations*, SBSs) de manera que se amplíe su área de cobertura. Esta técnica, conocida como Expansión del Rango de la Celda (*Cell Range Expansion*, CRE) [53], consigue reducir la carga asociada a las celdas grandes (*Macro-cell Base Stations*,

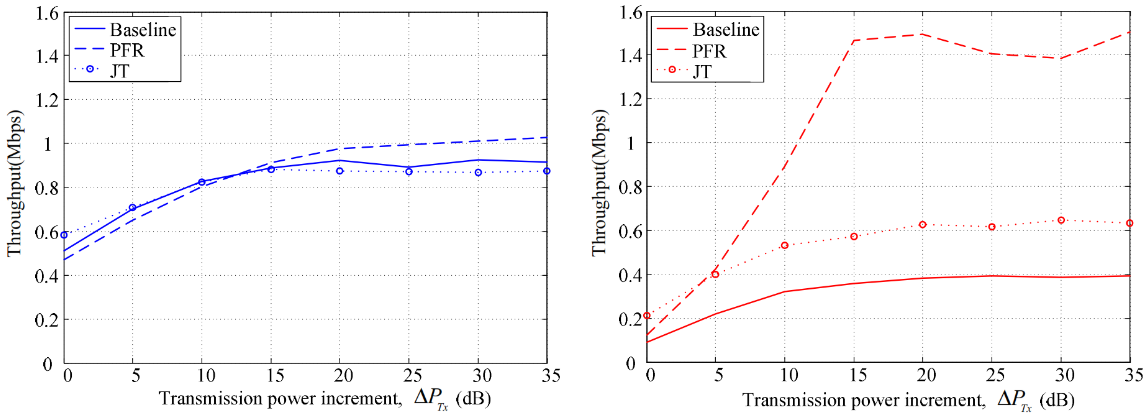


Figura 2.26: a) Tasa media de transmisión en toda la celda, b) Tasa media de transmisión para los usuarios del borde de la celda

MBSs) puesto que al aplicar el sesgo se fuerza a que UEs que estarían asociados a una MAP se asocien a una PAP. Ahora bien, estos usuarios forzados (CRE UEs), suelen tener unos valores de nivel de señal a interferencia (SINR) muy degradados debido a que la MAP a la que estarían asociados de forma natural actúa ahora como interferente.

A continuación se muestran los resultados mediante simulación de un escenario compuesto por 4 SBSs ubicados dentro del área de cobertura de una MBS. El objetivo de este estudio es analizar el efecto de la asociación de los usuarios en el rendimiento del sistema cuando se aplica la técnica CRE. Los parámetros de simulación se detallan en la Tabla 2.6.

El escenario de simulación consta de una MBS con cuatro SBS colocadas dentro de su área de cobertura. El área de cobertura de la MBS es de forma hexagonal con un radio interior de 250 m. Los SBS se ubican alrededor de la MBS a 0°, 90°, 180° y 270° en coordenadas polares, estando la MBS en el origen de coordenadas. Se analiza el efecto de la distancia entre MBS y SBS en un entorno urbano, es decir, desde una ubicación cercana a la MBS a una ubicación cerca del borde de la celda. En particular, las distancias tomadas en cuenta son 40, 100 y 160 m, denominados escenarios P-1, P-2 y P-3, respectivamente (ver Fig. 2.27 a 2.29). Un total de 60 usuarios están distribuidos uniformemente con una distancia mínima UE-MBS de 35 m. Se han generado 8 realizaciones espaciales de las ubicaciones de UE según un Proceso de

Tabla 2.6: Parámetros de simulación por defecto

Parameter	Value
Carrier Freq.	2 GHz
Bandwidth	10 MHz
Transmission Power	43 dBm (MBS), 30 dBm (SBSs)
Simulation Time	20000ms
Transmission Mode	SISO
Number of users	60
Association Criterion	Máx. RSRP+Bias
Distances MBS-SBSs	40, 100, 160 m
Hexagon Apothem	250 m
Spatial Distribution process	BPP
No. of Spatial Realization	8
Path Losses	ITU UMa, ITU Umi
Fast Fading	ITU UMa
Scheduler	Round Robin
Max. Allocation per user	15 PRBs
Source	FTP Model 3, λ = 2,5, 0.5 MB
Decoding Algorithm	SOVA
Link Adaptation Method	OLLA
Channel Estimation	Ideal
SNR Estimation	Error based

Punto Binomial. Para cada realización espacial, hemos realizado una simulación de 20000 subtramas. En la Fig. 2.30, hay un ejemplo de realización espacial de usuarios. Posteriormente, los resultados se promedian en el dominio espacial y temporal para obtener los indicadores de rendimiento previstos. Además, la asociación de los usuarios a las celdas tiene en cuenta tanto la potencia recibida como el sesgo CRE, con un rango de valores entre 0 dB y 30 dB.

La BS está configurada para transmitir con un ancho de banda de 10 MHz en una portadora de 2 GHz. La potencia de transmisión de las MBS y SBS es de 43 dBm y 30 dBm, respectivamente. Se usa AMC junto con la técnica OLLA y un planificador de paquetes Round Robin. Las pérdidas de propagación se calculan siguiendo la definición para Macro urbana (UMa) de la UIT y la Micro urbana (Umi) [54]. Se considera un modelo de Rayleigh para los desvanecimientos por propagación

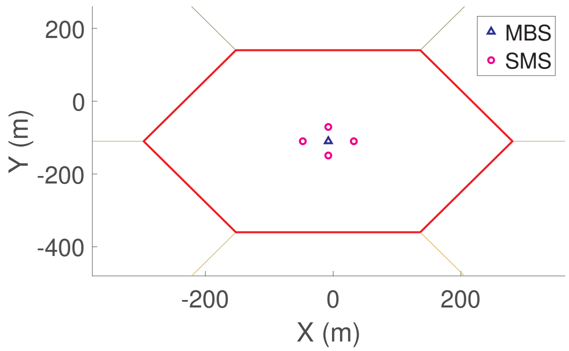


Figura 2.27: Escenario P-1

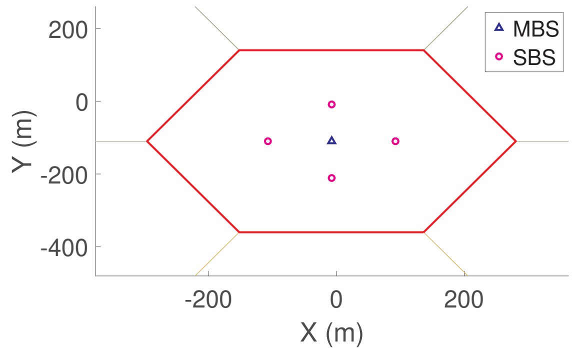


Figura 2.28: Escenario P-2

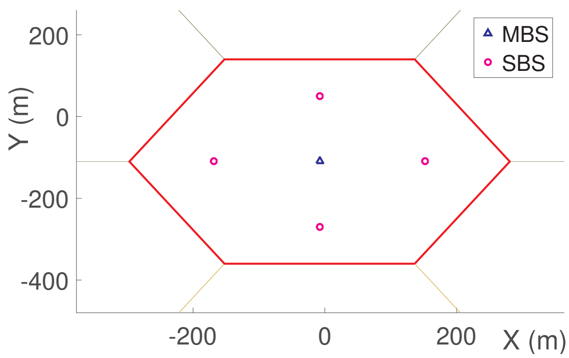


Figura 2.29: Escenario P-3

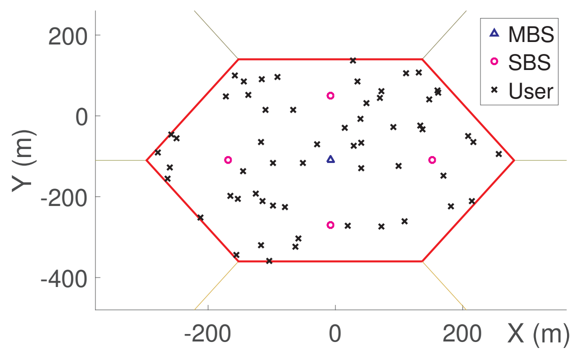


Figura 2.30: Realización espacial de usuarios

multicamino con la configuración UMa de la UIT [55]. Se utiliza un modelo de tráfico para el Protocolo de transferencia de archivos (FTP); en particular, consideramos el modelo de FTP 3 [54], donde los paquetes llegan de acuerdo con un proceso de Poisson con un tiempo de llegada de paquete que sigue una distribución exponencial. Además, los archivos tienen un tamaño de 0.5 Mbytes.

Las métricas que se analizan son las siguientes: asociación de usuarios por nivel, carga de PRBs por nivel, Tasa de transmisión agregada, porcentaje de usuarios de baja tasa de transmisión e imparcialidad.

Asociación de usuarios por nivel Representa el número de usuarios que están asociados al nivel 1 (MBS) y al nivel 2 (SBS). Cada usuario se asociará a la celda con el nivel de potencia recibido (RSRP) máximo medido más el sesgo CRE.

Carga de PRBs por nivel Este es el porcentaje de PRBs utilizados por nivel. Como el ancho de banda simulado es de 10 MHz, el uso máximo del nivel 1 es de 50 PRB, mientras que para el nivel 2 es de 200 PRB (4 celdas pequeñas).

Tasa de transmisión agregado Se calcula como los bits agregados recibidos por todos los usuarios (de ambos niveles) por segundo.

Porcentaje de usuarios de baja tasa de transmisión Se usa para mostrar el porcentaje de usuarios que tienen una tasa de transmisión baja, es decir, menos de 10kbps, nombrada como usuarios de tasa baja (LRU).

Imparcialidad Se usa para determinar si los usuarios reciben una parte justa de los recursos del sistema. El índice de Jain se utiliza para calcular la imparcialidad de las celdas simuladas de la siguiente forma:

$$J\left(x_1, x_2, ..., x_n\right) = \frac{\left(\sum_{i=1}^{n} x_i\right)^2}{n \cdot \sum_{i=1}^{n} x_i^2} \tag{2.16}$$

donde x_i es el número total de bits recibidos por el usuario i, y n representa el número de usuarios.

A continuación se presentan los resultados de simulación para diferentes valores de sesgo CRE. Tenga en cuenta que, al aumentar el sesgo, los usuarios se asociarán a una SBS con menos carga a costa de reducir su SINR. Es un compromiso entre los recursos disponibles y la SINR, ya que el sesgo aumenta la SINR más baja en la nueva celda de servicio. En consecuencia, la transferencia de usuarios del nivel 1 al nivel 2 debido a un alto valor de sesgo puede causar que los usuarios no pueden transmitir o tengan tasas de transmisión demasiado bajas.

En la Fig. 2.31, se muestra la asociación de usuarios para los escenarios P-1, P-2 y P-3 en función del sesgo. Por un lado, las líneas continuas representan la asociación del usuario a nivel 1 (MBS). Por otro lado, las líneas discontinuas representan la asociación a nivel 2 (SBS). Se puede ver que el punto de intersección donde el 50 % de los usuarios se dividen por igual en ambos niveles se alcanza antes, ya que las SBSs están más alejadas de la MBS. En P-1, la MBS está cerca de las SBSs y, por lo tanto, es difícil equilibrar a los usuarios de la MBS debido a su alta potencia.

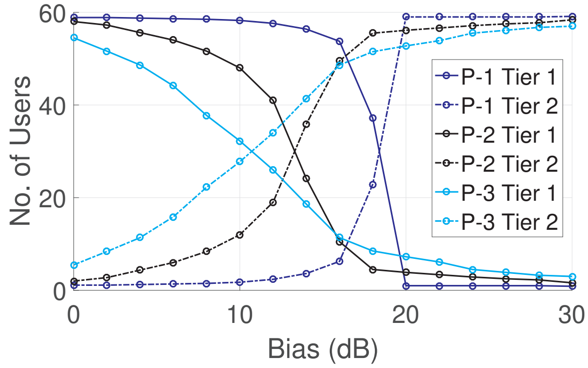


Figura 2.31: Asociación de usuarios por nivel

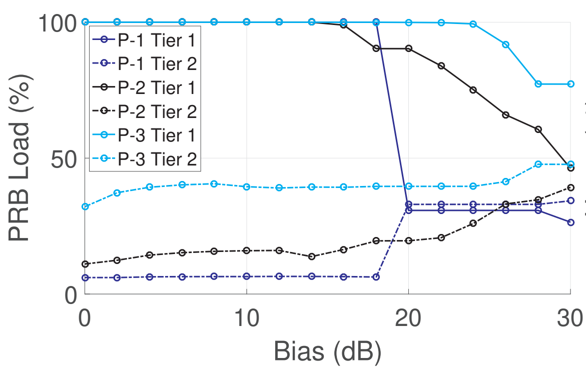


Figura 2.32: Porcentaje de Carga de PRB por nivel

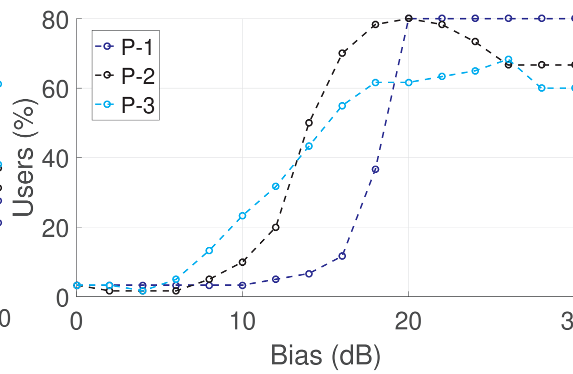


Figura 2.33: Porcentaje de usuarios de baja tasa de transmisión

Además, la carga por nivel (Fig. 2.32) es un resultado que está vinculado a la figura anterior. Al principio, cuando el sesgo es bajo, la MBS se carga al 100 % y cuando aumenta el número de usuarios asociados al nivel 2, la MBS se descarga. Tenga en cuenta que, aunque la cantidad de usuarios asociados al nivel 2 crece, no todos los usuarios pueden transmitir debido a su baja SINR. En la Fig. 2.33 se puede ver que el número de usuarios con baja tasa de transmisión aumenta cuando aumenta el sesgo porque estos usuarios se ven obligados a asociarse a una SBS pero con una SINR que no es suficiente para transmitir, principalmente por la interferencia de la MBS. La diferencia entre las tres curvas se produce para valores de alto sesgo. Para P-2 y P-3, el número de usuarios de baja tasa de transmisión disminuye con valores de alto sesgo. Esto se debe al hecho de que la interferencia se reduce cuando la MBS se carga por debajo del 90 % y, en consecuencia, los usuarios pueden transmitir con mejores condiciones. Por el contrario, para P-1, el número de usuarios de tasa baja permanece constante con un valor de 80 % porque el usuario más cercano a la MBS siempre está asociado a ella.

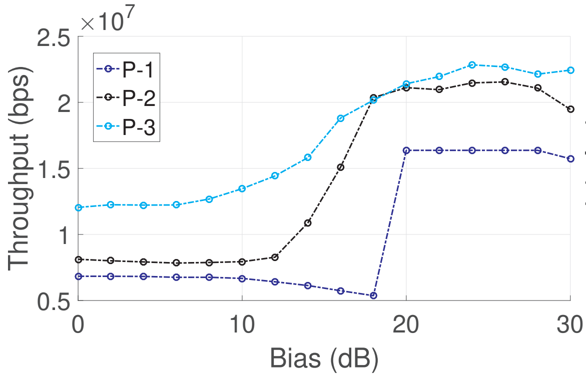


Figura 2.34: Tasa de transmisión agregada

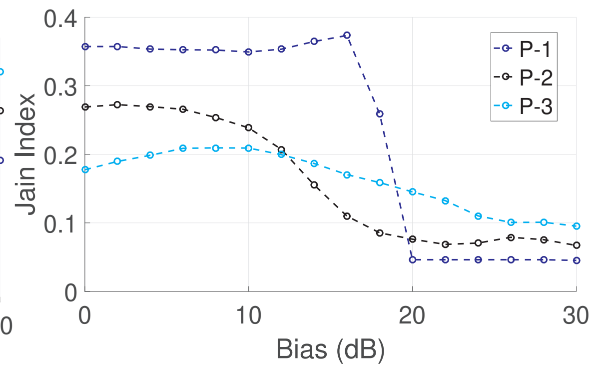


Figura 2.35: Imparcialidad (índice de Jain)

En la Fig. 2.34 se muestra la tasa de transmisión agregada de todas las celdas simuladas. Para valores de sesgo inferiores a 10 dB, el rendimiento de P-1 y P-2 disminuye progresivamente. Esto se debe a que los peores usuarios asociados a la MBS ahora están asociados a las SBSs y su velocidad de transmisión ahora es menor o nula. Sin embargo, para P-3, el rendimiento aumenta ligeramente para un sesgo de 2 dB como resultado de la nueva asociación de usuarios. Esto se debe a que con un sesgo inferior a 2 dB, los usuarios se asocian a una MBS bastante cargada. Sin embargo, con un sesgo de 2 dB, los usuarios pueden transmitir asociados a una SBS descargada. Por tanto, para valores de sesgo más altos, el rendimiento agregado crece rápidamente, pero esto se debe a dos factores. Por un lado, se asignan más usuarios al nivel 2, pero su SINR es tan baja que no pueden transmitir, por lo que hay más recursos para usar en la MBS. Por otro lado, la fuente principal de interferencia (la MBS) está menos cargada, por lo que la interferencia del área disminuye y, en consecuencia, la condición para la transmisión es mejor.

La imparcialidad se muestra en la Fig. 2.35, la cual está estrechamente relacionada con los usuarios de baja tasa. Para los tres casos, la imparcialidad óptima se logra cuando los usuarios de baja tasa están por debajo del 13 %. Al principio, la mayoría de los usuarios mejoran su rendimiento y otros usuarios pueden estar apagados, pero en general el sistema es más justo. Cuando se alcanza la imparcialidad óptima, más usuarios se convierten en usuarios de baja tasa y la imparcialidad disminuye.

2.3. Desarrollo de una herramienta de simulación para LTE-Advanced

Disponer de herramientas de simulación que permitan analizar las prestaciones de los sistemas de comunicaciones móviles es muy importante dado que permiten realizar pruebas reproducibles con un bajo coste asociado.

Por una parte, es importante aplicar un modelo lo suficientemente exacto como para que se contemplen las principales características del sistema, pero también debe ser lo más simplificado posible a fin de que el tiempo de simulación no sea excesivamente elevado. Esto exige, por tanto, un compromiso entre complejidad del modelo a utilizar y eficiencia en cuanto a tiempo de simulación.

En la literatura, existen distintas herramientas de simulación para sistemas LTE-A que cumplen este objetivo. Concretamente, en [56] se presenta un simulador LTE implementado en Matlab. Esta herramienta consiste en realidad en dos simuladores independientes: uno de enlace, que permite obtener tablas con resultados de capa física; y otro simulador de sistema, que usa estos resultados almacenados previamente en tablas. De este modo se consigue abstraer la capa física en las simulaciones que involucran capas superiores.

Otro ejemplo se encuentra en [57], donde se presenta la herramienta SimuLTE. Esta herramienta ofrece gran flexibilidad, ya que al estar basado en OMNeT++ se le pueden añadir distintos módulos. Como por ejemplo INET, el cual permite simular aplicaciones en tiempo real que usan diversas tecnologías de acceso radio. También cabe mencionar las herramientas presentadas en [58] y [59]. La primera de ellas, desarrollada en C++, contempla una gran cantidad de funcionalidades tales como movilidad, handover, diferentes planificadores, etc. La segunda, basada en ns-3, ha sido diseñada para permitir una simulación extremo a extremo asumiendo modelos de tráfico realistas.

Una característica común a los citados simuladores es que todos ellos aplican una estrategia de simulación de la capa física basada en tablas. De esta forma, se consigue reducir los tiempos de simulación, pero da lugar a una pérdida de precisión que deriva del mecanismo de abstracción.

El simulador Wireless Mobile SIMulation Advanced (WM-SIMA) [60] [61] ha sido diseñado para simular con gran detalle la capa física (PHY) y de acceso al medio (MAC) sin usar tablas que abstraigan la capa física. WM-SIMA considera el enlace descendente (DL) de un sistema multi-celda Long Term Evolution Advanced (LTE-A) en modo duplexación por división en frecuencia (FDD) considerando el acceso múltiple por división de frecuencias ortogonales (OFDMA) como técnica de acceso al medio [62]. Se trata de una evolución de la herramienta WM-SIM [63] en cuanto a que soporta esquemas de transmisión con múltiples antenas (Multiple Input Multiple Outputs, MIMO) con configuraciones de hasta 8x8, y con posibilidad de considerar varias Components Carriers (CCs) si se activa la opción de Carrier Aggregation (CA), entre otras nuevas funcionalidades.

2.3.1. Arquitectura del Simulador

Podemos distinguir dos niveles en la arquitectura del simulador: *nivel de sistema* y *nivel de enlace*, los cuales se detallan a continuación.

Nivel de sistema

Este nivel, implementado en Matlab, proporciona la información de despliegue de la red móvil a simular. Para ello, utiliza una serie de parámetros configurables como son número de terminales móviles (UEs), número y tipo de puntos de acceso (APs), potencia de transmisión de APs, área de celda, entre otros.

Se distinguen dos tipos de APs: macro (MAPs) y pico (PAPs). De esta forma, es posible considerar redes homogéneas donde sólo se tienen MAPs; o redes heterogéneas compuestas por ambos tipos de APs. En este caso, cada tipo de AP tendrá asociadas unas características en cuanto a potencia de transmisión y pérdidas de propagación. Otro aspecto de la red configurable es el tipo de grid a aplicar, que puede elegirse hexagonal o aleatorio, para el cual las celdas se sitúan siguiendo una distribución espacial uniforme.

Por otra parte, es posible indicarle al *nivel de enlace* que simule sólo un subconjunto de los APs y UEs que componen la red reduciendo, por tanto, el coste computacional de simulación. El subconjunto simulado estará compuesto por aquellos MAPs que se encuentren en la región central del área total definida por la red, así como los elementos que tenga asociados: UEs y PAPs. El efecto del resto de elementos no simulados en la red se modelará como interferencia Gaussiana.

Una vez posicionados los UEs y APs, se obtiene la siguiente información:

- Asociación de UEs a APs, que se puede llevar a cabo en base a dos criterios: pérdidas de propagación mínima o mayor potencia media recibida. Si se elige este último cuando se tiene un escenario heterogéneo, es posible aplicar un incremento (sesgo) sobre la potencia de transmisión de las PAPs de manera que se amplia su área de cobertura. Esta técnica, conocida como *Cell Range Expansion* (CRE) [53], consigue reducir la carga asociada a las MAPs puesto que al aplicar el sesgo se fuerza a que UEs que estarían asociados a una MAP se asocien a una PAP. Ahora bien, estos usuarios forzados o CRE UEs, suelen tener unos valores de nivel de señal a interferencia (SINR) muy degradados debido a que la MAP a la que estarían asociados de forma natural actúa ahora como interferente.

- Potencia media interferente que recibe cada elemento simulado desde el resto de elementos (UEs y APs) no simulados que componen la red celular.

- Ganancia del camino entre cada UE y AP simulado.

- Nivel de SINR medio para cada UE simulado.

- Potencia media de transmisión de cada AP simulado.

Toda esta información se almacena en un fichero .mat que será utilizado posteriormente por la parte de *nivel de enlace* para realizar la simulación.

Nivel de enlace

Este nivel, basado en C++, lleva a cabo el procesado de las capas PHY y MAC para la configuración de red establecida. La Fig. 2.36 muestra la arquitectura del simulador asociada a este nivel. Se observa cómo la herramienta se compone de cinco grandes bloques: transmisor (*Trasmitter*), receptor (*Receiver*), canal radio

MIMO (*Extended MIMO Channel*), planificador cross-layer (*MAC Scheduler*) y medidas (*QoS Statistics*).

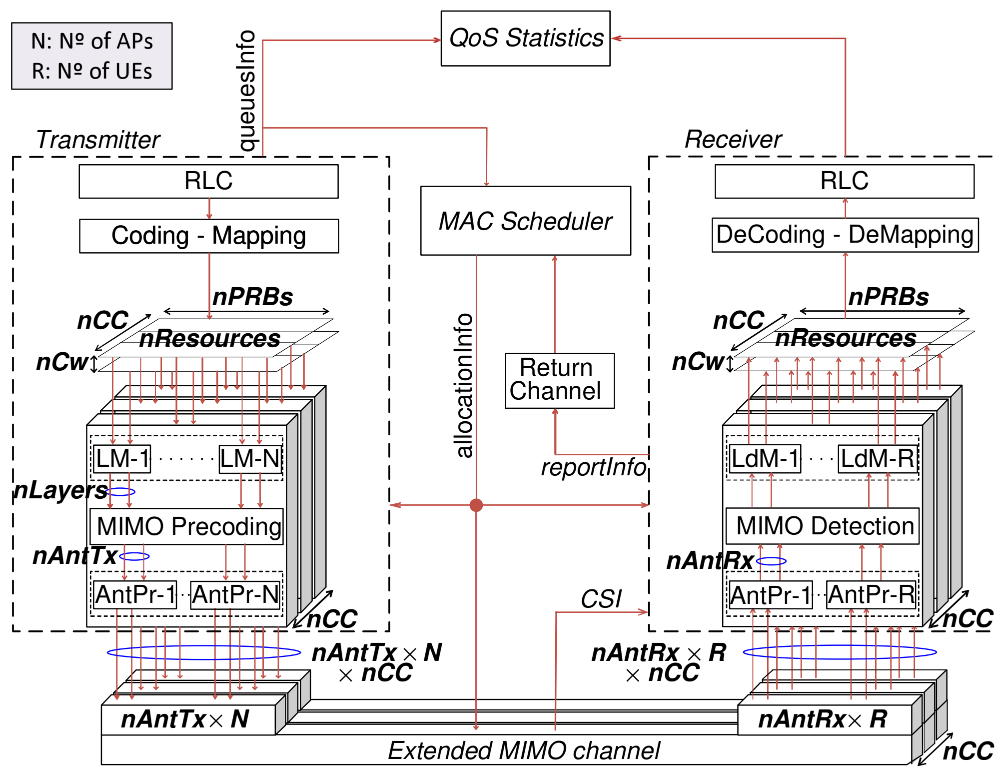


Figura 2.36: WM-SIMA: Arquitectura de nivel de enlace

Transmitter Es el encargado de generar los paquetes RLC (Radio Link Control), encolarlos y segmentarlos en bloques de transporte (TBs) en base a la información que recibe del planificador ($allocationInfo$). Para ello se dispone de 3 modelos de fuentes: *Full buffer*, en la que los UEs siempre tienen información para transmitir; *Streaming*, según una distribución de Pareto; o *Trazas*, que se cargan de un fichero .mat. Sobre los TBs generados, se aplica un esquema de modulación y codificación adaptativo para obtener dinámicamente los símbolos complejos que se mapearán sobre los recursos físicos disponibles. El número total de recursos ($nResources$) dependerá del ancho de banda considerado ($nPRBs$, número de Physical Resource Blocks), del número de CC (nCC) y de la configuración de antenas, que determina el número de *codewords*[1] (Cw) disponibles (nCw). A continuación, realiza un

[1]En este trabajo el término *codeword* hace referencia a un TB codificado y modulado de forma

procesado MIMO que implica: un mapeo de Cw a *layers*, realizado por el bloque Layer Mapping (LM) asociado a cada AP; y una precodificación MIMO, si procede. Por último, se realiza un procesado de antenas ($AntPro$) para la generación de los símbolos OFDM que se transmiten sobre el canal radio.

Receiver Realiza el proceso inverso al llevado a cabo en el transmisor con el objetivo de recuperar los TBs enviados. En efecto, a partir de los símbolos OFDM recibidos recupera la información transmitida haciendo uso tanto de la señalización ($allocationInfo$) como de la estimación de canal que realiza el propio bloque. Esta estimación puede ser: ideal, coincidiendo con el reporte recibido del canal (CSI); o estimada, a partir de los pilotos en la subtrama LTE-A.

Extended MIMO channel Modela el efecto que tiene el canal de comunicaciones móviles inalámbrico. Para ello, genera tantos canales independientes como caminos haya entre cada UE y AP simulados dependiendo de la configuración de antenas elegida. En cada uno de ellos se modelan los siguientes efectos:

1. Desvanecimiento selectivo en frecuencia, mediante un modelo de línea de retardo multitrayecto en el que la ganancia y retardo asociado a cada trayecto son configurables. Las muestras del canal pueden seguir una distribución Rayleigh o Rice. En caso de que se tenga una configuración MIMO, se utiliza el modelo de Kronecker [64] para determinar la correlación espacial entre antenas.

2. Interferencia, que se añade sobre la señal recibida en cada receptor desde el subconjunto de transmisores simulados.

3. Ruido Gaussino blanco (AWGN), que para cada receptor añade ruido Gaussiano en cada una de sus antenas. Hay que recordar que la potencia de ruido aquí añadida tiene en cuenta la interferencia del resto de elementos no simulados que componen la red.

Por último, cabe destacar que existe la posibilidad de que este bloque funcione en el dominio del tiempo, de manera que se simula el canal con gran exactitud, o de la frecuencia, con lo que los símbolos OFDM se ven afectados directamente por la respuesta en frecuencia del canal simulado. En caso de simular el canal en frecuencia

independiente.

se reduce el tiempo de simulación a costa de asumir que la transmisión está libre de interferencia entre símbolos (ISI) y entre portadoras (ICI).

MAC Scheduler Se encarga de llevar a cabo la planificación en cada intervalo de transmisión (TTI) para repartir los recursos tiempo-frecuencia entre los UEs en función del algoritmo seleccionado. Para ello tiene en cuenta la información que recibe por usuario basada en el estado de las colas ($queuesInfo$) y en el estado de los procesos HARQ y el canal radio ($reportInfo$).

Independientemente del algoritmo, se ha optado por dar prioridad a las retransmisiones pendientes frente a la transmisión de nueva información. En efecto, en cada TTI se intentan asignar recursos primero a las retransmisiones asociadas a cada UE. Para ello, se aplica un criterio de búsqueda cíclico sobre los usuarios con retransmisiones pendientes para determinar el siguiente UE a tratar. Una vez seleccionado, el planificador determina los recursos concretos sobre los que se hará la retransmisión. Tras planificar las retrasmisiones, los recursos restantes se asignan a la transmisión de nueva información. En este caso, el planificador determina tanto el usuario a tratar como los recursos concretos a asignarle.

Cabe destacar que la asignación de recursos se hace aplicando el Tipo 2 de los posibles criterios de *Resource Allocation* definidos por el 3GPP para el DL. Así, en cada *codeword*, los PRBs se asignan en bloques de PRBs consecutivos cuyo tamaño será múltiplo del tamaño mínimo de asignación asociado al ancho de banda de simulación [65].

Los algoritmos de planificación implementados son: Round Robin, Proportional Fair, Channel Dependent Earliest Deadline Due (CDD-EDD), CDD-EDD with postponent term y Opportunistic Hard Priority.

QoS Statistics Genera estadísticos asociados a la simulación tales como tasa de error de bit (Bit Error Rate, BER), de bloque (BLock Error Rate, BLER), throughput, información relacionada con las colas de usuario, etc. Para ello tiene en cuenta la información que recibe del receptor y del estado de las colas de los UEs ($queuesInfo$). Los estadísticos obtenidos se almacenan en un fichero .mat al finalizar la simulación, de manera que es posible hacer un procesado $offline$ de los mismos.

Capítulo 3

Conclusiones y Líneas Futuras de Trabajo

3.1. Conclusiones

En esta tesis se ha realizado un análisis de técnicas de gestión de recursos radio sobre redes celulares LTE-Advanced.

El primer bloque de técnicas está centrado en algoritmos de planificación de recursos radio, comenzado por aquellos que tienen en cuenta requisitos de retardo del tráfico. Los resultados de la simulación muestran que, para valores bajos de SNR (entre 0 y 15 dB), los tres algoritmos evaluados son capaces de reducir considerablemente el retardo medio y el percentil 95 del retardo a costa de descartar aquellos paquetes que excedan un cierto umbral. Para valores altos de SNR (de 15 a 30 dB), los resultados de retardo están influenciados principalmente por la función de utilidad de cada algoritmo. En este sentido, el algoritmo denominado *CD-EDD with postponed EDD term* logra el mejor rendimiento.

Posteriormente, se ha analizado la imparcialidad entre usuarios para un algoritmo PF con criterios de SNR y de tasa de transmisión, obteniendo expresiones de forma cerrada para la distribución de la SNR por usuario y del sistema para este segundo caso. Los resultados muestran diferencias notables en términos de distribución de probabilidad asociada con la SNR por usuario y por sistema. En particular, los resultados muestran que el criterio de PF basado en la tasa siempre es más justo que el basado en la SNR (especialmente a medida que aumenta el número de usuarios en la celda) aunque a costa de degradar la tasa de transmisión total del sistema.

También se ha abordado la gestión de recursos para una arquitectura C-RAN. En particular, se ha analizado el impacto del retardo en el informe H-ARQ sobre el rendimiento del usuario. Los resultados muestran que la tasa neta de transmisión disminuye cuando el retardo del informe H-ARQ es mayor que el tiempo requerido para la ejecución de los procesos HARQ.

El segundo bloque de técnicas se ha centrado en aquellas orientadas a la gestión de interferencias. En particular, se ha comenzado por el análisis del rendimiento del sistema en un despliegue celular basado en FFR cuando se utilizan tres esquemas de asignación de recursos diferentes: ERS, ETR, ETR Truncado. Se ha propuesto un método para determinar la tasa total de transmisión y el ancho de banda óptimo de partición de cada celda bajo cada estrategia de planificación de recursos. Se demuestra que la tasa total de tranmisión es máxima para la estrategia ERS mientras que ETR consigue la más baja. El análisis de imparcialidad muestra que la estrategia más justa es la política de ETR, que logra el menor valor del índice de Gini a costa de disminuir la tasa media de transmisión. Dicho compromiso puede solventarse usando la política ETR truncada a través del parámetro ϵ.

Posteriormente se han analizado las técnicas coordinadas para la gestión de interferencias cuando se aplica desde varias puntos de acceso (CoMP). En particular, se ha evaluado la técnica de Transmisión Conjunta (JT) sin precodificación y se ha comparado con la técnicas PFR para la transmisión de tráfico M2M. Los resultados han mostrado que ambas técnicas mejoran el rendimiento de los usuarios del borde de la celda en comparación con la configuración de referencia. Además, la técnica PFR presenta mejor resultado en términos de retardo y tasa media de transmisión para el tipo de tráfico considerado. De hecho, PFR mejora significativamente los resultados de retardo tanto para la celda observada como para los usuarios ubicados en el borde de la celda. Teniendo en cuenta los resultados, se puede concluir que el uso de JT sin precodificación no es adecuado cuando se considera el tráfico M2M.

En último lugar, se aborda una de las técnicas ICIC más importantes sobre redes heterogéneas: Expansión del Rango de la Celda (CRE). Se ha analizado la influencia del valor del sesgo asociado a la técnica CRE sobre las prestaciones medias de una red LTE-A heterogénea considerando varias densidades de PAPs. Los resultados obtenidos muestran que un aumento del sesgo produce siempre una descarga de la MAP mientras que existe un valor óptimo del mismo en términos de eficiencia espectral. Dicho valor óptimo no depende de la densidad de PAPs considerada. Ahora bien, un incremento del número de PAPs sí implica un incremento en el rendimiento medio de la red en términos de $bits/s/Hz/m^2$.

Finalmente, se ha presentado la herramienta WM-SIMA, la cual permite simular el enlace descendente de un sistema de comunicaciones móviles LTE-A. Para ello, se han implementado las principales funcionalidades asociadas a la capa PHY y MAC de este tipo de sistemas. Ha quedado reflejado tanto la configurabilidad como la variedad de resultados que permite obtener WM-SIMA.

3.2. Líneas Futuras de Trabajo

A continuación se describen algunas propuestas de futuras líneas de trabajo:

- Una de los cuestiones clave en la evolución de las redes celulares hacia 5G es el cambio a una banda de frecuencias más alta, en ondas milimétricas (mmW), lo que plantea desafíos adicionales. Uno de los desafíos es la gestión de las interferencias en redes basadas en mmW. Entre las técnicas que se barajan para estos escenarios se encuentran: la gestión dinámica de la potencia de transmisión (bajo demanda), así como el silenciamiento coordinado, la supresión de la señal interferente o el uso inteligente de múltiples antenas.

- Los sistemas inalámbricos de nueva generación (como 5G) están pensados para abarcar un amplio conjunto de escenarios con requisitos muy diversos, como son: muy baja latencia y alta fiabilidad, muy alta densidad de usuarios, muy alta eficiencia espectral, etc. Por este motivo, deben surgir nuevos algoritmos de reparto de recursos (*scheduling*) que se adapten a dichos requisitos y que aborden macro-celdas y pico/fempto-celdas. Una segunda línea de investigación podría ir enfocada a algoritmos específicos para cada escenario.

- Recientemente se han puesto en auge los algoritmos de transmisión de contenido multimedia adaptativos a la capacidad. Una última línea de investigación podría ir enfocada a la definición y evaluación de nuevos algoritmos de reparto de recursos en la capa de Control de Acceso al Medio (MAC) que tengan en cuenta la información de la capa de aplicación (*cross-layer*) con objeto de optimizar dicha adaptación y mejorar la calidad de experiencia final del usuario. Para ello, haría falta definir una capa encargada de traducir el impacto de los distintos indicadores de calidad en la red sobre la calidad de experiencia final del usuario.

Bibliografía

[1] 3GPP TS 36.201 v8.0.0, ”Physical Layer. General Description (Release 8)”, 2009.

[2] 3GPP TS 36.201 v10.0.0, “Physical Layer. General Description (Release 10)”, 2010.

[3] A. J. Goldsmith, “Wireless Communications”, Cambridge University Press, 2005.

[4] D. Gesbert, S. Hanly, H. Huang, S. Shamai Shitz, O. Simeone, Wei Yu, ”Multi-Cell MIMO Cooperative Networks: A New Look at Interference,” IEEE Journal on Selected Areas in Communications, vol.28, no.9, pp.1380-1408, Dec 2010.

[5] Shaohui Sun, Qiubin Gao, Ying Peng, Yingmin Wang, Lingyang Song, ”Interference management through CoMP in 3GPP LTE-advanced networks,” IEEE Wireless Communications, vol. 20, no. 1, pp. 59-66, 2013.

[6] S. Landström, A. Furuskar, K. Johansson, L. Falconetti, F. Kronestedt, ”Heterogeneous networks – Increasing cellular capacity”, Ericsson review, 2011.

[7] J.T. Entrambasaguas, M.C. Aguayo-Torres, G. Gómez, J.F. Paris, ”Multi-user capacity and fairness evaluation of channel/QoS aware multiplexing algorithms,” IEEE Network, Vol. 21, Issue 3, May 2007.

[8] Jun Cai, Xuemin Shen, Jon W. Mark, ”Downlink Resource Management for Packet Transmission in OFDM Wireless Communication Systems,” IEEE Transactions on Wireless Communications, Vol.4, No.4, Jul 2005.

[9] W. Ajib, D. Haccoun, "An Overview of Scheduling Algorithms in MIMO-based Fourth Generation Wireless Systems," IEEE Network, pp.43-48, Sep 2005.

[10] Q. Liu, S. Zhou, G. B. Giannakis, "Cross-Layer Scheduling with Prescribed QoS Guarantees in Adaptive Wireless Networks," IEEE Journal on Selected Areas in Communications, Vol.23, No.5, May 2005.

[11] T.D. Novlan, R.K. Ganti, A. Ghosh, J.G. Andrews, "Analytical Evaluation of Fractional Frequency Reuse for OFDMA Cellular Networks", IEEE Transactions on Wireless Communications, Vol.10, Issue 12, pp.4294–4305, Dec 2011.

[12] B. Soret, Wang Hua,K.I. Pedersen, C. Rosa, "Multicell cooperation for LTE-advanced heterogeneous network scenarios," IEEE Wireless Communications magazine, vol.20, no.1, pp.27-34, Feb 2013.

[13] K.I. Pedersen, Wang Yuanye, B. Soret, F. Frederiksen, "eICIC Functionality and Performance for LTE HetNet Co-Channel Deployments," IEEE Vehicular Technology Conference (VTC Fall), pp.1-5, Sep 2012.

[14] M. Cierny, H. Wang, R. Wichman, Z. Ding, C. Wijting, "On Number of Almost Blank Subframes in Heterogeneous Cellular Networks," IEEE Transactions on Wireless Communications, pp.1-13, Sep 2013.

[15] 3GPP TS 23.107 V8.2.0, "3rd Generation Partnership Project; Technical Specification Group Services and System Aspects; Quality of Service (QoS) concept and architecture (Release 8)", 2011.

[16] H. Luo, S. Ci, D. Wu, J. Wu, H. Tang, "Quality-driven cross-layer optimized video delivery over LTE". IEEE Communications Magazine, vol. 48, no. 2, pp. 102-109, February, 2010.

[17] B. Sadiq, R. Madan, A. Sampath, "Downlink Scheduling for Multiclass Traffic in LTE". EURASIP Journal on Wireless Communications and Networking, 2009.

[18] Y. Qian, C. Ren, S. Tang, M. Chen, "Multi-service QoS guaranteed based downlink cross-layer resource block allocation algorithm in LTE systems". Interna-

tional Conference on Wireless Communications and Signal Processing, pp. 1-4, November, 2009.

[19] 3GPP TS 23.203 V11.3.0, "3rd Generation Partnership Project; Technical Specification Group Services and System Aspects; Policy and charging control architecture (Release 11)", 2011.

[20] Viswanath, P., Tse, D. N. C., Laroia, R., "Opportunistic beamforming using dumb antennas", IEEE Transactions on Information Theory, 48(6), pp. 1277–1294, 2002.

[21] Shakkottai, S., Stolyar, A. L., "Scheduling algorithms for a mixture of real-time and non-realtime data in HDR", In Proceedings of 17th international teletraffic congress, Vol. 4, pp. 793-804, 2001.

[22] Valkova-Jarvis, Z. V., Stoynov, V. R., "Novel throughput quantile averaging methods for the proportional fair algorithm in indoor mobile networks", In 2017 Advances in wireless and optical communications (RTUWO), Riga, pp. 198-202, 2017.

[23] Liu, E., Leung, K., "Expected throughput of the proportional fair scheduling over Rayleigh fading channels", IEEE Communications Letters, 14(6), pp. 515–517, 2010.

[24] Pérez, J. and *et. al.*, "Performance analysis of SNR-based scheduling policies in asymmetric broadcast ergodic fading channel", EURASIP J. Wirel. Com. Netw., vol. 8, pp. 1-6, 2009.

[25] Yang, L., Alouini, M.-S., "Performance analysis of multiuser selection diversity", IEEE Trans. on Vehic. Technology, vol. 55, no. 6, pp. 1848-1861, 2006.

[26] D. Parruca, M. Grysla, S. Gortzen and J. Gross, "Analytical Model of Proportional Fair Scheduling in Interference-Limited OFDMA/LTE Networks," 2013 IEEE 78th Vehicular Technology Conference (VTC Fall), Las Vegas, NV, pp. 1-7, 2013.

[27] M. Kang, Y. J. Sang, H. G. Hwang, H. Y. Lee and K. S. Kim, "Performance Analysis of Proportional Fair Scheduling with Partial Feedback Information for Multiuser Multicarrier Systems," VTC Spring 2009 - IEEE 69th Vehicular Technology Conference, Barcelona, pp. 1-5, 2009.

[28] A. Senst, P. S. Rittich, G. Ascheid, and H. Meyr, "On the throughput of proportional fair scheduling with opportunistic beamforming for continuous fading states," in Proc. IEEE Vehic. Techn. Confer. (VTC), Vol. 1, pp. 300-304, Los Angeles, CA, 2004.

[29] N. Sharma and L. H. Ozarow, "A study of opportunism for multiple-antenna systems," in IEEE Trans. Inf. Theory, Vol. 51, no. 5, pp. 1804-1814, 2005.

[30] L. Yang, M. Kang, and M.-S. Alouini, "On the capacity-fairness tradeoff in multiuser diversity systems," in IEEE Trans. Veh. Technol., Vol. 56, no. 4, pp. 1901-1907, 2007.

[31] H. Tabassum, F. Yilmaz, Z. Dawy and M. S. Alouini, "On the modeling of uplink inter-cell interference based on proportional fair scheduling," 2012 IEEE Wireless Communications and Networking Conference Workshops (WCNCW), Paris, pp. 317-321, 2012.

[32] D. Parruca, J. Gross, "Throughput Analysis of Proportional Fair Scheduling for Sparse and Ultra-Dense Interference-Limited OFDMA/LTE Networks", IEEE Transactions on Wireless Communications, vol. 15, pp. 6857-6870, 2016.

[33] Isabel M. Delgado-Luque, M. Carmen Aguayo-Torres, Gerardo Gomez, Francisco J. Martin-Vega, Jose T. Entrambasaguas, "SNR- Versus Rate-based Proportional Fair Scheduling in Rayleigh Fading Channels", Wireless Personal Communications, pp. 1-13, 2018.

[34] IST-WINNER II Deliverable 1.1.2 v.1.2, "WINNER II Channel Models", IST-WINNER2, Tech. Rep., 2007 (available online: http://www.cept.org/files/1050/documents/winner2%20-%20final%20report.pdf)

[35] Sezer, S., Scott-Hayward, S., Chouhan, P. K., Fraser, B., Lake, D., Finnegan, J., Viljoen, N., Miller, M., Rao, N., "Are we ready for SDN? Implementation challenges for software-defined networks", IEEE Communications Magazine, 51(7), pp. 36-43, 2013.

[36] Mijumbi, R., Serrat, J., Gorricho, J. L., Bouten, N., Turck, F. D., Boutaba, R., "Network function virtualization: State-of-the-art and research challenges", IEEE Communications Surveys Tutorials, 18(1), 236-262, 2016.

[37] Wu, J., Zhang, Z., Hong, Y., Wen, Y., "Cloud radio access network (C-RAN): a primer", IEEE Network, 29(1), pp. 35-41, 2015.

[38] 3GPP TR 38.801 v2.0.0. Study on new radio access technology: Radio access architecture and interfaces (Release 14). http://www.3gpp.org/ftp/Specs/archive/38_series/38.801/38801-200.zip. Accessed 28 May 2018.

[39] NGMN Forum. Further study on critical C-RAN technologies. Technical Deliverable. Available at: https://www.ngmn.org/fileadmin/user_upload/NGMN_RANEV_D2_Further_Study_on_Critical_C-RAN_Technologes_v1.0.pdf. Accessed 28 May 2018.

[40] Rost, P., Bernardos, C., Domenico, A., Girolamo, M., Lalam, M., Maeder, A., Sabella, D., Wuebben, D., "Cloud technologies for flexible 5G radio access networks", IEEE Commun Magazine, 52(5), pp. 68-76, 2014.

[41] 3GPP TS 36.321 v15.1.0. Medium Access Control (MAC) protocol specification (Release 15). http://www.3gpp.org/ftp/specs/archive/36_series/36.321/36321-f10.zip. Accessed 28 May 2018.

[42] 3GPP TS 36.213 v15.1.0. Physical layer procedures (Release 15). http://www.3gpp.org/ftp/specs/archive/36_series/36.213/36213-f10.zip. Accessed 28 May 2018.

[43] ITU-R M.2135-1, "Guidelines for evaluation of radio interface technologies for IMT-Advanced", https://www.itu.int/dms_pub/itu-r/opb/rep/R-REP-M.2135-1-2009-PDF-E.pdf. Accessed 28 May 2018.

[44] Blanquez-Casado, F., Gomez, G., Aguayo-Torres, M. C., Entrambasaguas, J. T., "eOLLA: an enhanced outer loop link adaptation for cellular networks", EURASIP Journal on Wireless Communications and Networking, 20, 2016.

[45] Liu, X., Chong, E. K. P., Shroff, N. B. "A framework for opportunistic scheduling in wireless networks", Computer Networks, 41(4), pp. 451-474, 2002.

[46] Abdel-Baset Hamza *et al.*, "Inter-Cell Interference Coordination", In 4G++ Project: Advanced Performance Boosting Techniques in 4th Generation Wireless Systems, Deliverable D4.1, 2011.

[47] Bilio, D., Bouras, C., Kokkinos, V., Papazois, A., Tseliou, G., "Optimization of Fractional Frequency Reuse in Long Term Evolution Networks", In: Proceedings of IEEE Wireless Communications and Networking Conference, pp. 1853-1857, 2012.

[48] Necker, M.C., "Coordinated Fractional Frequency Reuse", Proceedings of ACM Symposium on Modelling, Analysis, and simulation of Wireless and Mobile Systems, pp. 296-305, 2007.

[49] Ali, S.H., Leung, V.C.M., "Dynamic Frequency Allocation in Fractional Frequency Reused OFDMA Networks", IEEE Transactions on Wireless Communications, vol. 8, issue 8, pp. 4286-4295, 2009.

[50] Barnett, N.S., Dragomir, S.S., "Advances in Inequalities from Probability Theory and Statistics", Nova Science Publishers, 2008.

[51] Shi, H., Seghu, H., Kanhere, S.S., "An Evaluation of Fair Packet Schedulers using a novel measure of Instantaneous Fairness", Compute Communications, vol. 28, pp. 1925-1937, 2005.

[52] A. Alsawah, I. Fialkow, "Optimal frequency-reuse partitioning for ubiquitous coverage in cellular systems," in 16th European Signal Processing Conference - publi-etis.ensea.fr, 2008.

[53] S. Singh and J. G. Andrews, ”Joint Resource Partitioning and Offloading in Heterogeneous Cellular Networks”, IEEE Trans. on Wir. Comm., vol. 13, no. 2, pp. 888-901, Feb. 2014.

[54] 3GPP, Evolved Universal Terrestrial Radio Access (E-UTRA); Further advancements for E-UTRA physical layer aspects. TR 36.814, 3rd Generation Partnership Project (3GPP), 2017.

[55] 3GPP, Technical Specification Group Radio Access Network; Coordinated multipoint operation for LTE physical layer aspects. TR 36.819, 3rd Generation Partnership Project (3GPP), 2013.

[56] Christian Mehlfuhrer, and *et al.* “The Vienna LTE simulators - Enabling reproducibility in wireless communications research in *EURASIP J. Wirel. Commun. Netw.,* Dec. 2011.

[57] A. Virdis, G. Stea and G. Nardini, “SimuLTE - A modular system-level simulator for LTE/LTE-A networks based on OMNeT++ ” 2014 *4th International Conference On Simulation And Modeling Methodologies, Technologies And Applications*, Vienna, Austria, pp. 59-70, 2014.

[58] G. Piro, and *et al.* “Simulating LTE Cellular Systems: An Open-Source Framework,” *in IEEE Transactions on Vehicular Technology*, vol. 60, no. 2, pp. 498-513, Feb. 2011.

[59] N. Baldo, and *et al.*, “An Open Source Product-Oriented LTE Network Simulator based on ns-3, ” in *Proc. 14th ACM International Conference of Modeling, Analysis and Simulation of Wireless and Mobile Systems,* pp. 293-298, Nov. 2011.

[60] *WM-SIM homepage,* [Online]. Available at:
http://riuma.uma.es/xmlui/handle/10630/7438

[61] Alberto Carreras, Isabel M. Delgado-Luque, Francisco J. Martin-Vega, Gerardo Gomez, Mari Carmen Aguayo-Torres, J. Tomas Entrambasaguas, ”A System-Level Simulator for the Downlink of LTE-A. Case of study: cell-offloading in

HetNets", Wireless Personal Communications, Vol. 100, Issue 1, pp. 177–191, 2018.

[62] *3GPP LTE-A Release 10 Specification,* [Online]. Available: http://www.3gpp.org/specifications/releases/70-release-10

[63] G. Gómez and *et al.*, "A Next Generation Wireless Simulator Based on MIMO-OFDM: LTE Case Study", EURASIP J. Wirel. Commun. Netw., vol. 2010, no. 15, pp. 1-12, Apr. 2010.

[64] D.-S. Shui, and *et al.*, "Fading correlation and its effect on the capacity of multielement antenna systems", IEEE Trans. on Comm., vol. 48, no. 3, pp. 502-513, 2000.

[65] 3GPP TS 36.213, "Evolved Universal Terrestrial Radio Access (E-UTRA); Physical layer procedures", Release 10, V10.13.0, Jun. 2015.

Apéndice A

Copia de los Trabajos

En este apéndice se adjunta una copia de las publicaciones que avalan a esta tesis doctoral, las cuales son detalladas a continuación:

- Isabel M. Delgado-Luque, M. Carmen Aguayo-Torres, Gerardo Gomez, Francisco J. Martin-Vega, Jose T. Entrambasaguas, “SNR- Versus Rate-based Proportional Fair Scheduling in Rayleigh Fading Channels”, Wireless Personal Communications, pp. 1-13, 2018

- Alberto Carreras, Isabel M. Delgado-Luque, Francisco J. Martin-Vega, Mari Carmen Aguayo-Torres, Gerardo Gomez, J. Tomas Entrambasaguas, Eneko Atxutegi, Ruben Solozabal, Bego Blanco, Jose Oscar Fajardo, Fidel Liberal, “Impact of Front-Haul Delays in Non-Ideal Cloud Radio Access Networks”, Wireless Personal Communications, 2018

- Alberto Carreras, Isabel M. Delgado-Luque, Francisco J. Martin-Vega, Gerardo Gomez, Mari Carmen Aguayo-Torres, J. Tomas Entrambasaguas, “A System-Level Simulator for the Downlink of LTE-A. Case of study: cell-offloading in HetNets”, Wireless Personal Communications, Vol. 100, Issue 1, pp. 177–191, 2018

- Isabel M. Delgado-Luque, M. Carmen Aguayo-Torres, Gerardo Gomez, Jose T. Entrambasaguas, “A Framework to Evaluate Fairness in Fractional Frequency Reuse Based Cellular Networks”, Wireless Personal Communications, Vol. 95, Issue 2, pp. 287–297, 2017

- I. M. Delgado-Luque, F. Blanquez-Casado, F.J. Martin-Vega, M. Garcia Fuertes, G. Gomez, M. C. Aguayo-Torres, J. T. Entrambasaguas, J. Banos, "Performance evaluation of cooperation-based techniques for M2M traffic over LTE", IEEE 24th International Symposium on Personal Indoor and Mobile Radio Communications (PIMRC), pp. 144-148, Sept. 2013

- I.M. Delgado-Luque, F. Blanquez-Casado, M. Garcia Fuertes, G. Gomez, M.C. Aguayo-Torres, J.T. Entrambasaguas, J. Banos, "Evaluation of Latency-aware Scheduling Techniques for M2M Traffic over LTE", 20th European Signal Processing Conference (EUSIPCO 2012) Bucharest, Romania, pp. 989-993, Aug. 2012

Wireless Personal Communications
https://doi.org/10.1007/s11277-018-5928-6

SNR- Versus Rate-Based Proportional Fair Scheduling in Rayleigh Fading Channels

Isabel M. Delgado-Luque[1] · **M. Carmen Aguayo-Torres**[1] · **Gerardo Gómez**[1] · **Francisco J. Martín-Vega**[1] · **José T. Entrambasaguas**[1]

Abstract
The initial definition of the Proportional Fair scheduling was based on a transmission rate criterion. However, a Signal-to-Noise Ratio (SNR) based definition is commonly used by researchers due to its mathematical tractability, based on the assumption that the performance of both definitions is equivalent. In this work, we derive closed-form expressions for the probability density function of the system output SNR (i.e. received SNR based on the selected scheduling algorithm) and the per-user SNR over Rayleigh fading channels when a rate-based criterion is used. Additionally, we analyze the common assumption of equivalence between both criteria by comparing their per-user performance. Our analysis reveals that their performance are only similar when users experience high values of average SNR. However, significant differences in terms of user fairness have been found in the lower SNR regime.

Keywords Proportional fair · Rate-based scheduling · SNR-based scheduling · Fairness

1 Introduction

The instantaneous capacity over wireless channels changes randomly over time due to the fading of the received signal. The scheduler is then responsible for allocating the radio resources to users considering their instantaneous capacity, among other possible criteria.

In that sense, it is well known that Best Channel (BC) scheduling attains maximum system capacity at the expense of degrading the user throughput fairness [1]. In order to avoid starvation and to improve the user fairness, a modification of previous strategy, called Proportional Fair (PF) scheduling, allocates resources to the user with the best channel conditions relative to its own average. The utility function to be maximized was initially defined using a transmission rate criteria [2–5], where resources are allocated to the user with the largest normalized transmission rate.

✉ Isabel M. Delgado-Luque
imdl@ic.uma.es

Gerardo Gómez
ggomez@ic.uma.es

[1] Departamento de Ingeniería de Comunicaciones, Universidad de Málaga, Campus de Teatinos s/n, 29071 Málaga, Spain

Published online: 27 September 2018

However, many other works use a Signal to Noise Ratio (SNR) criterion, whereby resources are allocated to the user with the largest pondered instantaneous SNR, since it is more mathematically treatable, making it possible to evaluate the per-user [6] and system [7] transmission rates. For instance, an analytical model for the SNR distribution of the scheduled users with PF scheduling assuming a SNR-based criterion is presented in [8]. Kang et al. [9] analyzes the PF scheduling with partial feedback information assuming a SNR-based criterion. The same approach has been assumed in many other works, like in [10, 11] for PF scheduling with multiple antenna techniques, or in [12], which proposes a new hybrid PF scheduler that includes a novel user grouping method. A semi-analytical approach to model the uplink interference considering proportional fair scheduling with SNR-based criterion is shown in [13].

As both criteria (SNR and rate) use the same idea [7], it could be inferred that the SNR version of the PF algorithm provide similar results than the rate-based one. However, to the best of our knowledge, the performance of the rate-based criterion has yet to be analyzed in detail. In a recent work, [14] presents closed-form expressions for the expected throughput of PF scheduling in interference limited scenarios considering the SNR as criterion, although pointing out that, when the rate is used as scheduling metric, one further needs to take into account the SINR-to-rate mapping of the corresponding system.

In this work, we analyze in detail the performance of the Rate-based PF criterion defined in [1]. The main contributions of this paper can be summarized as follows:

- The probability density function (pdf) of the system output SNR and of the per-user SNR considering a shared ergodic Rayleigh channel is obtained for this criterion assuming full side information.
- Results are compared to those obtained in [6, 7] for the SNR-based criterion in order to asses whether or not it is appropriate to assume an equivalent behavior in both schemes. It has been shown that the channel access time for different users is similar for the SNR-based criterion whereas it differs for the rate-based criterion.
- It has been also analytically proved that in the SNR-based PF criterion, the pdf of the per-user SNR only depends on the average SNR of a particular user, whereas in the rate-based criterion, it does not only depend on the individual user's average SNR but also on the average SNR of the rest of the users in the system.

The remainder of this paper is organized as follows. Section 2 describes the system model and general definitions. In Sect. 3, the expressions for the pdf of the per-user and the system SNR associated with a SNR-based PF scheme obtained in [6] and [7] are first presented for convenience; afterwards, we derive closed-form expressions for both pdfs when a rate-based PF scheme is considered. A performance comparison between both criteria is carried out in Sect. 4, both from a theoretical and simulation points of view. Finally, some concluding remarks are presented in Sect. 5.

2 System Model

We consider a single cell where L users (UEs) are communicating with a base station (BS). We assume that both BS and receivers have a single antenna. Independent and identically distributed (i.i.d.) sequence of zero-mean complex Gaussian noise with variance σ_i^2 is assumed at the receivers antennas. The fading channel gain from the BS to the ith user in time slot t is given by $h_i(t)$, where $i = 1..L$. We consider the flat Rayleigh fading model, assuming that the fading coefficients of all users are independent but not necessarily identically distributed. Therefore, $h_i(t)$ is a zero-mean complex Gaussian random variable and its amplitude, $\alpha_i(t) = \sqrt{|h_i(t)|^2}$, is Rayleigh distributed. Thus, the instantaneous (in time slot t) received SNR associated with user i, $\gamma_i(t) = \alpha_i^2(t)/\sigma_i^2$, is exponentially distributed with probability density function given by

$$f_{\gamma_i}(\gamma) = \frac{1}{\overline{\gamma}_i} \exp\left(-\frac{\gamma}{\overline{\gamma}_i}\right), \quad \gamma \geq 0 \tag{1}$$

where $\overline{\gamma}_i$ is the average SNR of the ith user. The cumulative distribution function (cdf) can be easily obtained as

$$F_{\gamma_i}(\gamma) = \Pr(\gamma_i \leq \gamma) = \int_0^u f_{\gamma_i}(\gamma)d\gamma = 1 - \exp\left(-\frac{\gamma}{\overline{\gamma}_i}\right), \quad \gamma \geq 0 \tag{2}$$

2.1 General Definitions

In time slot t, the PF scheduling scheme allocates radio resources to the user i with the best channel condition relative to its own average, i.e. the user with the largest utility, $u_i(t)$. The chosen criterion determines the concrete expression to compute the utility.

Hereinafter, the effective SNR associated with the ith user, $\gamma_i^*(t)$, refers to the per-user SNR of user i, which is 0 if no resources are allocated to the ith user, and is $\gamma_i(t)$, i.e. the instantaneous SNR of user i in time slot t, in case resources are allocated to him, i.e.,

$$\gamma_i^*(t) = \begin{cases} 0, & u_i(t) \leq u_{-i}(t) \\ \gamma_i, & u_i(t) > u_{-i}(t) \end{cases} \tag{3}$$

where $u_{-i}(t) = \max\limits_{k \neq i} \left(u_k(t)\right)$ is the maximum utility function for all users in time slot t, excluding user i.

Moreover, the system output SNR in time slot t is defined as

$$\gamma_s(t) = \gamma_i^*(t) \big| i = \arg \max_{k=1\ldots L} \left(u_k(t)\right) \tag{4}$$

Thus, $\gamma_s(t)$ is obtained from the instantaneous SNR of the user with the maximum utility at each time, i.e. the user for which resources have been allocated.

The potential transmission rate of user i can be computed by the well known Shannon limit as

$$r_i(t) = \log_2\left(1 + \gamma_i(t)\right). \tag{5}$$

This potential rate is only achievable for those time slots when radio resources are actually allocated to user i, that is, if his utility is maximum: $u_{-i}(t) > \max\limits_{k \neq i} \left(u_k(t)\right)$.

The ergodic capacity for user i, $\overline{r}_i$, is obtained by averaging that potential rate with the SNR distribution in (1), and it can be written for the Rayleigh fading as [15]

$$\overline{r}_i = \int_0^\infty r_i(t) f_{\gamma_i}(\gamma) = exp\left(1/\overline{\gamma}_i\right) \cdot E_1\left(1/\overline{\gamma}_i\right) \tag{6}$$

where $E_1(x)$ is the exponential integral.

However, the achievable mean transmission rate by the ith user is obtained by averaging that potential rate with the distribution for the effective SNR, that is,

$$R_i = \int_0^\infty \log_2(1+\gamma) f_{\gamma_i^*}(\gamma) d\gamma \tag{7}$$

where $f_{\gamma_i^*}(\gamma)$ is the pdf of the per-user SNR of the ith user, $\gamma_i^*(t)$.

As only one user is allocated resources at each slot, the average cell rate corresponds to the sum rate, $\sum R_i$. Thus, the sum rate can be also obtained as the average of the Shannon limit for the system SNR, γ_S:

$$R = \int_0^\infty \log_2(1+\gamma) f_{\gamma_S^*}(\gamma) d\gamma. \tag{8}$$

3 Proportional Fair Scheduling Analysis

In this section, a comparative analysis of the SNR- vs. Rate-based PF criteria is given.

3.1 SNR-Based Criterion

If the SNR criterion is assumed, the associated utility to the ith user in time slot t is given by

$$u_i^{SNR}(t) = \gamma_i(t)/\overline{\gamma}_i \tag{9}$$

The per-user SNR and system output SNR pdfs are presented in [6] and [7], respectively:

$$\begin{aligned} f_{\gamma_i^*}^{SNR}(\gamma) &= Pr(u_i^{SNR} < u_{-i}^{SNR})\delta(\gamma) \\ &\quad + \frac{1}{\overline{\gamma}_i} \sum_{k=0}^{L-1} (-1)^k \binom{L-1}{k} \exp\left(-(1+k)\frac{\gamma}{\overline{\gamma}_i}\right) \end{aligned} \tag{10}$$

$$f_{\gamma_s}^{SNR}(\gamma) = \sum_{i=1}^{L} \frac{1}{\overline{\gamma}_i} \cdot \sum_{k=0}^{L-1} (-1)^k \binom{L-1}{k} \exp\left(-(1+k)\frac{\gamma}{\overline{\gamma}_i}\right) \tag{11}$$

where $\delta(\gamma)$ is the Dirac delta function.

3.2 Rate-Based Criterion

In this case, the associated utility to user i in time slot t can be expressed as

$$u_i^{RATE}(t) = r_i(t)/\overline{r}_i \tag{12}$$

where $r_i(t)$ and $\overline{r}_i$ represent the potential transmission rate in time slot t and the ergodic capacity corresponding to the average fading level of the ith user, respectively.

The per-user SNR can be obtained from the normalized transmission rate from (5) and (12):

$$\gamma_i = 2^{u_i^{RATE} \cdot \bar{r}_i} - 1 \tag{13}$$

It should be noted that the original definition of the rate-based PF criterion is given by [2]

$$u_i^{RATE}(t) = r_i(t)/T_i(t) \tag{14}$$

being $T_i(t)$ the moving average throughput of user i in a past window of length t_c. However, both rules are equivalent whenever the moving average throughput $T_i(t)$ used in (14) tends to the average throughput in (12). This is fulfilled if there is a high amount of data to be transmitted, i.e. a full-buffer source model is considered, or if the averaging window to compute the throughput in (14) is long enough, as described in [3]. According to [5], the approximation in (12) is valid for moving average window length $t_c \geq 50$ (used in (14)) with an accuracy greater than 98%.

Since $\gamma_i(t)$ is exponentially distributed for Rayleigh channels, the pdf of the normalized transmission rate of user i, u_i, can be obtained as a continuous random variable change through the Jacobian transformation: the pdf of the SNR γ_i given by (1), evaluated at the inverse of (13), and multiplied by its derivative (i.e., $f_{\gamma_i}(\gamma) \cdot \gamma'$ with γ as given in (13) and γ' its derivative). After easy manipulation, the pdf of u_i can be expressed as

$$f_{u_i^{RATE}}(u) = \bar{r}_i \frac{\ln 2}{\bar{\gamma}_i} 2^{u \cdot \bar{r}_i} \exp\left(-\frac{2^{u \cdot \bar{r}_i} - 1}{\bar{\gamma}_i}\right) \tag{15}$$

Finally, its cdf can be evaluated by integrating (15) or, equivalently, from the cdf of γ_i (2) as:

$$F_{u_i^{RATE}}(u) = \Pr\left(u_i^{RATE} \leq u\right) = \Pr\left(\gamma_i \leq 2^{u \cdot \bar{r}_i} - 1\right) = 1 - \exp\left(-\frac{2^{u \cdot \bar{r}_i} - 1}{\bar{\gamma}_i}\right) \tag{16}$$

As stated in (12), the utility associated with this criterion is based on a transmission rate definition. In order to obtain the pdf of the per-user effective SNR under this criterion, we first obtain the expression of the pdf associated with the normalized effective transmission rate of user i, named as u_i^*, and afterwards, we derive the desired pdf by applying the change of variable expressed in (12) and (5).

The pdf of u_i^* can be computed using a similar process than that presented in [6] and [7]:

$$f_{u_i^*}(u) = \Pr\left(u_i^{RATE} < u_{-i}^{RATE}\right)\delta(u) + f_{u_i^{RATE}}(u)\widetilde{F}_{-i}(u) \tag{17}$$

The first term represents the probability that the effective transmission rate of the ith user is $u_i^* = 0$, i.e. resources are not allocated to user i, what happens as far as the utility function of the user is not the maximum one, u_{-i}^{RATE}. In that case, the effective rate and SNR will be 0 for that user.

The second term in Eq. (17) is the probability that a transmission is done to user i, since this user has the maximum utility, that is, the cumulative distribution function (cdf) of u_{-i}^{RATE} multiplied by the the pdf of the normalized transmission rate of each individual user, $f_{u_i^{RATE}}(u)$ (given by (15)). Transmission will be allocated to user i if the normalized rate for that user is maximum, that is:

$$\widetilde{F}_{-i}(u) = \Pr\left(u_k^{RATE} < u_i^{RATE}, k = 1 \ldots L, k \neq i\right) \tag{18}$$

Being all channels independent, this probability can be written as the product of each probability evaluated as the cdf in (16):

$$\widetilde{F}_{-i}(u) = \prod_{k=1,k\neq i}^{L} F_{u_k^{RATE}}(u) \tag{19}$$

Finally, the pdf of the per-user SNR of user i can be obtained by applying the Jacobian transformation defined by (13), yielding

$$\begin{aligned} f_{\gamma_i^*}^{RATE}(r) =& Pr(u_i^{RATE} < u_{-i}^{RATE})\delta(\gamma) \\ &+ \frac{1}{\bar{\gamma}_i}\exp\left(-\frac{\gamma}{\bar{\gamma}_i}\right)\cdot\prod_{\substack{k=1\\k\neq i}}^{L}\left(1-\exp\left(-\frac{\left((1+\gamma)^{\frac{\bar{r}_k}{\bar{r}_i}}-1\right)}{\bar{\gamma}_k}\right)\right) \end{aligned} \tag{20}$$

The pdf of the system output SNR coincides with the sum of the per user SNR pdf (20) over all users considering only the second term of the equation:

$$f_{\gamma_s}^{RATE}(\gamma) = \sum_{i=1}^{L}\frac{1}{\bar{\gamma}_i}\exp\left(-\frac{\gamma}{\bar{\gamma}_i}\right)\cdot\prod_{\substack{k=1\\k\neq i}}^{L}\left(1-\exp\left(-\frac{\left((1+\gamma)^{\frac{\bar{r}_k}{\bar{r}_i}}-1\right)}{\bar{\gamma}_k}\right)\right) \tag{21}$$

We can derive the analytical expression of the pdf associated with the system output transmission rate, r_s, from it by taking into account (5).

Remark 1 *(Different channel access time)* The first important difference between both criteria is reflected attending to the utility definition. When the SNR-based strategy is assumed, by applying the change of variable defined by its utility in Eq. (9), i.i.d. variables are obtained for the utility functions of all users. Thus, all users are identical when this criterion is used in a system in which a Rayleigh fading channel model is assumed, so the same percentage of channel access time is assigned to each one. It should be noted that it does not mean that all users achieve the same transmission rate, since channel utilization depends on the mean SNR associated with each user. By contrast, the utility functions associated with each user are not i.i.d. variables when the rate-based PF is used (by applying the change of variable defined by its utility in Eq. (12) to obtain (15)). This means that the use of this criterion yields non-identical users, so a different percentage of channel access time is assigned to each one.

Remark 2 *(Dependency with other users' average SNR)* Important differences can be also found in the per-user SNR pdf associated with each strategy. In a SNR-based PF criterion, the pdf of the per-user SNR (Eq. (10)) only depends on the average SNR of the user and on the number of users. However, when a rate-based PF criterion is used, this distribution (Eq. (20)) does not only depend on the individual user's average SNR (and the ergodic capacity) but also on the average SNR (and the ergodic capacity) of the rest of the users in the system. By analyzing Eqs. (10) and (20), it can be seen that both criteria are only equivalent when all users experience the same value of average SNR and, in this case, they both behave as the Best Channel strategy.

4 Performance Comparison between Both PF criteria

In this section we provide a performance comparison between the SNR-based and Rate-based PF scheduling criteria. Performance comparison of both schemes was carried out in terms of the achievable transmission rate of each user, which was computed using Eq. (7).

We consider a scenario in which each user has an average SNR, $\overline{\gamma}_i$, obtained from a lognormal distribution with an average Γ and standard deviation σ. The average associated with the lognormal distribution, Γ, represents the cell average SNR. We have assumed a standard deviation of 4 dB, which is a typical value in micro-cell environments [16].

In principle, two scenarios were considered: a 5-user scenario and a 9-user scenario. In both cases, the value of the average SNR associated with each user, $\overline{\gamma}_i$, was selected so that each of them represents a predefined percentile of the cell average SNR distribution. Specifically, users {1, 2, 3, 4, 5} correspond to the {10th, 30th, 50th, 70th and 90th} percentile in the 5-user scenario, while users {1, 2, 3, 4, 5, 6, 7, 8, 9} correspond to the {10th, 20th, 30th, 40th, 50th, 60th, 70th, 80th and 90th} percentile in the 9-user scenario. Thus, user 1 is associated with the lowest average SNR, whereas user 5 (in the 5 user scenario) or user 9 (in the 9-user scenario) is associated with the highest average SNR.

Figures 1 and 2 show the achievable per-user transmission rate, R_i, as a function of the cell average SNR, Γ, in a 5-user and a 9-user scenario, respectively. R_i has been computed numerically from the corresponding pdf expressions of the user rate and Eq. (7). Analyzing both figures, it can be observed that fairness (in terms of user rate) depends on the scenario considered when a rate-based criterion is used, and particularly, on the number of users. In fact, in both scenarios, rate-based criterion is actually fairer in terms of user rate for low values of Γ. Moreover, the rate-based scheme tends to be fairer as the number of users increases, as it is shown in Fig. 2.

Furthermore, it should be noted the different per-user rate behaviors as the cell average SNR, Γ, increases. If the SNR-based PF criterion is used, the per-user performance for the

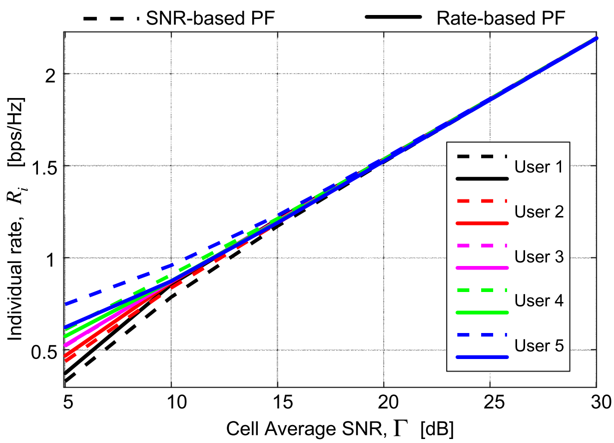


Fig. 1 Per-user average transmission rate in 5-user scenario

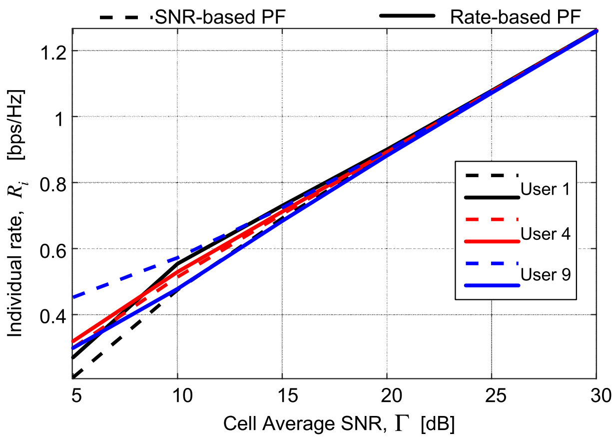


Fig. 2 Per-user average transmission rate in 9-user scenario. For clarity, only 3 users have been represented: best user (user 9), median user (user 4) and worst user (user 1)

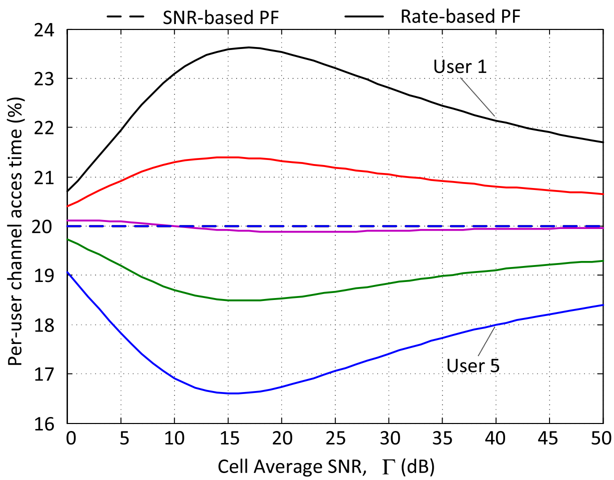


Fig. 3 Per-user channel access time (%) for the 5-users scenario using a SNR-based PF scheme (dotted line) and a rate-based PF scheme (solid line)

set of users is always the same for any value of Γ: higher average SNR $\overline{\gamma}_i$ represents higher allocated rates. However, when the rate-based PF criterion is applied, the user with the highest average SNR might not be the user with the highest allocated rates, but it depends on the value of the cell average SNR. This difference is due to the fact that the SNR-based PF criterion assigns the same percentage of channel access time to all users [6] whereas this percentage depends on the specific scenario if the rate-based PF criterion is used. By analyzing the utility expression, it is easy to see that the utility value is boosted for users with lower average potential rates (i.e. lower average SNR). Thus, they are assigned a higher percentage of channel access time to the extent that they could be allocated higher individual rates than other users with better channel conditions.

Figure 3 depicts the evolution of the per-user channel access time (measured in %) in a 5 users scenario. It can be seen that the SNR-based PF scheme is strictly fair in terms of channel access time. Indeed, all users receive the same percentage of channel access time (20% in this case) so all the curves are superimposed. However, the percentage of channel access time associated with each user is very different when a rate-based PF scheme is used. As it was mentioned before, the channel access time assigned to each user depends

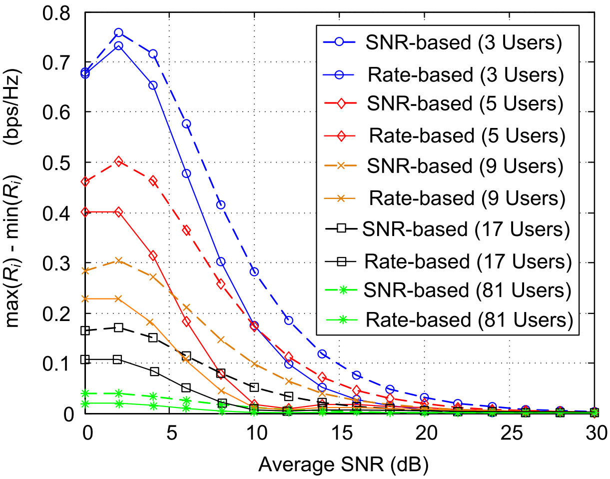


Fig. 4 Difference in the per-user rate (in bps/Hz) between the best and the worst user as a function of the average SNR for different number of users in the cell

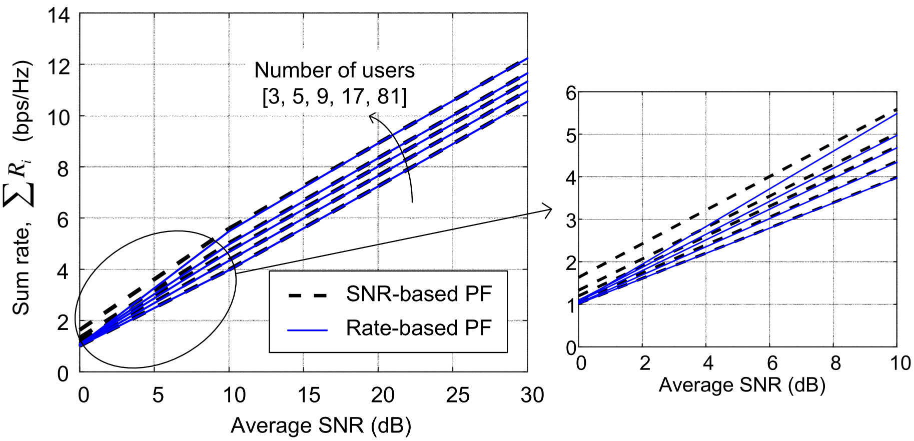


Fig. 5 Sum rate (in bps/Hz) as a function of the average SNR for different number of users in the cell

on its average SNR value, i.e. the higher the average SNR value the lower the percentage of channel access time assigned. This is due to the utility function used in the rate-based PF scheme. Note that individual rate results shown in previous figures already include the impact of the different channel access time allocated to each user. It should be noted that all curves tend to a constant value as the cell average SNR is increased. The reason is that, as the cell average SNR increases, the relative difference between the R_i value associated with each user is lower, as its deviation σ is kept constant. As result, the percentage of channel access time is similar for all users.

Nevertheless, it should be noted that if a more realistic computation of the potential transmission rate is used (e.g., including the impact of a limited set of modulation schemes) this effect occurs at lower values of average SNR.

Fig. 4 shows the difference in the per-user rate (in bps/Hz) between the best and the worst user as a function of the average SNR for different number of users in the cell (from 3 to 81 users). In other words, this figure represents the grade of fairness among users, measured as: $\max(R_i) - \min(R_i)$. As considered before, the value of the average SNR associated with each user, $\overline{\gamma}_i$, was selected so that each of them represents a predefined percentile of the cell average SNR distribution; in particular, the percentiles are set according to the following criteria: [10:40:90]th for 3 users, [10:20:90]th for 5 users, [10:10:90]th for 9 users, [10:5:90]th for 17 users and [10:1:90]th for 81 users. Results show that the rate-based PF criterion is always fairer than the SNR-based. In addition, we observe much higher differences between the best and the worst user under low average SNR values. Finally, it is worth noting that fairness improves as the number of users in the cell increases.

Note that, although the rate-based criterion provides a better fairness than SNR-based, this is achieved at the expense of degrading the sum rate of the system. This is confirmed in Fig. 5, which shows the sum rate (in bps/Hz) as a function of the average SNR for different number of users in the cell (from 3 to 81 users). From the one hand, results show that a higher number of users provide a sum rate gain due to the multi-user diversity gain. From the other hand, it is observed that SNR-based and Rate-based PF algorithms provide similar results only for high average SNR; for low average SNR values, the rate-based criterion achieves a lower sum rate (especially when the number of users is high) in order to improve the fairness.

5 Conclusions

In this work, we derive closed form expressions for the distribution of the per-user SNR and system output SNR for a PF algorithm based on a transmission rate criterion. Afterwards, we use these expressions to compare its performance with the SNR-based PF presented in most of previous works.

Analytic results show noticeable differences in terms of probability distribution associated with the per-user and system SNR. Additionally, the analysis of the simulation results reveals that the treatment that each criterion makes to each user differs for low values of the cell average SNR. In particular, results show that the rate-based PF criterion is always fairer than the SNR-based (specially as the number of users in the cell is increased) although at the expense of degrading the sum rate of the system.

Acknowledgements This work has been partially supported by the Spanish Government (Ministerio de Economía y Competitividad) under grant TEC2016-80090-C2-1-R, and by the University of Málaga.

References

1. Entrambasaguas, J. T. (2007). Multiuser capacity and fairness of channel/QoS-aware multiplexing algorithms. *IEEE Network*, *21*(3), 24–30.
2. Viswanath, P., Tse, D. N. C., & Laroia, R. (2002). Opportunistic beamforming using dumb antennas. *IEEE Transactions on Information Theory*, *48*(6), 1277–1294.
3. Shakkottai, S., & Stolyar, A. L. (2001). Scheduling algorithms for a mixture of real-time and non-real-time data in HDR. In *Proceedings of 17th international teletraffic congress* (vol. 4, pp. 793-804).
4. Valkova-Jarvis, Z. V., & Stoynov, V. R. (2017). Novel throughput quantile averaging methods for the proportional fair algorithm in indoor mobile networks. In *2017 advances in wireless and optical communications (RTUWO), Riga* (pp. 198-202).
5. Liu, E., & Leung, K. (2010). Expected throughput of the proportional fair scheduling over Rayleigh fading channels. *IEEE Communications Letters*, *14*(6), 515–517.
6. Pérez, J. (2009). Performance analysis of SNR-based scheduling policies in asymmetric broadcast ergodic fading channel. *EURASIP Journal on Wireless Communications and Networking*, *8*, 1–6.

7. Yang, L., & Alouini, M.-S. (2006). Performance analysis of multiuser selection diversity. *IEEE Transactions on Vehicular Technology*, *55*(6), 1848–1861.
8. Parruca, D., Grysla, M., Gortzen, S., & Gross, J. (2013). Analytical model of proportional fair scheduling in interference-limited OFDMA/LTE networks. In *2013 IEEE 78th vehicular technology conference (VTC Fall)* (pp. 1–7). Las Vegas, NV.
9. Kang, M., Sang, Y. J., Hwang, H. G., Lee, H. Y., & Kim, K. S. (2009). Performance analysis of proportional fair scheduling with partial feedback information for multiuser multicarrier systems. In *VTC Spring 2009 - IEEE 69th vehicular technology conference, Barcelona* (pp. 1–5).
10. Senst, A., Rittich, P. S., Ascheid, G., & Meyr, H. (2004). On the throughput of proportional fair scheduling with opportunistic beamforming for continuous fading states. In *Proceedings of IEEE vehicular technology conference (VTC)* (Vol. 1, pp. 300–304). Los Angeles, CA.
11. Sharma, N., & Ozarow, L. H. (2005). A study of opportunism for multiple-antenna systems. *IEEE Transactions on Informations Theory*, *51*(5), 1804–1814.
12. Yang, L., Kang, M., & Alouini, M.-S. (2007). On the capacity-fairness tradeoff in multiuser diversity systems. *IEEE Transactions on Vehicular Technology*, *56*(4), 1901–1907.
13. Tabassum, H., Yilmaz, F., Dawy, Z., & Alouini, M. S. (2012). On the modeling of uplink inter-cell interference based on proportional fair scheduling. In *2012 IEEE wireless communications and networking conference workshops (WCNCW), Paris* (pp. 317–321).
14. Parruca, D., & Gross, J. (2016). Throughput analysis of proportional fair scheduling for sparse and ultra-dense interference-limited OFDMA/LTE networks. *IEEE Transactions on Wireless Communications*, *15*, 6857–6870.
15. Günter, C. G. (1996). Comment on "Estimate of channel capacity in Rayleigh fading environment". *IEEE Transactions on Vehicular Technology*, *45*(2), 401–403.
16. IST-WINNER II Deliverable 1.1.2 v.1.2, "WINNER II Channel Models", IST-WINNER2, Tech. Rep., (2007) (available online: http://www.cept.org/files/1050/documents/winner2%20-%20final%20report.pdf).

Isabel M. Delgado-Luque received her B.Sc. degree in Telecommunications Engineering from the University of Malaga (Spain) in 2007. She has been working as a research at the Department of Communications Engineering at the University of Malaga since 2008. Her research interests include the field of mobile communications, especially, channel modeling and radio resource management strategies for LTE and LTE-Advanced.

Mari Carmen Aguayo-Torres received the Ph.D. degree in telecommunication engineering from the University of Málaga, Spain, in 2001, and the M.S. degree in 1994. Her M.S. thesis was on adaptive OFDM. She is currently an Associate Professor with the Department of Communications Engineering, University of Málaga. Her main research interests are 5G technologies with probabilistic QoS guarantees. Her work includes theoretical analysis with tools such as stochastic geometry besides statistical evaluation of current real or simulated networks. She is also involved in a number of public and private funded projects and actively collaborates with industry, mainly in the field of wireless communications.

Gerardo Gómez received the B.Sc. and Ph.D. degrees in telecommunication engineering from the University of Málaga, Spain, in 1999 and 2009, respectively. From 2000 to 2005, he was with Nokia Networks and Optimi Corporation (acquired by Ericsson), leading the area of quality of service (QoS) for 2G and 3G cellular networks. In 2005, he joined the University of Málaga, where he is currently an Associate Professor and the Deputy Director of the Communications Engineering Department. His research interests include the field of mobile communications, especially physical layer security, stochastic geometry, interference management, and QoS/QoE evaluation for multimedia services.

Francisco J. Martín-Vega received his B.Sc. degree in Telecommunication Engineering from the University of Malaga (Spain) in 2011, M.Sc. degree from the University of Vigo in 2014 and Ph.D. degree from the University of Málaga in 2017. He joined the Communication Engineering Department at the University of Malaga in 2011, as an associate researcher, where he participates in contracts with several industry partners related to cellular and satellite communications. His current research activity is focused on the mathematical modeling of heterogeneous networks and vehicular communications.

José Tomas Entrambasaguas received M.Sc. and Ph.D. degrees in telecommunication engineering from the Polytechnic University of Madrid (UPM), Spain, in 1975 and 1990, respectively. From 1975 to 1978 he was with E.T.S. Ingenieros de Telecomunicación, UPM, Madrid. From 1978 to 1993 he worked at Fujitsu-España R&D Center, mainly on the development of packet switching systems, data transmission, and hardware/software technology for computer local area networks. In 1993 he joined the University of Málaga, where he is now a professor and head of the Communication Engineering Department. His current research interests include digital signal processing, wireless access networks and methodologies for development and testing of complex communications systems.

Wireless Personal Communications
https://doi.org/10.1007/s11277-018-5898-8

Impact of Front-Haul Delays in Non-ideal Cloud Radio Access Networks

Alberto Carreras[1] · Isabel M. Delgado-Luque[1] · Francisco J. Martín-Vega[1] · Mari Carmen Aguayo-Torres[1] · Gerardo Gómez[1] · J. Tomás Entrambasaguas[1] · Eneko Atxutegi[2] · Ruben Solozabal[2] · Bego Blanco[2] · Jose Oscar Fajardo[2] · Fidel Liberal[2]

Abstract
The principle of Cloud Radio Access Network (C-RAN) is the split of traditional base stations into Radio Remote Units (RRU) as low-cost wireless access points, and Base Band Units (BBU) in a centralized location. This new RAN paradigm accepts several choices for a functional split of the protocol stack, with different latency requirements to the front-haul connecting a BBU with those RRUs under its control. In this paper, we focus on the functional split that implements the Medium Access Control layer at the BBU side. Particularly, we analyze the impact of delays on the report of the acknowledgment/negative acknowledgment messages for Hybrid Automatic Repeat reQuest (HARQ). In order to understand the trade-off between the HARQ report delay and user throughput, we define a new metric named as *Net Rate*. This metric is defined as the throughput that a user can reach after certain HARQ report delay while taking into account the actual channel conditions, the resource scheduling period to that user and the transmission window. The *Net Rate* metric can be used to determine the maximum HARQ report delay that can be tolerated by a user without throughput degradation. Our simulation results recommend the use for Mobile Edge Computing solutions to minimize the latencies of the front-haul connection and, thus, the impact of the HARQ delay on the throughput.

Keywords C-RAN · Functional split · Non-ideal front-haul

1 Introduction

Since the advent of packet data services with General Packet Radio Services (GPRS) in the 80's along with the appearance of the modern smartphones, mobile data subscriptions have not stopped growing and, moreover, the average data volume per user has significantly increased. According to the Ericsson mobility report [1], the demand of high-speed data traffic continues exponentially growing. Since 2011 the data traffic grew approximately 13% quarter-on-quarter and 55% year-on-year.

✉ Gerardo Gómez
ggomez@ic.uma.es

Extended author information available on the last page of the article

Published online: 26 June 2018

The users expect diverse and novel services with high data rates at the same cost. In addition, mobile networks are evolving towards ultra-dense deployments were different radio access networks such as macro-, micro-, pico- and femto-cells will coexist together. In the coexistence of different cellular access technologies, it is also necessary to take into account the arrival of a new 5G mobile generation and the increased use of Internet of Things (IoT) [2], machine-to-machine communications [3] and device mass use. Figure 1 illustrates an sketch of the envisioned future 5G network composed of different radio access technologies. The increase of traffic load and radio resource usage will lead to the increment of operating expenditure, energy consumption and investment in infrastructure.

The centralized Radio Access Network (RAN) design in 5G arises as a solution to the aforementioned problems moving some RAN-related functions back to shared hardware. The gathering of the processing resources in shared data centers boosts the reduction of deployment and managing costs and the CAPEX/OPEX efficiency is enhanced. If those resources run over virtualized infrastructures helped by Software Defined Network (SDN) [4] and Network Function Virtualization (NFV) technologies [5], the centralized RAN becomes a Cloud RAN (CRAN) [6].

In the context of CRAN, several centralization methods between evolved NodeBs (eNBs) can coexist. The 3GPP recommends different levels of functional division (see Fig. 2) in which one part of the logic is located at the Radio Remote Unit (RRU) while the

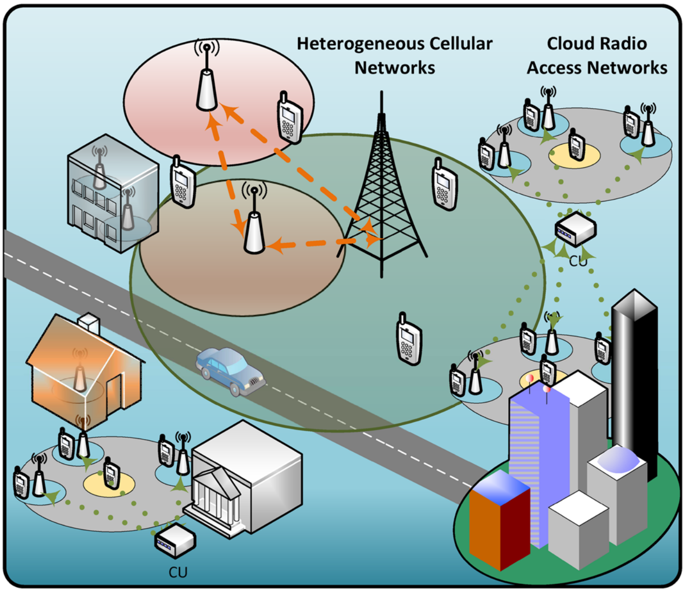


Fig. 1 Ultra-dense network example

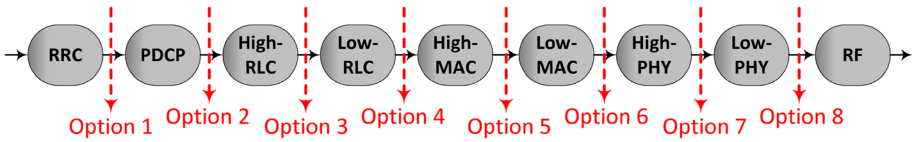


Fig. 2 3GPP functional splits proposal

rest is located at the centralized Base Band Unit (BBU) [7]. This centralization entails a new network architecture where all baseband processing is made by BBUs at centralized data centers and radio signals are exchanged with Remote Radio Units (RRUs) over low latency fiber-optic front-haul connections (Fig. 3). Depending on the functional split, the central cloud may take part in different layers of the softwarized protocols. For instance, the functional split given in option 6 [7] centralizes functionalities up to the Medium Access Control (MAC) layer at the BBU side [8], leaving only the physical layer in the RRUs. In terms of latency, the requirements of the front-haul connection are critical [9]. Since the deployment of optical fiber front-hauls involves costly operations, other solutions have been considered in order to facilitate the adaptation to the fully centralized RAN. This paper analyzes how the delay through non-ideal front-haul connection impacts on the Hybrid Automatic Repeat reQuest (HARQ) procedure and, consequently, on the user throughput.

Among the alternatives to resolve centralized RAN use-cases, one of the emerging technologies is the Mobile Edge Computing (MEC). It consists of evolving the traditional Small Cells (SCs) to CRAN Small Cells (CRAN-SCs) [10]. A CRAN-SC takes advantage of Physical Network Functions of the traditional SC and evolves its capabilities towards a virtualized execution platform that enables multi-operator ecosystems and allows deploying applications of vertical sectors with close-to-zero latency. In this regard, CRAN-SC clustering also enables the creation of a virtualized execution infrastructure in the form of a distributed data center. Centralized and distributed clouds will coexist during this evolution to a completely softwarized central C-RAN. In deployments with a multi-cloud distribution, different clouds are capable of working together due to the management of an orchestrator, becoming a hybrid cloud. In this case, centralized RAN can dynamically adjust the functional split to interact with heterogeneous eNBs.

Our work evaluates, in a multi User Equipment (UE) scenario and under different application requirements, the impact of the delay due to non-ideal connections while

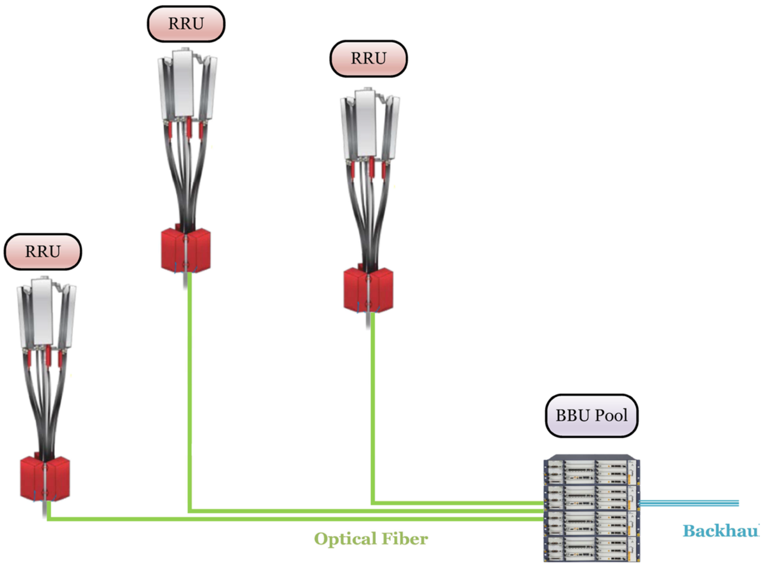


Fig. 3 Centralized RAN example

reporting the acknowledgment/negative acknowledgment (ACK/NACK) for HARQ messages. We study how such an increase in the latency affects the user throughput and define a new metric, named as *Net Rate*, able to evaluate the trade-off between the HARQ report delay and user throughput.

The remainder of this paper is organized as follows. Firstly, Sect. 2 briefly describes the MAC layer and Sect. 3 defines the *Net Rate* metric. In Sects. 4 and 5, the simulation setup and results are exposed. Finally, Sect. 6 summarizes the main conclusions of this work.

2 System Model

As previously introduced, the functional split option 6 (MAC-PHY split) as described in [7] is analyzed in this paper. In short, MAC layer is centralized while the lower layers of the stack are located at the RRU. As a consequence, a non-ideal front-haul would lead to reporting delays of the channel state as well as increments in the notification of HARQ acknowledgements (HARQ-ACKs).

Long Term Evolution (LTE) procedures for MAC and PHYsical (PHY) layers are described in [11, 12], respectively. In short, user data is encapsulated and coded in Transport Blocks (TB) whose sizes depend on the Modulation and Coding Scheme (MCS) and the number of Physical Resource Blocks (PRBs) assigned by the scheduler to that user transmission. Moreover, the Adaptive Modulation and Coding (AMC) along with the PRB scheduling make the MCS change over time. Users report their Channel Quality Indicator (CQI) as tracking the instantaneous Signal to Interference and Noise Ratio (SINR). The selected TB size depends on the MCS index finally assigned (i.e., the channel quality) and on the number of allocated PRBs (i.e., the resource sharing). The average TB size for a certain user is named as $\overline{\mathrm{TB}_{\mathrm{size}}}$.

In addition, the average time between two consecutive resource allocations to a given user is named as t_{TB}. The scheduler allocates resources to a user with a maximum periodicity upperbounded by the Transmission Time Interval (TTI), i.e., $t_{\mathrm{TB}} \geq 1$ ms. It must be noted that, from a radio resource assignment point of view, a user will only be taken into account by the scheduler if there is data to transmit. Thus, resource sharing does not only depend on available resources but also on the offered traffic pattern and the *greediness* of the data flow.

Specific mechanisms in the MAC layer are responsible for assuring a correct transmission of the TBs. HARQ functionality [11] uses a stop and wait protocol: when a TB is sent, the transmitting entity waits until an ACK or NACK is received. Then, if the TB is received without any error, the next TB is transmitted. However, the reception of an erroneous TB would lead to the retransmission of such TB until the packet is correctly received or the maximum number of retransmissions is reached. In order to optimize the transmission rate, several parallel HARQ processes are usually employed; each HARQ process sends one TB when resources are allocated to the user and waits for its associated acknowledgment. This procedure is repeated until the maximum number of HARQ processes, W, is reached after $W \cdot t_{\mathrm{TB}}$ seconds. In case of LTE FDD, there shall be a maximum of $W = 8$ parallel HARQ processes [12].

Moreover, HARQ uses Incremental Redundancy (IR) feature, whereby each retransmitted TB contains a different subset (or redundancy version) of coded bits [13]. Later, each subsequent transmission is combined with earlier transmissions to improve the decoding process through a Chase Combining (CS) procedure.

This paper aims to study the impact of HARQ-ACK reporting delay on the transmission rate. We name as t_{ACK} the time required to receive an HARQ-ACK report through the non-ideal front-haul, which is considered to be constant.

3 *Net Rate* Concept

To have a better understanding of the delay effects, we present the concept of *Net Rate* (NR), which is defined as the throughput that a given user can achieve taking into account the actual conditions of the transmission. As we will prove, the effect of exceeding a certain delay threshold may severely affect it.

Net Rate is determined using the following expression:

$$\mathrm{NR} = \frac{\overline{\mathrm{TB}_{\mathrm{size}}} W}{t_W}, \tag{1}$$

where W is the transmission window (i.e., the number of HARQ processes) and t_W is the time required to reuse an HARQ process, given by the maximum between the time needed to feedback the ACK and the time required for the execution of all W HARQ processes:

$$t_W = \max(t_{\mathrm{ACK}}, W \cdot t_{\mathrm{TB}}). \tag{2}$$

Note that if $t_{\mathrm{ACK}} > W \cdot t_{\mathrm{TB}}$, after $W \cdot t_{\mathrm{TB}}$, the HARQ process will be waiting for its ACK/NACK arrival. As a consequence, HARQ processes associated with the user will be blocked, thus reducing the *Net Rate*. On the other hand, if $t_{\mathrm{ACK}} < W \cdot t_{\mathrm{TB}}$, the HARQ process will be ready to transmit (or retransmit) a TB on the next TTI. The maximum throughput, $\overline{\mathrm{TB}_{\mathrm{size}}}/t_{\mathrm{TB}}$, would be thus reached.

Figure 4 shows this concept graphically for two different t_{ACK} examples. The first one–$t_{\mathrm{ACK-1}}$—represents an ACK that is received before all HARQ processes are handled, and hence, the HARQ entity is able to transmit without any further delay. However, the second case -$t_{\mathrm{ACK-2}}$—shows an ACK that is received after $W \cdot t_{\mathrm{TB}}$, thus causing the blockage of the HARQ process. Note that the minimum t_{TB} defined for an LTE system is 1 ms and, therefore, the minimum t_W is 8 ms if eight HARQ processes are being used. For this reason, a t_{ACK} lower than 8 ms will not affect the *Net Rate* for $W = 8$. Otherwise, the threshold from which the transmission rate decreases must be analyzed.

Once the *Net Rate* has been defined, the effect of higher report delay could be evaluated as an increment in t_{ACK}. However, the impact of such delay on the user throughput also depends on the Source Rate (SR), i.e. the data traffic rate transmitted to the user. As long as

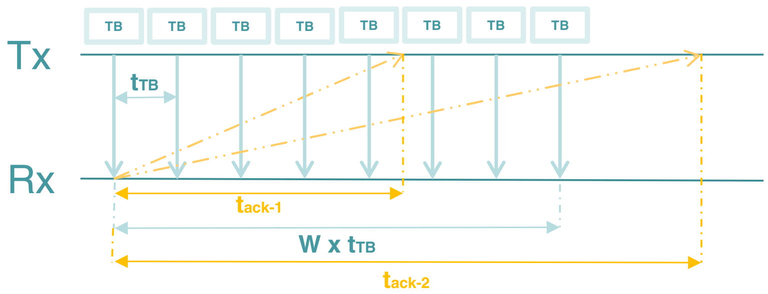


Fig. 4 HARQ transmission with different ACK reception times

the *Net Rate* is greater than the source rate transmitted to a given user, its throughput does not experience any degradation. In conclusion, HARQ-ACK delay will only have a negative impact on the user experience if two conditions are met: $t_{\mathrm{ACK}} > Wt_{\mathrm{TB}}$ and $\mathrm{NR} < \mathrm{SR}$. The delay threshold ($t_{\mathrm{W,th}}$) from which the throughput is degraded can be evaluated as:

$$t_{W,th} = \frac{\overline{\mathrm{TB}}_{\mathrm{size}} W}{SR} \tag{3}$$

As long as ACK delay t_W is lower than the threshold $t_{\mathrm{W,th}}$, the user throughput will not be affected. Otherwise, user throughput will decrease below SR.

4 Simulation Setup

In this section, the simulation setup is described. Simulations have been carried out on the Wireless Mobile SIMulator Advanced (WM-SIMA) [14], a link-level simulation framework for Long Term Evolution Advanced (LTE-A) mobile network. This tool has a wide range of configuration parameters for the physical and MAC layers of LTE-A without considering any link abstraction mechanism, leading to accurate and realistic results. In addition, the simulator is able to configure an heterogeneous network topology with different types of base stations.

We evaluate the DownLink (DL) of an eNB with 10 MHz bandwidth (equivalent to 50 PRBs) and six users within the coverage range of the cell, numbered as 1–6 in growing distances to the RRU. The average SINR values for those users are 27.8, 22.5, 18, 13.3, 9.7, and 4 dB. The average reported CQI indexes are 15, 13, 11, 9, 7 and 5, respectively.

We have considered the ITU Urban Macro channel model (UMa) [15] with a user speed of 4 km/h. Additionally, to account for state-of-the-art adaptive modulation and coding schemes, we have considered an outer loop link adaptation algorithm [16] to improve the selection of MCSs at each TTI. Hence, the instantaneous MCS index of each user, as well as their transport block size, varies along simulation according to the instantaneous SINR. Transport block sizes are averaged for each user independently to obtain its $\overline{\mathrm{TB}}_{\mathrm{size}}$, which is bigger for better channel quality users (lower numbering).

Simulations are carried out along 20,000 subframes. To ease results interpretation, a Round Robin (RR) [17] scheduler is configured with an allocation size per user and subframe of 15 PRBs (i.e., 3 users can be scheduled per subframe, each user scheduled after two subframes, $t_{\mathrm{TB}} = 2\,\mathrm{ms}$, if all buffers contain data).

A set of HARQ-ACK report delays has been simulated: 1, 4, 8, 16 and 24 ms. Furthermore, the influence of source rate has been studied. First, we assume a streaming source rate of 6 Mbps, characterized by packets of 1400 bytes every 1.9 ms. Lower source rates of 300, 950 kbps and 2.6 Mbps (sending data bursts each 2 s) have been also simulated. Main simulation parameters are summarized in Table 1.

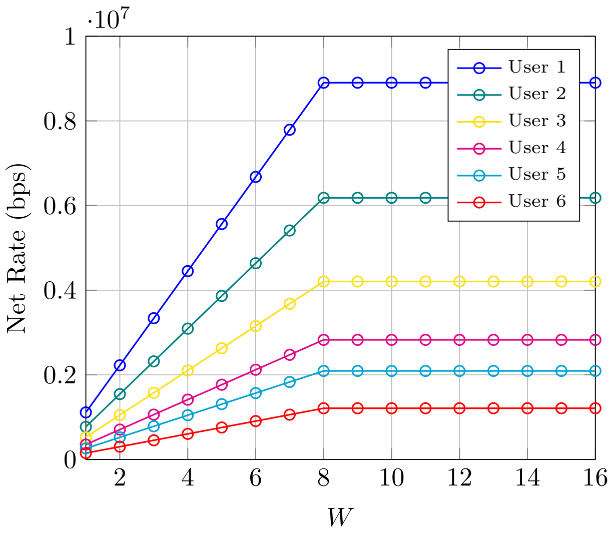
5 Simulation Results

First in this section, *Net Rate* is evaluated for each user from Eq. (1). The effect of the transmission window, i.e., the number of HARQ processes, is illustrated in Figs. 5 and 6 for $t_{\mathrm{TB}} = 1$ ms and $t_{\mathrm{TB}} = 2$ ms, respectively. In both figures, two operating regimes can be observed: (1) a regime where the *Net Rate* increases with *W*; and (2) a regime where the *Net Rate* is independent of *W*. The first regime comes when the ACK delay, t_{ACK}, is higher

Table 1 Simulation setup

Parameter	Value
Carrier frequency	2 GHz
Bandwidth	10 MHz
Transmission power	43 dBm
Simulation time	2×10^4 ms
Transmission mode	SISO
Channel model	ITU UMa
Mobile speed	4 km/h
Number of users	6
Set of avg. SINRs per user (dB)	27.8, 22.5, 18, 13.3, 9.7, 4
Set of avg. CQIs per user	15, 13, 11, 9, 7, 5
Scheduling policy	Round robin
Max. allocation per user	15 PRBs
Source rate	6 Mbps, 2.6 Mbps, 950 kbps, 300 kbps
HARQ processes	8
No. of retransmissions	3
Decoding algorithm	SOVA
Link adaptation method	OLLA

Fig. 5 *Net Rate* versus W for different channel qualities with $t_{\mathrm{TB}} = 1$ ms and $t_{\mathrm{ACK}} = 8$ ms



than the time between two consecutive transmission intervals, i.e., $t_{\mathrm{ACK}} > W \cdot t_{\mathrm{TB}}$. In this case, if we assume that there are no errors in the data transmission, each HARQ process is blocked waiting for its ACK message during $t_{\mathrm{ACK}} - W \cdot t_{\mathrm{TB}}$ seconds, thus increasing the window size (i.e., the number of parallel HARQ processes) improves the *Net Rate*. The second operating regime appears when $t_{\mathrm{ACK}} < W \cdot t_{\mathrm{TB}}$. In this case, the ACK delay is small enough to avoid blockage on HARQ processes. Therefore, the *Net Rate* is maximal for

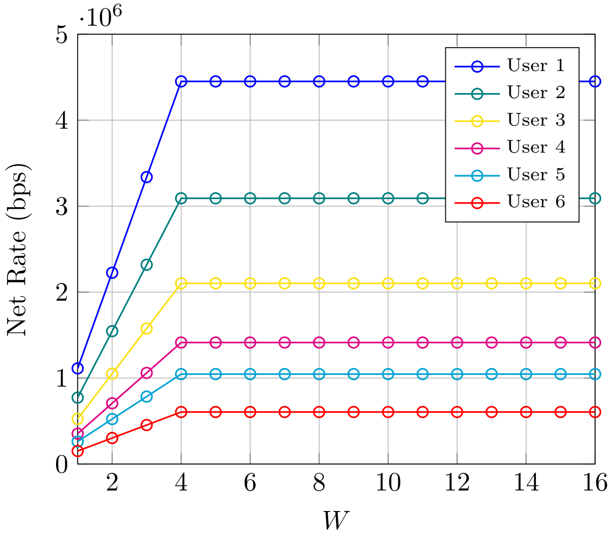


Fig. 6 *Net Rate* versus W for different channel qualities with $t_{\mathrm{TB}} = 2$ ms and $t_{\mathrm{ACK}} = 8$ ms

each user, and its value is the quotient between the average TB size (which depends on the channel quality of the user) and the time between two consecutive resource allocations, t_{TB}.

Interestingly, we observe that the value of W splitting these two regions also depends on t_{TB}. Let us recall that a higher t_{TB} value means that the transmission towards a given user occurs less frequently, which implies that the regime without blocked HARQ processes is reached with a smaller transmission window, W. This fact is also illustrated in Figs. 5 and 6. It is observed that the region without blocked processes is reached with $W = 8$ for $t_{\mathrm{TB}} = 1$ ms whereas it is reached with only $W = 4$ for $t_{\mathrm{TB}} = 2$ ms.

The theoretical delay thresholds $t_{W,th}$ as evaluated by Eq. (3) are given in Table 2 for different source rates and $W = 8$. This table gathers the required $t_{W,th}$ in order to achieve a *Net Rate* equal to the source rate so that $t_W > t_{W,th}$ causes a degradation of the user throughput. Under the previously described simulation setup, users with better channel quality are assigned bigger transport blocks, thus they can resist higher delays. As expected from Eq. (3), less strict source rate requirements also permit higher delays.

Figures 7 and 8 show the *Net Rate* behavior as a function of the ACK delay for two different t_{TB} scenarios. In the first case, we use $t_{\mathrm{TB}} = 1\,\mathrm{ms}$, and consequently, $W \cdot t_{\mathrm{TB}} = 8\,\mathrm{ms}$, so t_{ACK} does not affect users' *Net Rate* if $t_{\mathrm{ACK}} \leq 8\,\mathrm{ms}$. From 8 ms delay on, the *Net Rate* decreases along with t_{ACK}, being more noticeable for higher transmission

Table 2 Delay thresholds for the simulated users with different source rates according to the Eq. (3) with $W = 8$ and $t_{\mathrm{TB}} = 2\,\mathrm{ms}$

User	6 Mbps (ms)	300 kbps (ms)	2.6 Mbps (ms)	950 kbps (ms)
1 (best)	11.6	246.1	26.8	75.5
2	8.1	175.6	18.7	50.7
3	5.6	110.9	12.8	36.1
4	3.8	75.2	8.8	23.5
5	2.7	56.5	6.4	17.8
6 (worst)	1.6	33.0	3.7	9.9

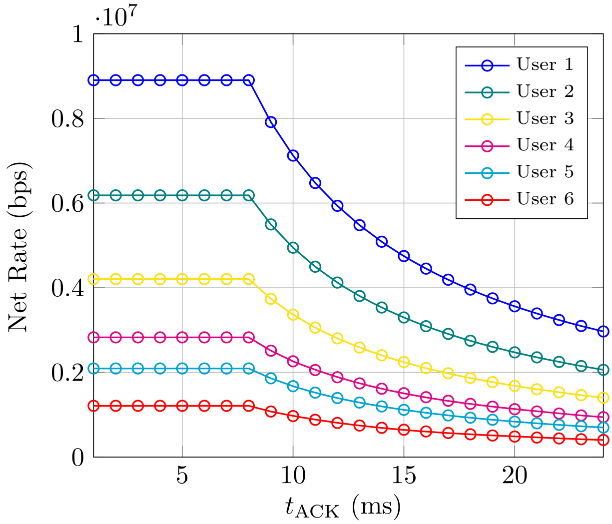
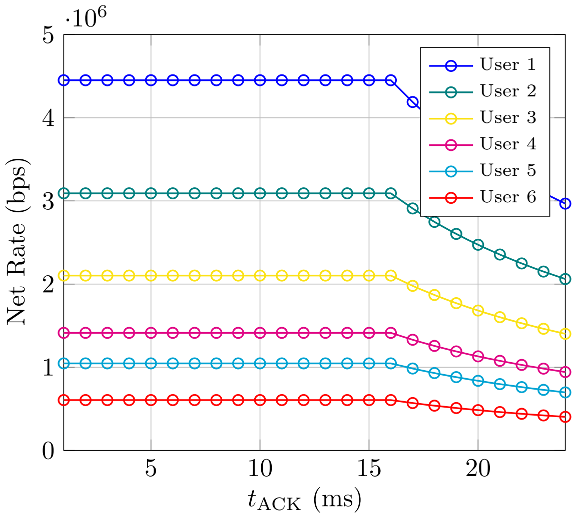


Fig. 7 *Net Rate* versus t_{ACK} for different channel qualities with $t_{TB} = 1$ ms and $W = 8$

Fig. 8 *Net Rate* versus t_{ACK} for different channel qualities with $t_{TB} = 2$ ms and $W = 8$



values. Likewise, calculations for $t_{\text{TB}} = 2\,\text{ms}$ ($W \cdot t_{\text{TB}} = 16\,\text{ms}$) are illustrated in Fig. 8. On the one hand, the initial *Net Rate* is half of that in Fig. 7 due to the fact that the time between two resources allocation has been doubled ($t_{\text{TB}} = 2\,\text{ms}$). On the other hand, the report delay required to have a negative impact on the *Net Rate* is also doubled.

Once the theoretical results have been measured, figures below analyze the simulation results for those sources previously described. Figures show the achieved throughput per user as the delay in the front-haul (t_{ACK}) grows due to a non-ideal connection. Results are different from previous ones as in this case the scheduler only assigns resources to those users having data in their queues. Users able to transmit with bigger transport blocks empty their queues earlier. The free resources are later employed by those users with worse channel quality, reducing time between resource allocations t_{TB}.

Figure 9 shows the results for the highest source rate, SR = 6 Mbps. From Eq. (2) it is evident that $t_W \approx \max(t_{\mathrm{ACK}}, 16\,\mathrm{ms})$. Delay thresholds, as given by Eq. (3) (shown in Table 2), are lower than 16 ms for all users. Therefore, no user will be able to reach their source rate (i.e., NR < SR for all users) and, therefore, the throughput will experience a significant degradation. Users 1 and 2 could improve their throughput if the scheduler would be capable of allocating resources to them at a subframe rate ($t_{\mathrm{TB}} = 1\,\mathrm{ms}$). However, users 3–6 are not able to reach the source rate due to the fact that $t_{W,th}$ is below the minimum of 8 ms imposed by the system. For them, even if t_{ACK} is not influencing the throughput, obtained rate is lower than the source rate as channel is not able to carry the whole 6 Mbps information flow due to: (1) the maximum achievable rate; (2) the fact that the transmission resources are shared between the six users; and (3) the effect of retransmissions. Hence, the same results (in terms of throughput) would be obtained if a higher source rate is used since the system is working on saturated traffic conditions.

Contrary to the previous case, a source rate of 300 kbps is slow enough to ensure that all users achieve the upperbounded source rate as NR > SR. Figure 10 depicts the above mentioned results, where no impact of delay on the final throughput is shown regardless the selected user. As a confirmation step, the calculated values for the delay thresholds (in Table 2) show that t_{ACK} delays would need to be too high to perceive its effects over the user throughput (33 ms for the weakest UE in terms of radio quality whereas 246.1 ms are needed for UEs receiving good quality signal).

Results for 2.6 Mbps source rate are shown in Fig. 11. This source rate is high enough to have all user buffers with data pending to be transmitted. In this regard, since the maximum number of users scheduled at the same TTI is 3, each user is allocated resources every 2 ms, leading to increased queuing periods for each packet. Users 1 and 2 reach the source rate of 2.6 Mbps. However, the throughput of users 3 to 6 is lower than the source rate due to their $NR < SR$. This is a result that can be contrasted with Table 2, where $t_{W,th}$ is lower than 16 ms for such users (12.8, 8.8, 6.4 and 3.7 ms, respectively). Users 4 and 5 could improve their throughput if the scheduler would be capable of allocating resources to them at a subframe rate (e.g. $t_{\mathrm{TB}} = 1$ ms); however, the $t_{W,th}$ of users 4 and 5 is below 8 ms (6.8

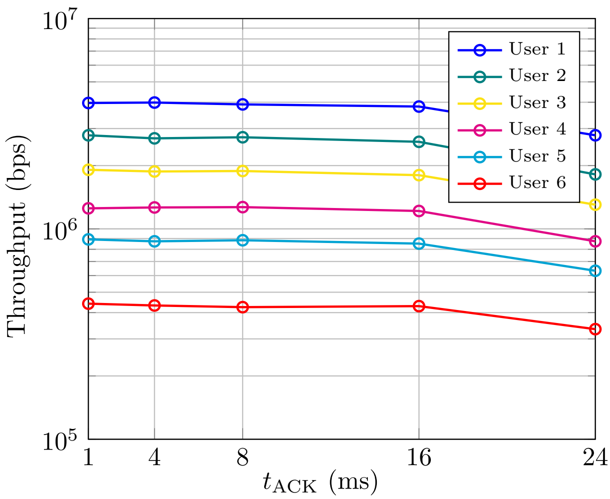


Fig. 9 Throughput per user, SR = 6 Mbps

Fig. 10 Throughput per user, SR = 300 kbps

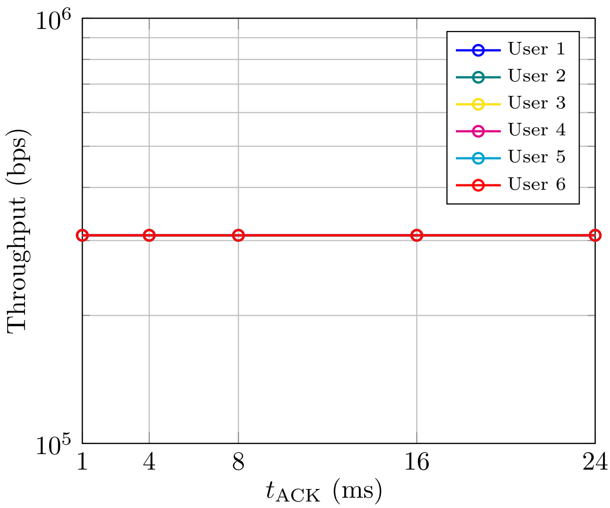
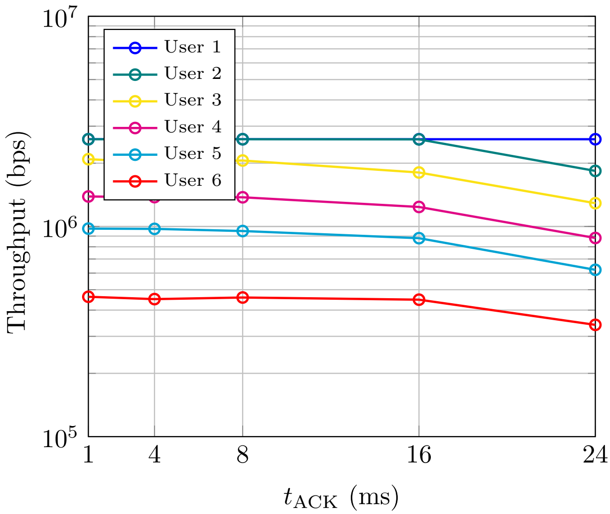


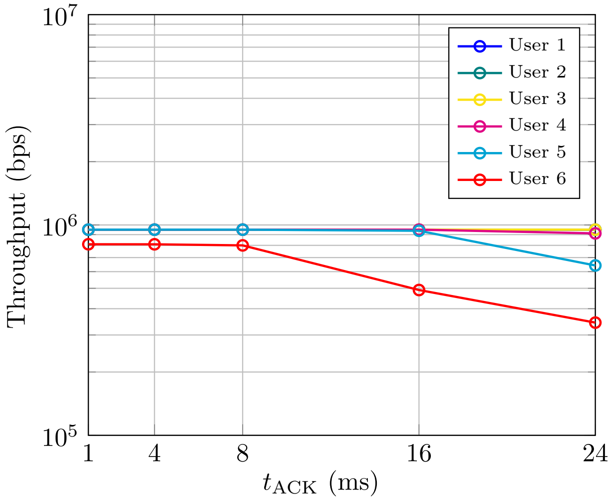
Fig. 11 Throughput per user, SR = 2.6 Mbps



and 3.7 ms), so they will never achieve 2.6 Mbps throughput. User 1 is not affected by delays below 26.8 ms (Table 2), which is out of the range of the figure. Users 2 to 6 experience a degradation of their throughput when $t_{\mathrm{ACK}} > 16$ ms, as given in Table 2.

Figure 12 shows the results for an intermediate source rate of 950 kbps per user. Results show outcome patterns that allow classifying the users into three groups according to their behavior. The best three users (labeled as 1–3) experience a channel quality good enough to keep their *Net Rate* over the source rate for all evaluated delays, i.e. NR > SR. In such cases, the product of the window size by the average time between consecutive transmissions is $W \cdot t_{\mathrm{TB}} \approx 16$ ms. Thus, t_{ACK} does not have an impact on their performance until it reaches 16 ms. At this value, any increment of t_{ACK} will reduce their *Net Rate*. As shown in Table 2, 1–3 users' throughput is not degraded until their delays reach 75.5, 50.7 and 36.1 ms, respectively. In contrast, users 4 and 5 have a *Net Rate* higher than the source

Fig. 12 Throughput per user, SR = 950 kbps



rate NR > SR for non-delayed front-haul. However, the throughput of these users decreases as t_{ACK} grows over 23.5 and 17.8 ms, respectively. This is because $t_{\mathrm{ACK}} > W \cdot t_{\mathrm{TB}}$ and t_W value is high enough to make NR < SR. For the worst user, the average transport block size is so low that any upper layer packet is segmented in a set of TBs. Therefore, when other users have finished their transmission, this user is still sending TBs. That is, its t_{TB} is approximately 1 ms because most of the time it is transmitting alone in the cell, so $W \cdot t_{\mathrm{TB}} \approx 8$ ms. Consequently, its NR lowers if $t_{\mathrm{ACK}} > 8$ ms. According to Table 2, its throughput decreases as $t_{\mathrm{ACK}} > 9.9$ ms as under this condition $NR < SR$.

6 Conclusions

This paper has analyzed the impact of HARQ report delay on the user throughput under centralized and MAC splitting conditions. In order to clarify the understanding of the trade-off between report delay and user throughput, a new metric named as *Net Rate* (NR) has been proposed. This metric represents the achievable throughput by a given user under the actual transmission conditions, i.e., the quality of the available resources, how often the user receives resources from the scheduler, the transmission window of the system, and the HARQ report delay.

We analyze the *Net Rate* behavior for two values of the scheduling period t_{TB} with different channel qualities. *Net Rate* decreases when the HARQ report delay is longer than the time required for the execution of HARQ processes, $W \cdot t_{\mathrm{TB}}$. In addition, delay thresholds that fulfill the condition NR = SR have been calculated for source rates from 300 kbps to 6 Mbps.

Later, we obtain from simulations the conditions under which the user throughput is degraded due to a too long reporting delay. Simulation results show that user throughput gets worse if *Net Rate* is lower than the source rate (NR < SR). It is also highlighted that the actual *Net Rate* is influenced by the resource sharing. Since good quality users finish their transmissions earlier, free resources can be used by the scheduler to allocate resources to worse quality users, thus reducing their time between consecutive resource allocations and improving their allowable delay.

Acknowledgements This work has been partially supported by the Spanish Government (Ministerio de Economía y Competitividad) under Grant TEC2016-80090-C2 and by the Universidad de Málaga.

References

1. Ericsson (2017). Ericsson mobility report. https://www.ericsson.com/assets/local/mobility-report/documents/2017/ericsson-mobility-report-november-2017.pdf. Accessed 28 May 2018.
2. Zanella, A., Bui, N., Castellani, A., Vangelista, L., & Zorzi, M. (2014). Internet of things for smart cities. *IEEE Internet of Things Journal*, *1*(1), 22–32.
3. Lien, S. Y., Chen, K. C., & Lin, Y. (2011). Toward ubiquitous massive accesses in 3GPP machine-to-machine communications. *IEEE Communications Magazine*, *49*(4), 66–74.
4. Sezer, S., Scott-Hayward, S., Chouhan, P. K., Fraser, B., Lake, D., Finnegan, J., et al. (2013). Are we ready for SDN? Implementation challenges for software-defined networks. *IEEE Communications Magazine*, *51*(7), 36–43.
5. Mijumbi, R., Serrat, J., Gorricho, J. L., Bouten, N., Turck, F. D., & Boutaba, R. (2016). Network function virtualization: State-of-the-art and research challenges. *IEEE Communications Surveys Tutorials*, *18*(1), 236–262.
6. Wu, J., Zhang, Z., Hong, Y., & Wen, Y. (2015). Cloud radio access network (C-RAN): a primer. *IEEE Network*, *29*(1), 35–41.
7. 3GPP TR 38.801 v2.0.0 (2018). Study on new radio access technology: Radio access architecture and interfaces (Release 14). http://www.3gpp.org/ftp/Specs/archive/38_series/38.801/38801-200.zip. Accessed 28 May 2018.
8. NGMN Forum (2015). Further study on critical C-RAN technologies. Technical Deliverable. https://www.ngmn.org/fileadmin/user_upload/NGMN_RANEV_D2_Further_Study_on_Critical_C-RAN_Technologes_v1.0.pdf. Accessed 28 May 2018.
9. Rost, P., Bernardos, C., Domenico, A., Girolamo, M., Lalam, M., Maeder, A., et al. (2014). Cloud technologies for flexible 5G radio access networks. *IEEE Communications Magazine*, *52*(5), 68–76.
10. Fajardo, J. O., Liberal, F., Giannoulakis, I., et al. (2016). Introducing mobile edge computing capabilities through distributed 5G cloud enabled small cells. *Mobile Networks and Applications*, *21*(4), 564–574.
11. 3GPP TS 36.321 v15.1.0. (2018). Medium access control (MAC) protocol specification (Release 15). http://www.3gpp.org/ftp/specs/archive/36_series/36.321/36321-f10.zip. Accessed May 28, 2018.
12. 3GPP TS 36.213 v15.1.0. (2018). Physical layer procedures (Release 15). http://www.3gpp.org/ftp/specs/archive/36_series/36.213/36213-f10.zip. Accessed May 28, 2018.
13. Sesia, S., Toufik, I., & Baker, M. (2009). *LTE—The UMTS long term evolution: From theory to practice*. New York: Wiley.
14. Carreras, A., Delgado-Luque, I. M., Martin-Vega, F. J., Gomez, G., Aguayo-Torres, M. C., & Entrambasaguas, J. T. (2018). A system-level simulator for the downlink of LTE-A: Case of study—Cell-offloading in HetNets. *Wireless Personal Communications*, *100*(1), 177–191.
15. ITU-R M.2135-1. (2009). Guidelines for evaluation of radio interface technologies for IMT-advanced. https://www.itu.int/dms_pub/itu-r/opb/rep/R-REP-M.2135-1-2009-PDF-E.pdf. Accessed May 28, 2018.
16. Blanquez-Casado, F., Gomez, G., Aguayo-Torres, M. C., & Entrambasaguas, J. T. (2016). eOLLA: An enhanced outer loop link adaptation for cellular networks. *EURASIP Journal on Wireless Communications and Networking*, *2016*, 20.
17. Liu, X., Chong, E. K. P., & Shroff, N. B. (2002). A framework for opportunistic scheduling in wireless networks. *Computer Networks*, *41*(4), 451–474.

Alberto Carreras was born in Las Palmas de Gran Canaria, Spain. He received the B.E. degree in telecommunications engineering from the University of Málaga, Spain, in 2015, and the M.Tech. degree in mobile communication engineering in 2017. In 2015, he joined the Department of Communication Engineering, University of Málaga, where he is currently associated researcher. His main areas of research interest are channel estimation techniques, load balancing and interference coordination in mobile Heterogeneous Networks and C-RAN.

Isabel M. Delgado-Luque received her B.Sc. degree in Telecommunications Engineering from the University of Malaga (Spain) in 2007. She has been working as a research at the Department of Communications Engineering at the University of Malaga since 2008. Her research interests include the field of mobile communications, especially, channel modeling and radio resource management strategies for LTE and LTE-Advanced.

Francisco J. Martín-Vega received his B.Sc. degree in Telecommunication Engineering from the University of Malaga (Spain) in 2011, M.Sc. degree from the University of Vigo in 2014 and Ph.D. degree from the University of Málaga in 2017. He joined the Communication Engineering Department at the University of Malaga in 2011, as an associate researcher, where he participates in contracts with several industry partners related to cellular and satellite communications. His current research activity is focused on the mathematical modeling of heterogeneous networks and vehicular communications.

Mari Carmen Aguayo-Torres received the Ph.D. degree in telecommunication engineering from the University of Málaga, Spain, in 2001, and the M.S. degree in 1994. Her M.S. thesis was on adaptive OFDM. She is currently an Associate Professor with the Department of Communications Engineering, University of Málaga. Her main research interests are 5G technologies with probabilistic QoS guarantees. Her work includes theoretical analysis with tools such as stochastic geometry besides statistical evaluation of current real or simulated networks. She is also involved in a number of public and private funded projects and actively collaborates with industry, mainly in the field of wireless communications.

Gerardo Gómez received the B.Sc. and Ph.D. degrees in telecommunication engineering from the University of Málaga, Spain, in 1999 and 2009, respectively. From 2000 to 2005, he was with Nokia Networks and Optimi Corporation (acquired by Ericsson), leading the area of quality of service (QoS) for 2G and 3G cellular networks. In 2005, he joined the University of Málaga, where he is currently an Associate Professor and the Deputy Director of the Communications Engineering Department. His research interests include the field of mobile communications, especially physical layer security, stochastic geometry, interference management, and QoS/QoE evaluation for multimedia services.

J. Tomás Entrambasaguas received M.Sc. and Ph.D. degrees in telecommunication engineering from the Polytechnic University of Madrid (UPM), Spain, in 1975 and 1990, respectively. From 1975 to 1978 he was with E.T.S. Ingenieros de Telecomunicación, UPM, Madrid. From 1978 to 1993 he worked at Fujitsu-España R&D Center, mainly on the development of packet switching systems, data transmission, and hardware/software technology for computer local area networks. In 1993 he joined the University of Málaga, where he is now a professor and head of the Communication Engineering Department. His current research interests include digital signal processing, wireless access networks and methodologies for development and testing of complex communications systems.

Eneko Atxutegi received B.Sc. and M.Sc. degrees in telecommunications engineering from the University of the Basque Country in 2011 and 2014, respectively. In 2018 he obtained the Ph.D. in Telecommunications Engineering from the same university, mainly focused on analysing the open-issues regarding the interaction of mobile networks and transport protocols and proposing cost-effective solutions to solve them. His research interests also include MEC and C-RAN topologies as well as NFV.

Ruben Solozabal received his M.Sc. in Telecommunications engineering in 2015 from UPV/EHU University. He has worked at the R&D division of Ingeteam S.A. for 3 years focusing on Industial Communications. Currently, he is a Ph.D. candidate with particular interest in Public Safety over 5G architecture.

Dr. Bego Blanco received her BS and MS in Telecommunications Engineering from the University of the Basque Country, Spain, in 2000, and her Ph.D. in Telecommunications Engineering in 2014 from the same university. She currently works as a lecturer and researcher in the Faculty of Engineering in Bilbao. Her research interests include PQoS/QoE/QoS assessment as well as multicriteria optimization in 5G networks.

Dr. Jose Oscar Fajardo works as research fellow in the Department of Communications Engineering of the University of the Basque Country (UPV/EHU), at the Faculty of Engineering in Bilbao. He received his Ph.D. degree in 2016. He works in adaptive management of mobile multimedia services in 4G and 5G networks and under the framework of IMS. He has co-authored more than 25 journal and conference papers since 2005, mainly in areas of QoS/PQoS/QoE and service performance assessment, and QoS-aware networking.

Dr. Fidel Liberal received his Ph.D. from University of the Basque Country (UPV/EHU) in 2005 for his work in the area of holistic management of quality (both QoS and QoE) in telecommunications services. His research interests include quality management and multicriteria optimization for multimedia and mission critical services in NGNs and 5G, particularly the evolution of service provision over C-RAN and MEC architectures together with SDN and NFV. He has co-authored more than 75 conference and journal papers.

Affiliations

Alberto Carreras[1] · Isabel M. Delgado-Luque[1] · Francisco J. Martín-Vega[1] · Mari Carmen Aguayo-Torres[1] · Gerardo Gómez[1] · J. Tomás Entrambasaguas[1] · Eneko Atxutegi[2] · Ruben Solozabal[2] · Bego Blanco[2] · Jose Oscar Fajardo[2] · Fidel Liberal[2]

Alberto Carreras
acm@ic.uma.es

Isabel M. Delgado-Luque
imdl@ic.uma.es

Francisco J. Martín-Vega
fjmvega@ic.uma.es

Mari Carmen Aguayo-Torres
aguayo@ic.uma.es

J. Tomás Entrambasaguas
jtem@ic.uma.es

Eneko Atxutegi
eneko.atxutegi@ehu.eus

Ruben Solozabal
ruben.solozabal@ehu.eus

Bego Blanco
begona.blanco@ehu.eus

Jose Oscar Fajardo
joseoscar.fajardo@ehu.eus

Fidel Liberal
fidel.liberal@ehu.eus

[1] Dpto. Ingeniería de Comunicaciones, Universidad de Málaga, Malaga, Spain

[2] Dpto. Electrónica y Telecomunicaciones, Universidad del País Vasco, País Vasco, Spain

Wireless Pers Commun (2018) 100:177–191
https://doi.org/10.1007/s11277-018-5616-6

A System-Level Simulator for the Downlink of LTE-A: Case of Study—Cell-Offloading in HetNets

Alberto Carreras[1] · Isabel M. Delgado-Luque[1] · Francisco J. Martín-Vega[1] · Gerardo Gómez[1] · Mari Carmen Aguayo-Torres[1] · J. Tomás Entrambasaguas[1]

Published online: 20 March 2018

Abstract In this paper, we present a novel and efficient data-flow oriented simulation platform for the downlink of Long Term Evolution Advanced (LTE-A). This tool accounts for a wide set of configuration parameters from LTE-A physical and medium access control layers as well as different traffic sources. The simulator, which is implemented in C++, does not consider any link abstraction mechanism, leading to accurate and realistic results. As a case of study, we investigate the performance of cell-offloading for different deployments of small cell base stations that differ in the distance towards their nearest macro base station under File Transfer Protocol type traffic. Additionally, the impact of the spatial user distribution on the performance of cell-offloading is assessed. Results reveal that aggregated throughput is maximized when macro base stations are highly offloaded, which means that a high bias for cell association is advisable in terms of aggregated throughput. However, if the fairness among users is also taken into account, the selected bias for cell-offloading should be smaller.

Keywords Simulator · LTE-A · Heterogeneous cellular systems · Offloading · Cell range expansion

1 Introduction

Current and next generation cellular systems require efficient and accurate simulators to assess the performance in realistic settings and identify trade-offs between Key Performance Indicators (KPIs). These simulations can be a preliminary step before real

This work has been supported by the Spanish Government (Ministerio de Economía y Competitividad) and FEDER under grant TEC2016-80090-C2-1-R and by the Universidad de Málaga.

✉ Mari Carmen Aguayo-Torres
aguayo@ic.uma.es

[1] Dpto. de Ingeniería de Comunicaciones, Universidad de Málaga, Málaga, Spain

implementations or deployments that require a measurement campaign to obtain insights in a costly and time-consuming manner.

Firstly, simulating tools should be flexible enough to account for many practical configurations of the involved layers, e.g., physical (PHY) and Medium Access Control (MAC) layers, and different deployments of Macro-cell Base Stations (MBSs) and Small-cell Base Stations (SBSs). This flexibility avoids the need to change the simulator in order to simulate a different scenario.

Secondly, a good simulator should provide realistic results. To obtain realistic results it would be appealing to simulate every layer and functionality instead of applying abstraction of the layers like Effective SNR mapping (ESM) techniques [1]. Additionally, practical impairments like the occurrence of errors or delay in the feedback information provided by the users or a spatial correlation between antenna elements in Multiple Input Multiple Output (MIMO) systems should be considered.

Thirdly, simulating tools should be efficient and require a short time to perform the simulations in order to obtain a quick evaluation of the main metrics. However, it is difficult to have efficient simulating tools without using an abstraction model to reduce the computational complexity. Moreover, it is appealing if the simulators can be downloaded by the research and teaching communities.

In the literature, there are different simulation tools for LTE-A systems that aim at fulfilling these objectives. Specifically, in [2] an LTE simulator implemented in Matlab is presented. This tool actually consists of two independent simulators: (i) a link-level simulator, which allows obtaining tables with PHY layer results; and (ii) a system-level simulator, which uses these results previously stored in tables to obtain the desired KPIs. With this approach, it is possible to make an abstraction of the PHY layer in simulations involving upper layers. This approach reduces the time required to perform simulations; however, the cost to pay is a reduction on the accuracy. A flexible tool named SimuLTE is presented in [3]. This tool is based on OMNeT ++, which allows to add different modules easily. An example is INET module, which allows simulating real-time applications using various radio access technologies. A simulator developed in C ++, which contemplates a great number of radio resource management and MAC functionalities such as mobility, handover and different scheduling algorithms, is presented in [4]. However, the PHY modelling is not so flexible since it is not possible to consider MIMO schemes, to select different decoding algorithms, or channel estimation methods. In [5], a simulator based on ns-3 has been designed to allow end-to-end simulation by assuming realistic traffic patterns.

However, a common limitation of these simulators is that they resort on the abstraction of the PHY layer in order to reduce the computational complexity. PHY layer abstraction is normally implemented using ESM techniques that require some stored tables with the Block Error Rate (BLER) versus SNR for the AWGN channel. Additionally, these techniques need some calibration coefficients (i.e., correction parameters) which have to be obtained by simulation for a given PHY configuration, i.e., channel profile, modulation, coding rate, decoding algorithm, MIMO processing, equalization, etc. Unfortunately, this solution has two main drawbacks: (i) it leads to a smaller accuracy; and (ii) it reduces the flexibility, since the stored calibration coefficients are highly sensitive to the PHY configuration that was used to obtain them.

In this paper we present Wireless Mobile SIMulation Advanced (WM-SIMA) which aims at fulfilling the aforementioned objectives without using PHY abstraction mechanisms: (1) the simulator is widely flexible since it considers many PHY and MAC configurations that are defined in the 3GPP Long Term Evolution Advanced (LTE-A) as well as many types of deployment and frequency planning strategies; (2) it is accurate since it does not use any link abstraction or semi-analytic results, and it allows to consider many impairments; (3) it is very efficient thanks to its data-flow oriented architecture that has been developed in C++; and (4) it is available to be downloaded in [6].

To show its potential benefits, it is considered in this paper the case of Heterogenous Networks (HetNets) with cell-offloading and File Transfer Protocol (FTP) traffic. In HetNets it is normally used a cell association criteria based on maximum averaged received power from DownLink (DL) pilot signals. Nevertheless, since MBSs transmit with a higher power and have better propagation conditions than SBSs, this criteria would lead to an scenario with highly congested MBSs and almost empty SBSs [7]. For that reason, it is normally used a bias in the association to offload MBSs, so that a better use of available resources is done at the SBSs. This bias, which is also known as Cell Range Expansion (CRE) bias, increases the interference of the system since cell ranged expanded users tend to suffer from a high interference from their nearest MBS. In [8] it is shown that this tradeoff between making a better use of the bandwidth and interference leads to an optimal value of the bias in terms of throughput for full-buffer traffic. Besides this, the available literature have not investigated in deep the effect of cell-offloading with FTP traffic in terms of user fairness [7–10] which is a paramount metric in cellular systems. Hence, in this work we investigate with WM-SIMA the performance of a LTE-A system with cell-offloading and FTP traffic. To provide a better understanding about the tradeoff between system throughput and fairness in such a scenario, different deployments for the SBSs are considered.

The rest of the paper is structured as follows. Section 2 describes the main features of WM-SIMA and its architecture. Section 3 illustrates the performance of cell-offloading for different setting and deployments. Finally, Sect. 4 summarizes the main findings of the presented work.

2 Wireless Mobile SIMulator Advanced (WM-SIMA)

The simulator architecture is organized into two levels: *System level*, implemented in Matlab; and *Link level*, implemented in C++.

2.1 WM-SIMA: System Level

This level, which is implemented in MATLAB, generates the information associated with the topology of the cellular network to be simulated. Large-scale parameters as the number of user terminals (UEs), number and type of Base Stations (BSs), BSs transmission power, cell area, frequency planning and path loss models among others are selected and used at this level. A common feature of these parameters is that they can be considered as time-invariant during a small scale of time (e.g., 20 s).

Two types of BSs can be configured, i.e., Macro (MBSs) and Small (SBSs). Therefore, two types of network topologies can be considered: homogeneous topology, in which only MBSs are simulated; or heterogeneous, in which there exits two tiers with different types of BSs. In this case, each type of BS will have different characteristics related to the power transmission and path loss model.

Among the set of BSs and UEs that are placed in the considered area, only a subset of those BSs and UEs maybe simulated by the *link level*. This is due to the fact that this level implements all the processes associated to the PHY & MAC layers of LTE/LTE-A and, therefore, the computational cost is high. The average interference generated by the non-simulated elements (BSs and UEs) is taken into account in the *link level* as Gaussian interference, added to the Additive White Gaussian Noise (AWGN) associated to the wireless channel. However, it should be noticed that the interference caused by the simulated subset is accurately considered in the *link level* part, taking into account all the PHY & MAC layer LTE/LTE-A features.

The tool allows to configure the number of MBSs which will be simulated. It should be noted that the final simulated set will consist of the selected number of MBSs and its related parameters. Nevertheless, it must be taken into account that the *link level* will simulate only those BSs with at least one associated UE.

The position of the elements in the network, i.e. MBSs, SBSs and UEs, can be configured separately. On the one hand, related to the BSs, the position of MBSs, which determines the layout of the network, can follow either an hexagonal grid or an irregular deployment, determined by a Poisson Point Process (PPP). The position of SBSs can be given by: (1) a PPP, so that the number of SBSs falling inside each MBS region is random; (2) a Binomial Point Process (BPP), where the number of SBSs per MBS is deterministic; or (3) it is possible select the location of each SBS in the plane. On the other hand, the locations of the simulated UEs can follow either a BPP or they can be selected as deterministic locations, both in the whole simulation area or creating hotspots areas with a high density of users.

Once the elements are positioned, the *system level* generates the following information:

- UEs to BSs association. This process can be done using two different criteria: smallest path loss [11] or higher Reference Signal Received Power (RSRP). A bias can be chosen in an heterogeneous deployment; it is possible to apply such an increment over the received power from SBSs to increase their coverage area. This technique, known as *Cell Range Expansion* (CRE) [8, 9], is used to balance the load between tiers, i.e., UEs are shifted from highly loaded MBSs to SBSs. However, these offloaded users will have a degraded SINR, since the bias yields to a situation where the UEs are not associated to the BS that provides the highest average power. As discussed in detail in Sect. 3, the higher the bias, the higher the degradation of the SINR of a given UE. Nevertheless, offloading may be beneficial for the UEs, since their reward is being granted with more bandwidth from partially loaded SBSs.
- Average received interfering power which affects to each simulated UE received from the rest of non-simulated BSs in the network.
- Signal gain between each simulated BS and UE.
- Average SINR associated to each simulated UE.
- Average transmission power for each simulated BS. It is possible to consider the classical case where all the BSs that belong to the same tier transmit with the same power. However, it is also possible to consider the case of DL power control and DL power setting [12], where each BS can adjust its transmission power so as to reach an appropriate balance between desired power and generated interference.
- Frequency planning. The frequency planning especifies the frequency bands that can be allocated for DL transmission towards each UE. The simulator includes the case of partial and fractional frequency resuse [13], as well as resource partitioning in the frequency domain to mitigate the inter-tier interference [8].

This information will be stored in a .mat file which will be used by the *link level*.

2.2 WM-SIMA: Link Level

This level has been developed in C++ and implements functionalities related to PHY and MAC of LTE-A systems. Figure 1 depicts the architecture associated to this level, which is based on five main blocks: Transmitter, Receiver, Extended MIMO Channel, MAC Scheduler and QoS Statistics.

Transmitter: This block is related to the BSs in the DL and it performs functionalities involved in the delivery of packets, from Radio Link Control (RLC) layer to PHY layer. More specifically, RLC packets are generated following one of these models: (1) *full-buffer*, where data is always available at the transmission queues; (2) *streaming*, in which packets are generated based on a Pareto distribution [14]; or (3) *trace files*, that contains arrival times and packet lengths.

Once packets have been generated, they are extracted from the queues according to the corresponding Transport Block (TB) size, which depends on the scheduling decisions (i.e., number of PRBs allocated to a given UE) and the selected Modulation and Coding Scheme (MCS). These scheduling decisions are signalled with the signal *allocationInfo*. Hence, the RLC layer segments and/or concatenates the received packets to accommodate the difference between RLC packet and TBs sizes related to a given UE.

The MAC layer is responsible for scheduling the PRBs and UEs, performing Adaptive Modulation and Coding (AMC) and Hybrid Automatic Repeat Request (HARQ) protocol. However, the scheduler is placed at a separated block in WM-SIMA for the sake of implementation flexibility. The receivers deliver Channel State Information (CSI) towards the transmitter by means of a return channel that can add error and/or delay in WM-SIMA.

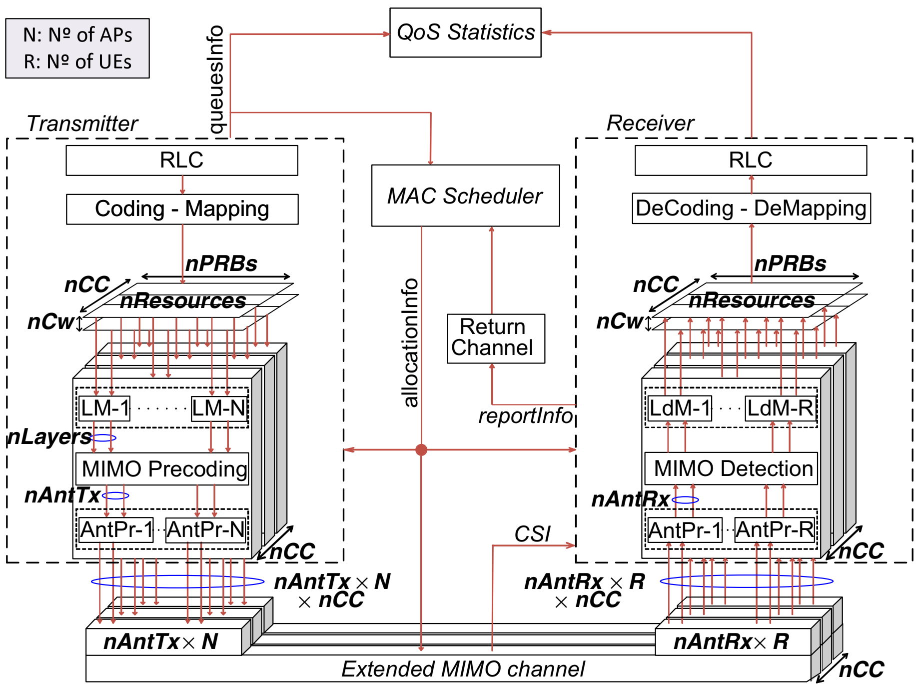


Fig. 1 Sketch of WM-SIMA *link level*

CSI consists on the Channel Quality Indicator (CQI), which is computed at the receiver side, and it is related with the MCS that maximizes the spectral efficiency while guaranteeing a Block Error Rate (BLER) below 10% of received TBs [15]. In the case of MIMO antenna configuration with a closed loop technique, the CSI information is also related to Rank Indicator (RI) and the Precoding Matrix Indicator (PMI), which is the index to the appropriate precoding matrix. HARQ consists on the retransmission of previously transmitted TBs, provided that they have not been successfully received, which is assumed if a NACK is received or if it is not received an ACK after a certain time (stop-and-wait time). In WM-SIMA the maximum number of retransmissions that can be performed and the number of HARQ channels is configurable.

Then, the channel coding is performed based on LTE specification. The Cyclic Redundancy Check (CRC) is computed from the bits of the TB. Then, if the TB is greater than 6144 bits, i.e. the maximum information block length, the TB si segmented into Code Blocks (CBs) smaller than 6144 bits and an additional CRC sequence is attached to each code block. After that, each CB is coded (e.g. turbo coding for user channels) with a mother coding rate of 1 / 3 and the rate matching is applied to reach the target MCS. Finally, the different CBs from the rate matching are concatenated and the resulting bits are mapped into M-Quadrature Amplitude Modulation (QAM) symbols according to the selected MCS (i.e., M can be 4, 16 or 64 in WM-SIMA).

Afterwards it is performed MIMO precoding as long as the number of antenna elements in the transmitter is greater than 1. In such a case the M-QAM symbols are multiplied by the selected precoding matrix if we have selected a closed-loop MIMO technique. WM-SIMA considers both closed-loop and open-loop Single-User (SU) and Multi-User (MU) MIMO multiplexing. In case of SU MIMO multiplexing, it is possible to transmit two TBs (codewords) per UE; whereas in case of MU MIMO, a single TB from different UEs is transmitted. Additionally, WM-SIMA considers beamforming and Space Frequency Block Coding (SFBC).

Finally, the OFDM signal is generated to be sent towards the radio interface. Firstly, each modulation symbol is mapped to a given resource element located in time and frequency. Then, the OFDM symbols are converted into the time domain by means of an N-point Inverse Fast Fourier Transform (N-IFFT). Afterwards, a cyclic prefix (CP) is added to each OFDM symbol. Finally, the power of the transmitted signal is normalized.

Most of PHY layer parameters are configurable, including FFT size, which is related to the system bandwidth in LTE, and the CP length, which can be either normal or extended.

Receiver This block simulates the UE at the receiver side and it is responsible for two main tasks. One is to maintain periodic reports with the transmitter in order to perform the link adaptation. These reports contain: (1) the information for the SINR estimation quantified in a CQI value; (2) the RI and the PMI for the MIMO adaptation in transmission; and (3) the ACK/NACK messages for HARQ processes. All this information can be ideally estimated, using the frequency response of the channel, or through the interpolation of the reference signals to have a realistic estimation.

The second task is the processing of the OFDM symbols received from the different antennas. Basically, the process is same as in the transmitter side but in inverse order. In the time domain, the CP is removed and then, in the frequency domain the OFDM symbols are converted through an N-FFT to obtain the modulated symbols from the resource elements.

Then, based on the information provided by control channels, each UE selects the PRBs that contains its related data. After performing channel estimation, which can be done either using the ideal frequency response or by estimating the channel from DL pilot symbols, equalization is performed. In the case of Single-Input and Single-Output (SISO) it is performed 1-tap Zero Forcing (ZF) equalization. For the case of multiple antennas at the receiver, WM-SIMA includes several MIMO detection techniques (i.e., ZF, maximum ratio combining, minimum mean squared error, etc.)

After that, received symbols are demodulated and decoded to obtain the transmitted bits and recover the original information block. In WM-SIMA, decoding algorithms can be either Soft-Output-Viterbi Algorithm (SOVA) or Max-Log MBS algorithms [16]. After turbo decoding, it is computed the CRC to check whether the decoded block is free of errors or not, and an ACK or NACK message is generated accordingly. WM-SIMA also implements an Outer Loop Link Adaptation (OLLA) algorithm to improve the accuracy in the selection of appropriate CQI index to report towards the transmitter.

Finally, the correctly received blocks are concatenated to form the RLC packets.

Extended MIMO Channel WM-SIMA considers an extended network MIMO channel, which considers all the links between antenna elements of each simulated BS and each simulated UE. Hence, having N BSs with n_T antennas each and R UEs with n_R antennas, the extended MIMO channel have $N \times n_T$ inputs and $R \times n_R$ outputs. The extended channel also considers the Kronecker model [17] among the n_T and/or n_R antenna elements that belong to the same transmitter or receiver, respectively, to add antenna correlation. It is considered a Tapped Delay Line (TDL) model, which allows to consider different channel power delay profiles with both Rayleigh and Rician fading. Each coefficient of the TDL model can be computed either by sum of sinusoids or filtered Gaussian noise methods.

Scheduler This block makes scheduling decisions based on the CSI information delivered by the receiver (i.e., UE) through a return channel. The scheduling policies in WM-SIMA are: (1) Round Robin, which equally assigns the radio resources to users in a cyclic manner; (2) Proportional Fair, which aims at allocating the resources to the UEs in their best time instants; and (3) channel-dependent earliest deadline due, which aims at reducing the average delay and packet loss [18]. The scheduling decisions are delivered towards the transmitter and receiver by means of the *allocationInfo* signal.

3 Use case: Cell Range Expansion

To show a use case with WM-SIMA, the simulation of an HetNet scenario with four SBSs in the coverage area of a MBS is performed. The aim of this study is to show the effect of the user association on the system performance due to the use of the CRE technique with a sweep of bias values. In this case, the MBS tier (Tier 1) and SBS tier (Tier 2) are configured in co-channel deployment, i.e. the same carrier is shared, leading to a trade-off between available BW and intercell interference. For this study, no interference management is employed to check whether this technique can be used by itself without any interference cancellation. Moreover, different distances between MBS-SBS cells have been taking into account.

3.1 Simulation Setup

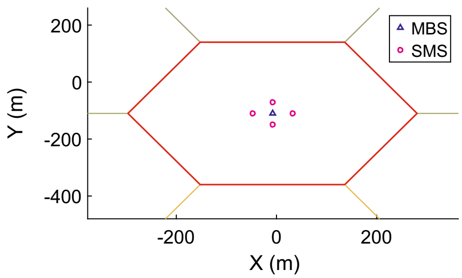
In this section, the system configuration as well as the scenario of simulations are described. All simulation parameters are shown in Table 1.

Table 1 Simulation setup

Parameter	Value
Carrier freq.	2 GHz
Bandwidth	10 MHz
Transmission power	43 dBm (MBS), 30 dBm (SBSs)
Simulation time	20,000 ms
Transmission mode	SISO
Number of users	60
Association criterion	Mx. RSRP+Bias
Distances MBS-SBSs	40, 100, 160 m
Hexagon apothem	250 m
Spatial distribution process	BPP
No. of spatial realization	8
Path losses	ITU UMa, ITU Umi
Fast fading	ITU UMa
Scheduler	Round Robin
Max. allocation per user	15 PRBs
Source	FTP model 3, $\lambda = 2.5$, 0.5 MB
Decoding algorithm	SOVA
Link adaptation method	OLLA
Channel estimation	Ideal
SNR estimation	Error based

Scenario The scenario has been chosen to identify the impact of the user association to different cells in heterogeneous networks. The simulation set consists of one MBS with four SBSs placed within the macro coverage area. The macro coverage area is hexagonal shaped with an apothem (inner radius) of 250 m. The SBSs are located around the macro cell at 0°, 90°, 180° and 270° in polar coordinated, being the MBS at the origin of coordinates. The effect of the distance between MBS and SBSs in an urban environment is analysed, i.e. from a location close to the MBS to a location near the cell edge. In particular, the distances taken into account are 40, 100 and 160 m, named as P-1, P-2 and P-3, respectively. In Figs. 2, 3 and 4 the scenario is shown graphically. A total number of 60 users are uniformly distributed with a minimum distance UE-MBS of 35 m. It has been generated 8 spatial realizations of the UE locations according to a Binomial Point Process. For each spatial realization, we have performed a simulation of 20,000 subframes. In Fig. 5, there is an example of spatial realization of users. Afterwards, results are averaged

Fig. 2 Scenario: P-1



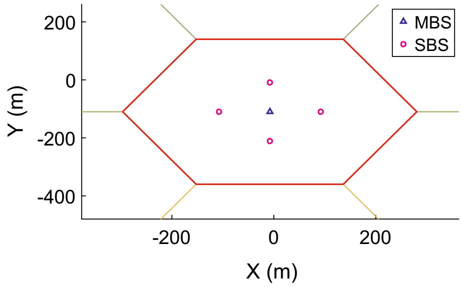


Fig. 3 Scenario: P-2

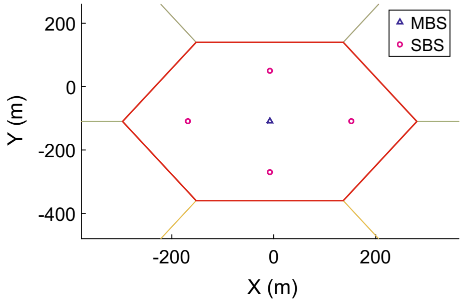


Fig. 4 Scenario: P-3

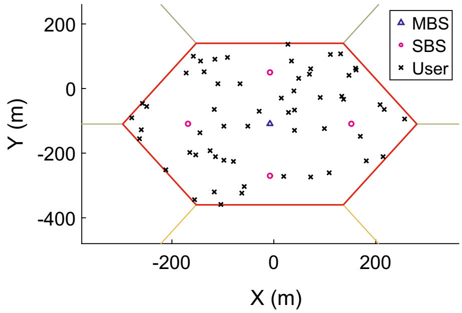


Fig. 5 Spatial Realization example with P-3

in the spatial and temporal domain to obtain the intended performance indicators. Moreover, the user association to cells is done by RSRP and taking into account the CRE bias with a range of values between 0 and 30 dB.

System Configuration The DL of the BSs is configured to transmit over a 2 GHz carrier with a bandwidth of 10 MHz. The transmission power of MBS and SBSs are 43 dBm and 30 dBm, respectively. For the layer 2 configuration, AMC along with the OLLA technique are used and the scheduler chosen is Round Robin. Path losses are calculated following the ITU Urban Macro (UMa) and Urban micro (Umi)[19]. The multi-path fading propagation conditions are generated according to a Rayleigh model with the ITU UMa configuration [20].

Data Source A File Transfer Protocol (FTP) model suggested by the 3GPP is used as data source. In particular, we consider the FTP model 3 [19], where packets arrive according to a Poisson process with a packet arrival time that follows an exponential distribution. Furthermore, the files have a size of 0.5 Mbytes.

3.2 Performance Indicators

In this section, the metrics under analysis are defined: user association per tier, PRB Load per Tier, aggregated throughput, percentage of low rate users and fairness.

User Association per Tier Represents the number of users that are associated to the tier 1 (MBS) and the number associated to the tier 2 (SBSs). Each user will be associated to the cell with the maximum RSRP measured plus the CRE bias.

PRB Load per Tier This is the percentage of PRBs used per tier. As the simulated bandwidth is 10 MHz, the maximum use of the tier 1 is 50 PRBs, whereas for the tier 2 is 200 PRBs (due to the four small cells).

Aggregated Throughput This is calculated as the aggregated bits received by all users (i.e both tiers) per second.

Percentage of Low Rate Users This metric is used to show the percentage of users that have a low transmission rate, i.e. less than 10 kbps, named as Low Rate User (LRU).

Fairness This metric is used to determine whether users are receiving a fair share of system resources. The Jain index is used to calculate the fairness of the simulated cells by

$$J(x_1, x_2, \ldots, x_n) = \frac{\left(\sum_{i=1}^{n} x_i\right)^2}{n \cdot \sum_{i=1}^{n} x_i^2} \tag{1}$$

where x_i is the total bits received by the user i and n is the number of users.

3.3 Numerical Results

In this section, simulation results are presented through different figures that show the network behaviour when CRE is used with different bias values. Note that, by increasing the bias, the users will be associated to a SBS with less load at the expense of lowering its SINR. It is a trade-off between available resources and SINR, as the bias grows the lowest SINR in the new serving cell. In other words, a high bias re-association causes users with low SINR. Consequently, the transfer of users from tier 1 to tier 2 due to a high bias value may cause users in outage (i.e. users that cannot transmit) or with too poor transmission rates.

In the Fig. 6, the user association for P-1, P-2 and P-3 versus bias is shown. On the one hand, solid lines represent the user association to tier 1 (macro cell). On the other hand, dashed lines represent the association to tier 2 (small cells). It can be seen that the intersection point where the 50% of the users are divided equally into both tiers is granted earlier as the small cells are further from the macro cell. In P-1, the macro cell is close to

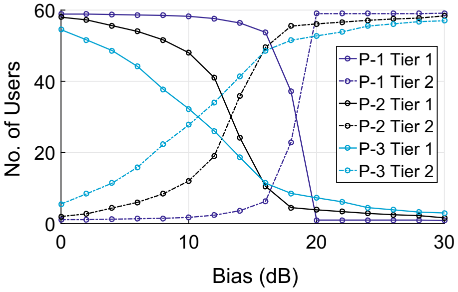


Fig. 6 User association per tier

the small cells and, therefore, it is difficult to balance users from the MBS due to the high power of the macro cell.

In addition, the load per tier (Fig. 7) is a result that is linked to the previous figure. At the beginning, when the bias is low, the macro cell is loaded at 100% and, when the number of users associated to the tier 2 increases, the macro cell is offloaded. Note that, although the number of users associated to tier 2 grows, not all users can transmit due to their low SINR. In Fig. 8, it can be seen that the number of users with low transmission rate increases when the bias grows because these users are being forced to be associated to a small cell but with a SINR that is not enough to transmit, mainly due to the macro cell interference. The difference between the three curves is at high bias values. For P-2 and P-3, the number of low rate users decreases with high bias values. This is due to the fact that the interference is reduced when the macro cell is loaded below 90% and, consequently, the users can transmit with better conditions. On the contrary, for P-1 the number of low rate users remains constant with a value of 80% because the closest user to the macro is always associated to the MBS.

In Fig. 9 the aggregated throughput of all simulated cells is plotted. For bias values lower than 10 dB, the throughput of P-1 and P-2 decreases progressively. This is because the worst users associated to the macro cell now are associated to the small cells and their transmission rate now is lower or null. However, for P-5 the throughput slightly increases for a bias of 2 dB as a result of the new user association. This is because with bias lower than 2 dB the users were associated to a high loaded macro cell. However, with a bias of 2 dB the users can transmit associated to a offloaded small cell. Then, for higher bias values the aggregated throughput grows quickly but this is due to two factors. On the one hand, more users are assigned to tier 2 but their SINR is so low that they can not transmit, so there are more resources to use in the macro cell. On the other hand, the main interference source (the macro cell) is less loaded so the interference of the area decreases and, consequently, the condition for the transmission is better.

Fig. 7 PRB Load per Tier

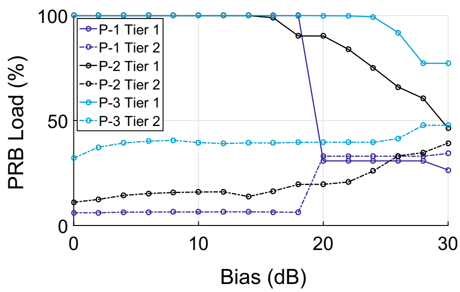
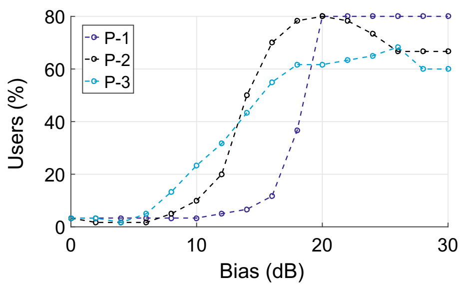


Fig. 8 Percentage of low rate users



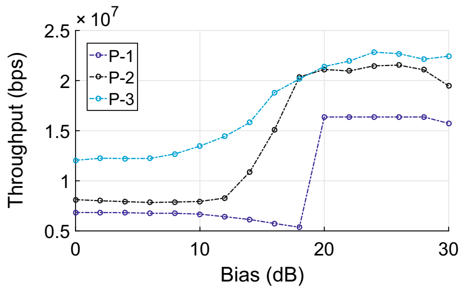


Fig. 9 Aggregated Throughput

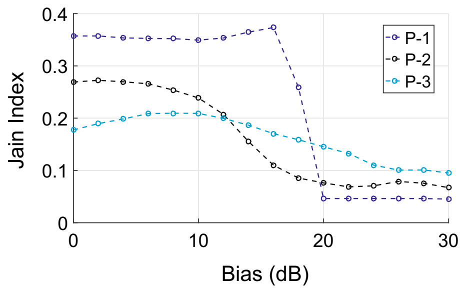


Fig. 10 Fairness

In addition, the fairness of the system is shown in Fig. 10. The fairness is closely related to the low rate users. For all of the three cases the optimum fairness is achieved when the low rate users are below 13%. At the beginning, most of the users improve their throughput and other users can be in outage but in general the system is fairer. When the optimum fairness is reached, more users become low rate/outage users and the fairness decreases.

4 Conclusions

In this paper it has been presented WM-SIMA, a flexible, accurate and efficient data-oriented LTE-A simulator. It has been depicted the architecture of this tool by means of its main building blocks. Then, as a case of study, it has been investigated the performance of cell offloading under FTP traffic. Results reveal that it is necessary to select appropriately the bias for cell association to achieve a good balance between fairness and aggregated throughput. Besides this, the appropriate bias greatly depends on the locations of small cell with respect towards the nearest macro access point. Results have remarked the necessity to apply enhanced interference mitigation techniques to cope with the SINR degradation of offloaded users.

References

1. Carreras, A., Aguayo-Torres, M. C., Martin-Vega, F. J., Gomez, G., Blanquez, F., Delgado-Luque, I. M., & Entrambasaguas, J. T. (2018). Link abstraction models for multicarrier systems: A logistic regression approach. *International Journal of Communication Systems*. https://doi.org/10.1002/dac.3436

2. Mehlfhrer, C., Colom Ikuno, J., Šimko, M., Schwarz, S., Wrulich, M., & Rupp, M. (2011). The Vienna LTE simulators-enabling reproducibility in wireless communications research. *EURASIP Journal on Advances in Signal Processing*, *2011*(1), 29. https://doi.org/10.1186/1687-6180-2011-29.
3. Virdis, A., Stea, G., & Nardini, G. (2014). SimuLTE – A modular system-level simulator for LTE/LTE-A networks based on OMNeT++. In *2014 4th International Conference On Simulation And Modeling Methodologies, Technologies And Applications (SIMULTECH)*, pp. 59–70. https://doi.org/10.5220/0005040000590070.
4. Piro, G., Grieco, L. A., Boggia, G., Capozzi, F., & Camarda, P. (2011). Simulating LTE cellular systems: An open-source framework. *IEEE Transactions on Vehicular Technology*, *60*(2), 498–513. https://doi.org/10.1109/TVT.2010.2091660.
5. Baldo, N., Miozzo, M., Requena-Esteso, M., & Nin-Guerrero, J. (2011). An open source product-oriented LTE network simulator based on ns-3. In *Proceedings of the 14th ACM International Conference on Modeling, Analysis and Simulation of Wireless and Mobile Systems*. ACM, New York, MSWiM '11, pp. 293–298. https://doi.org/10.1145/2068897.2068948.
6. Aguayo-Torres, M.C. et Al. (2016). WM-SIM LTE link simulator version 2. Tech. rep., Universidad de Málaga. https://riuma.uma.es/xmlui/handle/10630/11057?show=full.
7. Andrews, J. G., Singh, S., Ye, Q., Lin, X., & Dhillon, H. S. (2014). An overview of load balancing in hetnets: Old myths and open problems. *IEEE Wireless Communications*, *21*(2), 18. https://doi.org/10.1109/MWC.2014.6812287.
8. Singh, S., & Andrews, J. (2014). Joint resource partitioning and offloading in heterogeneous cellular networks. *IEEE Transactions on Wireless Communications*, *13*(2), 888. https://doi.org/10.1109/TWC.2013.120713.130548.
9. Damnjanovic, A., Montojo, J., Wei, Y., Ji, T., Luo, T., Vajapeyam, M., et al. (2011). A survey on 3GPP heterogeneous networks. *IEEE Wireless Communications*, *18*(3), 10. https://doi.org/10.1109/MWC.2011.5876496.
10. Ali, M. S., Coucheney, P., & Coupechoux, M. (2016). Load balancing in heterogeneous networks based on distributed learning in near-potential games. *IEEE Transactions on Wireless Communications*, *15*(7), 5046. https://doi.org/10.1109/TWC.2016.2551744.
11. Martin-Vega, F. J., Gomez, G., Aguayo-Torres, M. C., & Renzo, M. D. (2016). Analytical modeling of interference aware power control for the uplink of heterogeneous cellular networks. *IEEE Transactions on Wireless Communications*, *15*(10), 6742. https://doi.org/10.1109/TWC.2016.2588469..
12. Martin-Vega, F. J., Delgado-Luque, I. M., Gomez, G., Aguayo-Torres, M. C., & Entrambasaguas, J. T. (2016). Downlink power setting for energy efficient heterogeneous cellular networks. In *2016 8th International congress on ultra modern telecommunications and control systems and workshops (ICUMT)*, pp. 147–151. https://doi.org/10.1109/ICUMT.2016.7765348.
13. Delgado-Luque, I. M., Aguayo-Torres, M. C., Gómez, G., & Entrambasaguas, J. T. (2017). A framework to evaluate fairness in fractional frequency reuse based cellular networks. *Wireless Personal Communications*, *95*(2), 287. https://doi.org/10.1007/s11277-016-3892-6.
14. Xie, M., Ye, W., Feng, S., & Song, H. (2005). Performance of a queuing model with Pareto input traffic for wireless network nodes. In *Proceedings of international conference on wireless communications, networking and mobile computing, 2005*, Vol. 2, pp. 867–870. https://doi.org/10.1109/WCNM.2005.1544188.
15. Sesia, S., Toufik, I., & Baker, M. (2009). *LTE, The UMTS long term evolution: From theory to practice*. New York: Wiley Publishing.
16. Martin-Vega, F., Blanquez-Casado, F., Lopez-Martinez, F., Gomez, G., & Entrambasaguas, J. (2015). Further improvements in SOVA for high-throughput parallel turbo decoding. *IEEE Communications Letters*, *19*(1), 6. https://doi.org/10.1109/LCOMM.2014.2371041.
17. Han, B., & Zheng, Y. R. (2015). Higher rank principal Kronecker model for triply selective fading channels with experimental validation. *IEEE Transactions on Vehicular Technology*, *64*(5), 1654. https://doi.org/10.1109/TVT.2014.2332518.
18. Delgado-Luque, I. M., Blanquez-Casado, F., Fuertes, M. G., Gomez, G., Aguayo-Torres, M. C., Entrambasaguas, J. T., & Baños, J. (2012). Evaluation of latency-aware scheduling techniques for M2M traffic over LTE. In *Proceedings of the 20th European signal processing conference (EUSIPCO)*, pp. 989–993.
19. 3GPP, Evolved Universal Terrestrial Radio Access (E-UTRA); Further advancements for E-UTRA physical layer aspects. TR 36.814, 3rd Generation Partnership Project (3GPP) (2017).
20. 3GPP, Technical Specification Group Radio Access Network; Coordinated multi-point operation for LTE physical layer aspects. TR 36.819, 3rd Generation Partnership Project (3GPP) (2013).

Alberto Carreras received the B.E. degree in telecommunications engineering from the University of Málaga, Spain, in 2015, and the M.Tech. degree in mobile communication engineering in 2017. In 2015, he joined the Department of Communication Engineering, University of Málaga, where he is currently associated researcher. His main areas of research interest are channel estimation techniques, load balancing and interference coordination in mobile heterogeneous networks and C-RAN.

Isabel M. Delgado-Luque received her B.Sc. degree in Telecommunications Engineering from the University of Malaga (Spain) in 2007. She has been working as a researcher at the Department of Communications Engineering at the University of Malaga since 2008. Her research interests include the field of mobile communications, especially, channel modeling and radio resource management strategies for LTE and LTE-Advanced.

Francisco J. Martín-Vega received his B.Sc. degree in 2011, M.Sc. in 2014 and PhD degree in 2017. He joined the Communication Engineering Department at the University of Malaga in 2011, as an associate researcher, where he participates in contracts with several industry partners related to cellular and satellite communications. He was awarded by the Official College of Telecommunication Engineers of Spain with the best master thesis in accessible technologies. His current research activity is focused on the mathematical modeling of heterogeneous networks and vehicular communications.

Gerardo Gómez received the B.Sc. and Ph.D. degrees in telecommunication engineering from the University of Málaga, Spain, in 1999 and 2009, respectively. From 2000 to 2005, he was with Nokia Networks and Optimi Corporation (acquired by Ericsson), leading the area of quality of service (QoS) for 2G and 3G cellular networks. In 2005, he joined the University of Málaga, where he is currently an Associate Professor and the Deputy Director of the Communications Engineering Department. His research interests include the field of mobile communications, especially physical layer security, stochastic geometry, interference management, and QoS/QoE evaluation for multimedia services.

Mari Carmen Aguayo-Torres received the Ph.D. degree in telecommunication engineering from the University of Málaga, Spain, in 2001, and the M.S. degree in 1994. She is currently a e Professor with the Department of Communications Engineering, University of Málaga. Her main research interests are 5G technologies with probabilistic QoS guarantees. Her work includes theoretical analysis with tools such as stochastic geometry besides statistical evaluation of current real or simulated networks. She is also involved in a number of public and private funded projects and actively collaborates with industry, mainly in the field of wireless communications.

J. Tomás Entrambasaguas received M.Sc. and Ph.D. degrees in telecommunication engineering from the Polytechnic University of Madrid (UPM), Spain, in 1975 and 1990, respectively. From 1975 to 1978 he was with E.T.S. Ingenieros de Telecomunicación, UPM, Madrid. From 1978 to 1993 he worked at Fujitsu-España R&D Center, mainly on the development of packet switching systems, data transmission, and hardware/software technology for computer local area networks. In 1993 he joined the University of Málaga, where he is now a professor and head of the Communication Engineering Department. His current research interests include digital signal processing, wireless access networks and methodologies for development and testing of complex communications systems.

Wireless Pers Commun (2017) 95:287–297
DOI 10.1007/s11277-016-3892-6

A Framework to Evaluate Fairness in Fractional Frequency Reuse Based Cellular Networks

Isabel M. Delgado-Luque[1] · M. Carmen Aguayo-Torres[1] · Gerardo Gómez[1] · José T. Entrambasaguas[1]

Published online: 8 December 2016

Abstract The focus in Mobile communication networks has been addressed to high data rate networks that offer high quality of service. Long Term Evolution Advanced (LTE-A) technology has been the first to completely fulfill 4G requirements. However, performance gains remain limited due to severe interference levels. For this reason, the Inter-Cell Interference (ICI) problem is one of the main challenges in LTE-A systems. In order to deal with this ICI problem, in this paper we propose a novel framework to analyze the fairness of several scheduling techniques when Fractional Frequency Reuse (FFR) schemes are used. We evaluate the system performance of a typical FFR-based LTE-A scenario. Performance results show the optimum values of the Signal to Interference plus Noise Ratio and partition bandwidth for frequency reuse that maximizes the spectral efficiency. Fairness evaluation has been done in terms of the Gini index, showing a trade-off between spectral efficiency and fairness among users. In that sense, the proposed Truncated-Equal Transmission Rate scheduling policy is able to adjust this trade-off by means of a parameter ε.

Keywords Fairness · FFR · Gini index · Scheduling

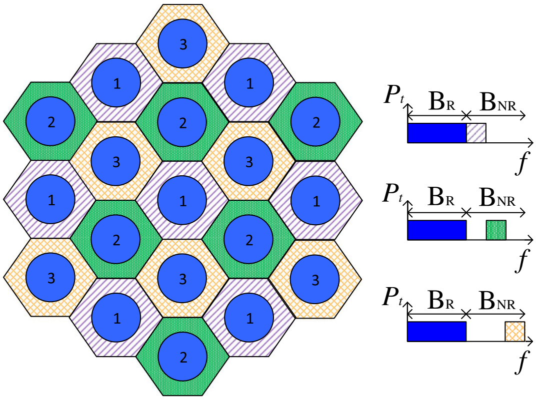
1 Introduction

Mobile networks, like WiMax or Long Term Evolution Advanced (LTE-A) are targeted to support aggressive frequency reuse patterns in order to maximize the spectral efficiency. However, this type of frequency planning leads to a considerably increase of Inter-Cell

✉ Isabel M. Delgado-Luque
imdl@ic.uma.es

[1] Departamento de Ingeniería de Comunicaciones, Universidad de Málaga, Campus de Teatinos s/n, 29071 Málaga, Spain

Fig. 1 Example of cellular topology with a FFR scheme



Interference (ICI), which may cause an important performance degradation, especially to users located at the cell edge.

ICI Coordination (ICIC) [1] has been widely investigated as a key technology to alleviate the impact of interference. Among the existing ICIC techniques, this paper focuses on Fractional Frequency Reuse (FFR). These schemes aim at reducing the interference at the cell edge by applying a looser frequency reuse pattern to cell-edge users, thus reducing the interference for such users [2]. However, high efficiency is kept by aggressive reuse 1 pattern for those users with high Signal to Interference plus Noise Ratio (SINR), typically located near the base station (BS).

In LTE-A, Orthogonal Frequency Division Multiple Access (OFDMA) eases the use of FFR schemes as the whole band can be quite straightforwardly divided into two different reuse patterns. Figure 1 shows an example of FFR scenario: a common frequency sub-band, B_R, is employed in all cells (i.e. with a frequency reuse of 1) whereas the use of the remaining band, B_{NR}, is coordinated among neighboring cells in order to create sub-bands with frequency reuse higher than 1 (commonly 3) [2].[1] Users are assigned to reuse 1 or reuse 3 sets by being signaled to receive data on resources allocated in the appropriate sub-band.

It is clear that one objective of the FFR is to provide a fairer treatment to users located in the edge of the cells. However, only a small set of works on FFR addresses the issue of fairness among users. Certain interest is shown in [3], where 5% worse user throughput is studied besides aggregate sector throughput for a coordinated FFR scheduling algorithm. The performance of an opportunistic scheduling over an FFR scheme is evaluated in [4], considering the fairness requirements of the users. In [2] all possible bandwidth allocations for reuse 1 and 3 are evaluated, and that maximizing the 'user satisfaction' (a metric on how close the users throughput is to the maximum throughput in the area) is selected. However these works do not allow easy comparison among allocation techniques with different average rates.

In order to evaluate fairness in terms of user rate, several metrics can be considered. The well known Gini index [5], originally proposed to study inequality in economics, is able to

[1] It would be also possible that each cell transmits in the whole frequency band adjusting the power allocation, P_t, of each sub-band in order to minimize the interference [2], but this issue is out of the scope of this work.

capture the overall fairness of distribution and was also proposed to study the fairness of schedulers in the Internet [6].

This paper provides a general framework for an FFR-based evaluation and deployment. A general system model allowing user classification in reuse 1 or reuse 3 sets is first given in Sect. 2. Section 3 describes the scheduling policies analyzed in this paper while in Sect. 4 Gini index is proposed as a fairness evaluation figure. Detailed insight is given for different resource allocation schemes in a specific scenario of hexagonal cells in Sect. 5. Finally, Sect. 6 concludes the work.

2 System Model

We consider a system with an arbitrary number P of cells. One of the cells will be the observed cell while the remaining $P-1$ cells will be treated as interfering cells; nevertheless, performance results could be extended to any other cell if a cell wrap around model is assumed. An FFR method is used as the frequency allocation scheme, where each cell is divided into two geographical regions referred as reuse factor 1 (reuse 1) and reuse factor 3 (reuse 3) regions.

The total available system bandwidth B is split into four non-overlapped orthogonal frequency bands denoted by B_R, B_{NR1}, B_{NR2} and B_{NR3} obeying $B = B_R + B_{NR}$ and $B_{NR} = B_{NR1} + B_{NR2} + B_{NR3}$. B_R represents the available bandwidth for the reuse 1 region and B_{NRi} (where $i \in [1, 2, 3]$) are three sets of equal-sized sub-bands that represent the available bandwidth for the reuse 3 region; each one of these sub-bands will be allocated to cells in a manner that no other surrounding cell is using the same sub-band, like in a classical frequency planning.

In the observed cell we consider N users, each of them with an associated location s_k, $k \in [1 \ldots N]$.

2.1 Mean SINR

The mean SINR associated to the kth user depends on the specific cell region (reuse 1 or reuse 3) the user has been associated to. In a generic way, mean SINR associated to the kth user, $SINR_k$, will be equal to $SINR^{R1}(s_k)$ if it is located in the reuse 1 region or will be equal to $SINR^{R3}(s_k)$ if it is located in the reuse 3 region. $SINR^{R1}$ and $SINR^{R3}$ represents the mean SINR associated to reuse 1 and reuse 3 region of the cell, respectively.

The above affirmation reflects the fact that $SINR^{R1}$ and $SINR^{R3}$ regions will have different values because of the different amount of interference that could be expected within each region. Indeed, users located in the reuse 1 region of a cell receive the interference of the rest of the cells in the network because the reuse 1 region of all the cells uses the same spectrum B_R (reuse factor of 1); however, if a user is located in the reuse 3 region the number of interfering cells is reduced by a factor of three because the adjacent cells are assigned different frequency sub-bands for their transmission in that region.

2.2 Potential Rate

We defined the *potential rate*, r_{0k}, associated to the kth user as the achievable throughput of user k if all the available resources are allocated to him. Since r_{0k} is a function of the mean SINR associated to the user, r_{0k} may have different values depending on the region of the cell to which the user is associated. Specifically, if user k is associated to reuse 1 region, r_{0k} will be equal to $c(SINR^{R1}(s_k))$, whereas if it is in reuse 3 region r_{0k} will be equal to $\frac{1}{3}c(SINR^{R3}(s_k))$. The concrete expression to compute the achievable throughput, $c(SINR)$, will be presented in Sect. 5.

2.3 Allocation of Users into Each Cell Region

The objective is to find an optimal classification that leads to a maximum rate for each user. This is equivalent to find an optimal SINR threshold since r_{0k} depends on the mean SINR associated to the user. Thus, the following expression will be applied to classify users:

$$\begin{aligned}&\text{If } c\big(SINR^{R1}(s_k)\big) \geqslant \frac{1}{3}\big(SINR^{R3}(s_k)\big), \quad k \in U^{R1} \Rightarrow r_{0k} = c\big(SINR^{R1}(s_k)\big)\\ &\text{If } c\big(SINR^{R1}(s_k)\big) < \frac{1}{3}\big(SINR^{R3}(s_k)\big), \quad k \in U^{R3} \Rightarrow r_{0k} = \frac{1}{3}c\big(SINR^{R3}(s_k)\big)\end{aligned} \tag{1}$$

where U^{R1} and U^{R3} represents the set of users associated to reuse 1 and reuse 3 regions, respectively.

3 Scheduling Policy

Let ρ_k be the fraction of resources (bandwidth and time) assigned to the kth user, the following constraint must be satisfied if all available resources are allocated to users: $\sum_{k=1}^{N}\rho_k = 1$. Moreover, if we consider an FFR-based deployment, the following two additional constraints must be taken into account in order to dimension appropriately the frequency bands associated to reuse 1: $B_R/B = \sum_{k\in U^{R1}}\rho_k$, and reuse 3: $B_{NR}/B = \sum_{k\in U^{R3}}\rho_k$, regions, according to the user distribution within each region.

The real amount of resources allocated to a user depends on the scheduling policy selected, this is, the value of ρ_k. Thus, the *allocated transmission rate* for user k can be computed as $r_k = \rho_k \cdot r_{0k}$. Taking the above definition into account, the *sum rate* of the cell, R_N, is determined by the concrete scheduling policy applied and it can be calculated as $R_N = \sum_{k=1}^{N} r_k = \sum_{k=1}^{N}\rho_k \cdot r_{0k}$. Regarding the specific scheduling policy, which determines the value of ρ_k, many different schemes can be applied, some of them are described next.

3.1 Equal Resource Sharing (ERS)

Time/frequency resources are equally shared among users independently of their potential rate, i.e. $\rho_k = 1/N$. Thus, the sum rate is $R_N^{ERS} = 1/N\sum_{k=1}^{N} r_{0k}$ and the ratio between reuse 1 and 3 frequency bands and B should satisfy $B_R/B = N^{R1}/N$ and $B_{NR}/B = N^{R3}/N$ where N^{R1} and N^{R3} is the number of users allocated to reuse 1 and 3 frequency bands, respectively.

3.2 Equal Transmission Rate (ETR)

Resources allocation is inversely proportional to the user potential rates (i.e. more resources are allocated to users with poor SINR) so that actual transmission rates are equally distributed among users, i.e. $\rho_k = K_1/r_{0k}$, where $K_1 = \left(\sum_{k=1}^{N} 1/r_{0k}\right)^{-1}$ to satisfy resource allocation constraint. In this case, the sum rate can be computed as $R_N^{ETR} = N \cdot K_1$ and the ratio between reuse 1 and 3 frequency bands and B shall satisfy $B_R/B = 1/N \cdot \sum_{k \in U^{R1}} R_N^{ETR}/r_{0k}$ and $B_{NR}/B = 1/N \cdot \sum_{k \in U^{R3}} R_N^{ETR}/r_{0k}$, respectively.

3.3 Truncated-ETR

This is a modified version of the previous scheduling policy that considers a maximum value of ρ_k to avoid allocating too many resources to users receiving very low SINR, i.e.: $\rho_k = K_2 \cdot \min\{K_1/r_{0k}, \varepsilon/N\}$, where $K_2 = \left(\sum_{k=1}^{N} \min K_1/r_{0k}, \varepsilon/N\right)^{-1}$ and $\varepsilon \in [0, \infty]$ is a truncation factor with respect to an ERS approach. Then, ε/N means that users with poor SINR will get at maximum ε times the resources that would be allocated to them if an ERS were used. By varying the value of ε, this scheduling policy may behave very differently. Indeed, the sum rate $R_N^{\varepsilon} = K_2 \sum_{k=1}^{N} \min R_N^{ETR}/(N \cdot r_{0k}), \varepsilon/N$ will be within the limits of previous algorithms $R_N^{ETR} = \lim_{\varepsilon\to\infty} R_N^{\varepsilon} \leqslant R_N^{\varepsilon} \leqslant \lim_{\varepsilon\to 0} R_N^{\varepsilon} = R_N^{ERS}$.

4 Fairness Indicators

Analysis of fairness could involve studying the distribution of the amount of resources (bandwidth and time) allocated to each user. However, as different SINRs lead to distinct potential rate, allocated rate instead of allocated resources is analyzed in this section by Lorenz curve and Gini index evaluation [5].

Let $\langle r_k, k = 1\ldots N\rangle$ be the ordered set of allocated rates obtained by resources allocated to users under certain scheduling policy. Rate accumulated by the first n of those users is given by $R_n = \sum_{k=1}^{n} \cdot$

Lorenz curve (see Fig. 2a) plots the percentage of accumulated number of users normalized to the total number of users, *n*/*N*, on the x-axis whereas the accumulate rate of those n users normalized to the sum rate, R_n/R_N is shown on the y-axis. The line at 45 degrees represents perfect equality in allocated rate Gini coefficient, i.e. ETR policy, captures the statistical dispersion of r_k by accumulating the deviation of capacity share

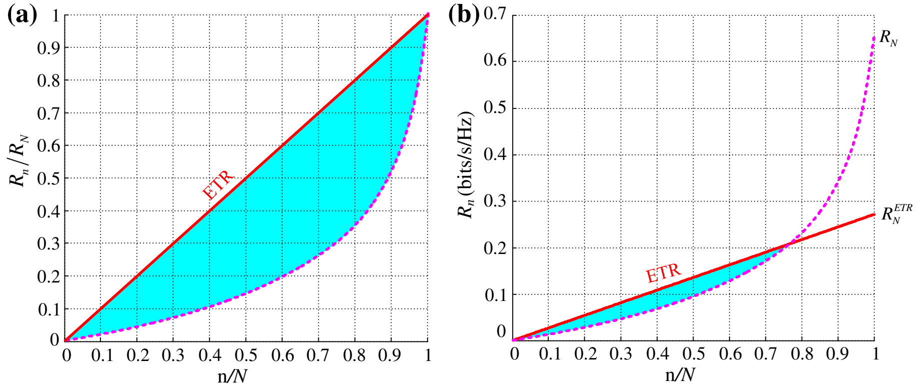


Fig. 2 **a** Lorenz curve **b** Generalized Lorenz curve

from the capacity share that corresponds to perfect equality. Gini index can be thought as the ratio of the area that lies between the line of equality and the Lorenz curve over the total area under the line of equality. This coefficient can range from zero (perfect equality, i.e., all values are the same) to one (maximal inequality among values).

Lorenz curve can be generalized using the non normalized value of R_n in the y-axis, resulting in Fig. 2b). As will be seen in the next section, this representation is able to show the allocated rate exchange between advantaged and disadvantaged users for certain ε.

5 Performance Evaluation

In this section, we evaluate the fairness of the resource allocation schemes presented in Sect. 3 over the proposed FFR-based deployment.

5.1 Analyzed Scenario

We consider the downlink of a LTE-A cellular network with 61 hexagonal cells, each cell with a radius, R, of 1 km. We assume that BSs using a Single Input Single Output (SISO) scheme with omnidirectional antennas are located at the cell centers and with the same transmit power, $P_t = 49$ dBm. The scenario assumed is a dense urban area served by macro-cells.

There are a number of users randomly distributed in the observed cell whose positions are assumed to be i.i.d. random variables uniformly distributed in the cell.

Signal received by users from the observed cell suffers from path loss, which is computed using (2), and small scale fading, which is Rayleigh distributed. Additionally, in order to ease the numerical evaluation, we have considered that the interference received from the rest of the cell in the system is Gaussian distributed [4] and its signal strength is only affected by path loss. Moreover, we assume that the system is interference limited for simplicity. Thus, SINR can be described as Rayleigh.

$$\begin{array}{ll} 98.42 + 20log_{10}(d) & \text{if} \quad d < 0.5\,\text{km} \\ 140.35 + 38log_{10}(d) & \text{if} \quad 0.5 \leq d < 1\,\text{km} \\ 140.36 + 35.04log_{10}(d) & \text{if} \quad d \geq 1\,\text{km} \end{array} \tag{2}$$

In order to determine the potential rate, we have applied an accurate computation considering the constraints established in an LTE-A system. In such systems, AMC is employed in order to maximize the throughput while keeping the Block Error Rate (BLER) under a predefined target ($BLER_t$). This goal can be achieved by selecting the optimal Modulation and Coding Scheme (MCS) subject to the $BLER_t$, 0.1 in this work, among a set of 16 possible schemes. Thus, the transmission rate can be computed by considering the probability of using each channel MCS given an average SINR. Assuming that MCS i is associated with a modulation rate of b_i bits/symbol and a coding rate c_i, the effective data rate when using MCS i can be computed as $r_i = b_i \cdot c_i$. In a Rayleigh fading channel, the probability of selecting a particular MCS i is given by:

$$p_i = \exp\left(-\frac{\gamma_i}{\bar{\gamma}}\right) - \exp\left(-\frac{\gamma_{i+1}}{\bar{\gamma}}\right). \tag{3}$$

where γ_i and γ_{i+1} are the lower and upper SINR thresholds associated to MCS i, respectively; and $\bar{\gamma}$ is the average SINR. Then r_{0k} can be obtained by averaging individual transmission rates at each MCS: $r_{0k} = \sum_{i=0}^{15} r_i \cdot p_i$.

5.2 Results and Discussions

Two well-differentiated set of results will be presented. Firstly, the analysis of system performance over the FFR scheme proposed will be discussed. Secondly, the fairness aspects of the scheduling policies presented will be evaluated.

5.2.1 Analysis of System Performance

Figure 3a depicts the evolution of the SINR associated to each region as a function of the normalized distance between a user and the BS in the observed cell. It can be seen that $SINR^{R3}$ is always greater, as it was expected, than $SINR^{R1}$. In Fig. 3b it can be seen how the reduction in the interference level is achieved at the expense of reducing the achievable throughput in this region (dashed line). Indeed, Fig. 3b represents the achievable throughput if a reuse factor of 1 (continuous line)/reuse factor of 3 (dashed line) is used in the whole cell as a function of the SINR. As stated in (1), it can be seen that the way to maximize the potential rate is to select the higher value of both expressions at each SINR. It should be noted that the SINR threshold which leads to the maximum potential rate for each user is the same independently of the scheduling policy selected and it can be seen from Fig. 3b) that has a value of 9.3 dB. In our work, since shadowing is not considered, a SINR threshold is equivalent to a distance threshold, X_f , with a value of $0.515R$. This means that reuse 1 and reuse 3 regions can be renamed as inner and outer region, respectively, as it is usually done [2].

Once the optimal SINR threshold has been established, we are going to analyze the performance of the system under the three scheduling policies considered. Figure 4 depicts the attainable rate Fig. 4a and the split of the system bandwidth Fig. 4b associated to each scheduling policy considered in our work as a function of the parameter ε. It can be seen in Fig. 4a that the value of the sum rate associated to ERS strategy is higher than that

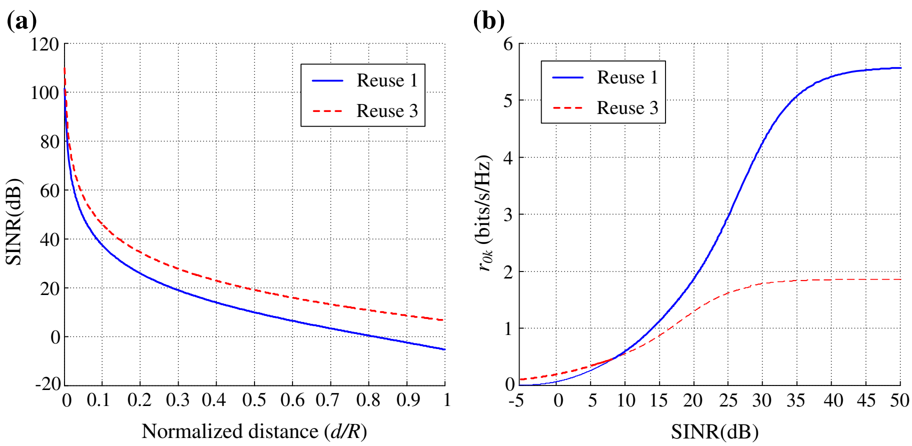


Fig. 3 **a** SINR reuse 1 and 3 **b** potential rate reuse 1 and 3

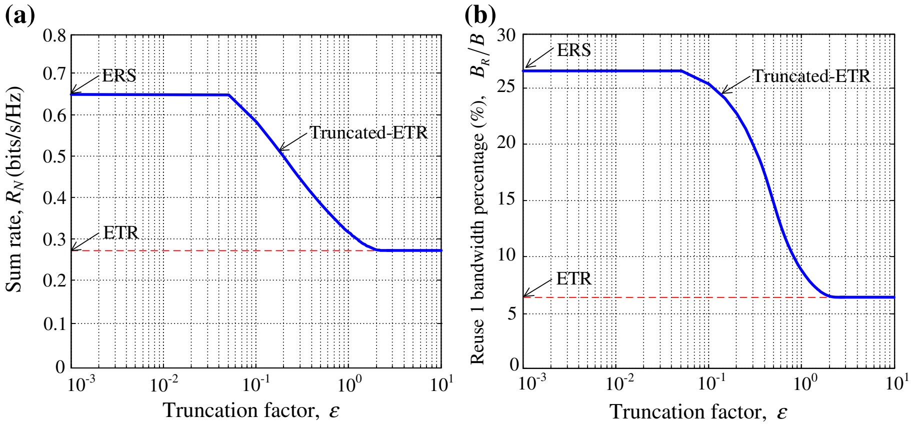


Fig. 4 **a** System attainable rate, **b** partition of system bandwidth

associated to ETR scheduling policy. Additionally, in Fig. 4b, it can be seen that ETR policy assigns almost the 95% of the bandwidth to reuse 3 users while ERS strategy assigns more than the 25% of the whole bandwidth to reuse 1 users. It should be noted how results associated to Truncated-Equal Transmission Rate (Truncated-ETR) scheduling strategy evolves from those obtained with ERS scheduling policy (where $\varepsilon \to 0$) to those associated to ETR technique (where $\varepsilon \to \infty$).

5.2.2 Analysis of Fairness

The Gini index associated to the Truncated-ETR scheduling policy as a function of the parameter ε and as a function of the sum rate of the observed cell is shown in Fig. 5a, b, respectively. It can be seen in Fig. 5a how the Gini index increases as the parameter ε decreases. This is because, as ε decreases, certain resources are transferred from users with lower potential rates r_{0k} to others whose use of them is more efficient, thus increasing the

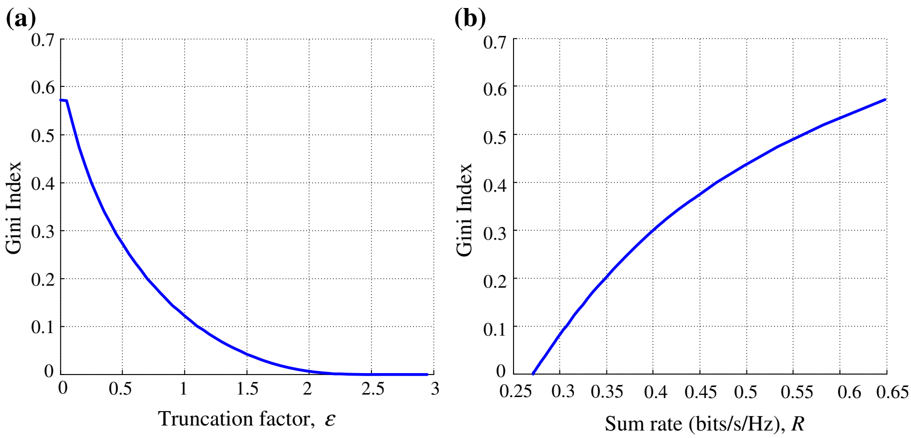


Fig. 5 Truncated-ETR Gini index evolution **a** versus truncation factor, **b** versus sum rate

inequalities among them. Finally, when $\varepsilon \rightarrow 0$, the truncated-ETR policy allocates all users the same amount of resources, but those with higher SINR exploit them in a more efficient way. Thus, ERS policy achieves a higher accumulated rate R_{ERS} at the expense of providing unfairness among users.

6 Conclusions

In this paper, we have focused on the analysis of the system performance over an FFR based deployment under three resource allocation schemes. Moreover, the fairness evaluation for scheduling policies considered over this deployment has been presented. We have proposed a method to determine the SINR threshold in order to maximize the rate of each user and we have found that the value of this threshold is the same for the three scheduling policies analyzed. The sum rate and the partition bandwidth of each cell of the system under each scheduling strategy have been calculated. In addition, the fairness evaluation has been done in terms of the Gini index. The highest sum rate is achieved with the ERS strategy whereas the lowest is associated to the ETR scheduling policy.

Furthermore, the fairness analysis shows that the fairest strategy is the ETR scheduling policy, which achieves the lowest Gini index value. This result reveals that fairness is achieved at the expense of a lower average efficiency. A trade-off between fairness and average efficiency can be found using the Truncated-ETR scheduling policy with the parameter ε. The concrete value of this parameter can be chosen in order to meet the requirements of the system, e.g. in terms of limiting the number of disadvantaged users; the relation between the gain achieved in the average efficiency and the maximum admissible sum rate lost for disadvantaged users, etc.

Acknowledgements This work has been partially supported by the Spanish Government under the Project TEC2013-44442-P.

References

1. Hamza, A.-B., et al. (2011). Inter-cell interference coordination. In *4G++ Project: Advanced performance boosting techniques in 4th generation wireless systems*, Deliverable D4.1.
2. Bilio, D., Bouras, C., Kokkinos, V., Papazois, A., & Tseliou, G. (2012). Optimization of fractional frequency reuse in long term evolution networks. In *Proceedings of IEEE wireless communications and networking conference* (pp. 1853–1857).
3. Necker, M. C. (2007). Coordinated fractional frequency reuse. In *Proceedings of ACM symposium on modelling, analysis, and simulation of wireless and mobile systems* (pp. 296–305).
4. Ali, S. H., & Leung, V. C. M. (2009). Dynamic frequency allocation in fractional frequency reused OFDMA networks. *IEEE Transactions on Wireless Communications*, *8*(8), 4286–4295.
5. Barnett, N. S., & Dragomir, S. S. (2008). *Advances in inequalities from probability theory and statistics*. New York: Nova Science Publishers.
6. Shi, H., Seghu, H., & Kanhere, S. S. (2005). An evaluation of fair packet schedulers using a novel measure of instantaneous fairness. *Compute Communications*, *28*, 1925–1937.

Isabel M. Delgado-Luque received her B.Sc. degree in Telecommunications Engineering from the University of Málaga (Spain) in 2007. She has been working as a research at the Department of Communications Engineering at the University of Málaga since 2008. Her research interests include the field of mobile communications, especially, channel modeling and radio resource management strategies for LTE and LTE-Advanced.

M. Carmen Aguayo-Torres received M.S. and Ph.D. degrees in Telecommunications Engineering from the University of Mĺaga, Spain, in 1994 and 2001, respectively. Currently, she is working at the Department of Communications Engineering, at the same university. Her main research interests include adaptive modulation and coding for fading channels, multi-user OFDM, SC-FDMA, cross-layer design, and probabilistic QoS guarantees for wireless communications.

Gerardo Gómez received his B.Sc. and Ph.D. degrees in Telecommunications Engineering from the University of Málaga (Spain) in 1999 and 2009, respectively. From 2000 to 2005, he worked at Nokia Networks and Optimi Corporation (recently acquired by Ericsson), leading the area of QoS for 2G and 3G cellular networks. Since 2005, he is a professor at the Department of Communications Engineering at the University of Málaga. His research interests include the field of mobile communications, especially QoS/QoE evaluation for multimedia services, radio resource management strategies and interference management for LTE and LTE-Advanced.

José T. Entrambasaguas received the M.S. and Ph.D. degrees, in 1975 and 1990, respectively, both from the Polytechnic University of Madrid, Spain. From 1975 to 1978, he worked at the Polytechnic University of Madrid. In 1978 he joined Fujitsu-España R&D center, where he worked in the development of packet switching systems, data transmission systems, and computer local area networks. In 1993, he joined the University of Málaga at the Communication Engineering Department, where he is currently a full professor. His current interests include digital signal processing techniques applied to digital communication and methodologies for efficient and integrated development of complex communications systems.

Performance Evaluation of cooperation-based techniques for M2M traffic over LTE

Isabel M. Delgado-Luque[(2)], *Francisco Blánquez-Casado*[(2)], *Francisco Javier Martín-Vega*[(2)], *Manuel Garcia Fuertes*[(1)], *Gerardo Gomez*[(2)], *Mari Carmen Aguayo-Torres*[(2)], *J. Tomás Entrambasaguas*[(2)], *Janie Baños*[(1)]

[(1)]AT4 wireless, S.A., Málaga, Spain
[(2)] University of Málaga, Málaga, Spain

***Abstract*—Long Term Evolution (LTE) and its evolved version, LTE-Advanced (LTE-A), are the standards developed by 3GPP to avoid the limitations associated to 3G networks. Efforts have been put on the improvement of system performance in terms of data rates. Coordinated Multipoint (CoMP) and inter-cell interference avoidance techniques have been widely studied in order to do that. However, not only is the increase in data rate important, but also latency reduction is a crucial factor to be considered for some applications. For instance, Machine-to-Machine (M2M) communications are nowadays one of the most promising applications, which are very sensitive to delay. In this work, we evaluate one of the most promising CoMP techniques, i.e. non-precoded Joint Transmission (non-precoded JT) and one of the most widely studied Inter-Cell Interference (ICI) avoidance technique, i.e. Partial Frequency Reuse (PFR). The objective is to determine whether, in addition to increase the data rates, they also reduce the latency. Simulation results show that PFR presents a greater reduction in the mean packet delay than this associated to non-precoded JT.**

***Index Terms*— Coordinated Multipoint, Joint Transmission, latency, LTE-A, Machine-to-Machine, Partial Frequency Reuse**

I. Introduction

Long Term Evolution (LTE) is a mobile communication system designed for data services that will replace the existing 3G system. Indeed, LTE is the standard developed by the 3GPP in order to solve the limitations associated to 3G networks. The standard is based on a new access technology, Orthogonal Frequency Division Multiplexing (OFDM), in order to support wideband transmission and use Multiple-Input Multiple-Output (MIMO) techniques that allow improving the spectral efficiency.

Nowadays, the 3GPP group is working in an evolved version of LTE called LTE-Advanced (LTE-A). This new version is expected to achieve higher data rates in uplink (UL) (0.5 Gbps instead of 50 Mbps) and downlink (DL) (1 Gbps instead of 100 Mbps), lower latency and scalable bandwidth up to 100 MHz (instead of 20 MHz). In order to do that, several new features have been envisioned for LTE-A, including carrier aggregation, advanced MIMO techniques, wireless relays, enhanced Inter-Cell Interference Coordination (eICIC) and coordinated multipoint (CoMP) transmission/reception [1].

One of the main reasons behind the use of the features mentioned above is to achieve LTE-A target data rates. In that respect, CoMP is one of the most important features to be taken into account in LTE-A. CoMP can be considered as a distributed MIMO system, in which geographically distributed nodes form multiple antennas and they cooperate to transmit and/or to receive from user equipments (UEs) [2-4]. The objective is to increase the system throughput, especially at cell edge areas where Inter-Cell Interference (ICI) is severe. Other group of techniques which have been investigated in order to improve system performance, especially at cell edge, is ICI avoidance techniques. These schemes aim at reducing the interference at the cell edge by applying a looser frequency reuse pattern to cell-edge users [5-6].

Beside higher data rates, latency reduction is an important factor to be considered in LTE-A since new types of latency-constrained applications are increasingly using mobile networks. One type of these latency-constrained applications is Machine-to-Machine (M2M). It is expected that the market for these applications will grow in a few years; according to some estimations indicating that 50 billion M2M devices will be active in year 2020 [7][8]. For this reason, 3GPP is working to improve its LTE standard for M2M applications in next LTE-A releases [9] and this kind of traffic is part of the Machine Type Communication (MTC) framework which describes the exchange of data between two machines. Efforts are being put on the provision of enhancements for RAN overload control MTC [10] and on the provision of low-cost MTC UEs based on LTE [11]. However, there is little or no work identified on the provision of enhancements for latency-constrained M2M applications. Thus, it is interesting to analyze the features of LTE-A in order to evaluate the latency performance when M2M applications are considered.

In this paper, we focus on two widely studied techniques used to increase the system throughput and evaluate their performance for M2M communications over the DL in a LTE-A system. The first technique considered is the non-precoded Join Transmission (non-precoded JT), where a cooperating evolved NodeB (eNB) and a Remote Radio Head (RRH) unit jointly transmit data to one or more corresponding UEs. The second technique studied is the Partial Frequency

978-1-4577-1348-4/13/$31.00 ©2013 IEEE

Reuse (PFR), where adjacent cells have to coordinate in order to divide the whole bandwidth into two reuse patterns, namely reuse 1 and reuse 3, and in order not to use the same reuse 3 frequencies simultaneously. The former technique, non-precoded JT, has been selected because it has been pointed as a major category of CoMP techniques to improve the overall system performance [12] while the later, PFR, seems to be one of the most promising techniques to reduce ICI and to increase the transmission rate [13-14].

The remainder of this paper is organized as follows. Section II describes the techniques considered. Section III presents the simulated scenario proposed to evaluate these techniques. Simulation results are presented in Section IV. Finally, some concluding remarks are given in Section V.

II. TECHNICAL DESCRIPTION

LTE and LTE-A are based on OFDM, which implies immunity to intra-cell interference because of the orthogonality between subcarriers. However, due to the frequency reuse factor is 1, users will experience ICI, especially in a fully-loaded OFDM cellular environment. This ICI will have a higher impact on users located at the cell boundary.

Assuming that the eNB uses the same transmit power for all users, a user close to the eNB will receive (in average) a higher desired signal power than a user located at the cell boundary. Also, within a cell, two regions can be differentiated: one region where the average Signal to Interference plus Noise Ratio (SINR) is dominated by the noise, so the reception is noise-limited; and other region where the interference is the dominant term, so the reception is interference-limited. The first region will be located close to the eNB (cell-center region), whereas the second one will be located at the boundary of the cell (cell-edge region). Thus, a user close to the eNB is expected to have better transmit conditions than a user at the cell boundary, which normally experiences lower throughput and higher delay. In order to improve the reception process of cell-edge users, two techniques are considered.

A. *Partial Frequency Reuse (PFR)*

PFR improves the reception process of cell-edge users by preventing them from interferences. Figure 1 shows the PFR scenario considered in this work: the total available system bandwidth B is split into two non-overlapped orthogonal frequency bands denoted by B_{center} and B_{edge}. B_{center} represents the available bandwidth for users near the eNB and B_{edge} is divided into three sets of equal-sized sub-bands that represent the available bandwidth for cell-edge users; each one of these sub-bands will be allocated to cells in a manner that no other surrounding cell is using the same sub-band. That way, each cell uses the spectrum B_{center} to serve cell-center users and a third of the B_{edge} spectrum to serve cell-edge users. This frequency allocation ensures that cell-edge users will not receive interference from the closest neighbor cells because these cells are assigned different frequency sub-bands for their transmission in that region. This reduction in the interference level is achieved at the expense of reducing the amount of allocable bandwidth per cell (only one third of B_{edge} is used in each cell).

B. *Non-precoded Joint Transmission (non-precoded JT)*

Unlike the above technique, JT tries to exploit the fact that a user receives signals from adjacent cells. Indeed, this technique consists of transmitting the same information simultaneously from different eNBs to cell-edge users in order to improve the reception process. This information has to be coherently combined at the UE. Figure 2 depicts the process done when this technique is applied. As it can be seen, the same frequency resources of the adjacent cell are dedicated to the transmission of the same information to cell-edge users. This means that the capacity of these frequency resources is divided by the number of entities involved in the JT process. Thus, it is important to decide which users are considered to be at the cell-edge region in order to avoid the use of this technique by users whose reception process is not interference-limited. This decision can be taken according to the average power received from the serving cell and from the other cells. If both values are similar, it means that the user is at the cell boundary. Otherwise, the user is close to the eNB and interferences do not significantly affect the reception process.

The efficiency of this technique depends on whether signals coming from different cells are added in phase or counterphase (as no precoding is assumed). In the worst case,

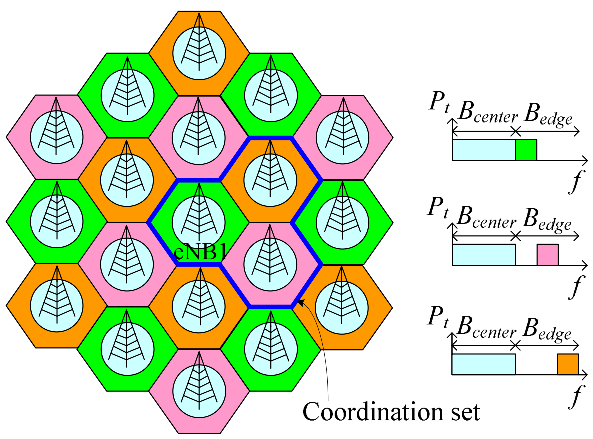


Fig. 1. Cellular topology with the PFR scheme considered

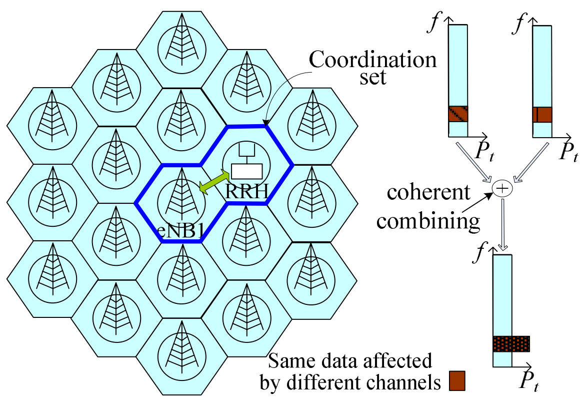


Fig. 2. Joint processing at transmission scheme considered

addition might be destructive and the received signal will be null. Anyway, the reception will be improved as the number of transmitting cells is increased.

It should be noted that neighboring cells shall be coordinated in both techniques but the grade of coordination is very different in each case. On one hand, in the PFR technique adjacent cells have to be coordinated in order to define the non-overlapped frequency bands, B_{center} and B_{edge}, and to decide which sub-band of B_{edge} eNBs can use. However, this exchange of information is typically performed in a large scale time (i.e. several days). On the other hand, a perfect coordination among adjacent cells must exist when non-precoded JT is applied since coordinated entities need to know not only the frequency resources assigned to each cell-edge users but also the complex data to be transmitted. In both cases, we call *coordinated set* the group of cells that share information.

III. SIMULATION ENVIRONMENT

The downlink of a LTE cellular network with 61 hexagonal cells is considered in the simulations. It is assumed that each cell has a radius of 0.5 km and that an eNB is located at the center of each cell. All eNBs use the same transmit power. A Single-Input Single-Output (SISO) antenna configuration has been used. A summary of the simulation parameters is shown in Table I.

In order to reduce the simulation time, we have only simulated one *coordinated set* while the effect of the interference signal coming from the rest of the cells has been modeled. It should be noted that when PFR is analyzed, this coordinated set is composed of three cells, as can be seen from Figure 1: the observed cell (named as eNB1) and two other cells, each with an eNB. Nevertheless, only two cells form the coordinated cell when non-precoded JT is studied (see Figure 2). Here again, the cell labeled as eNB1 is the observed cell whereas the second cell, named as *aiding cell*, includes an RRH unit located at the cell center. Thus, two different sources of interferences are implemented: the interference from the simulated adjacent cells, which is obtained instantaneously and depends on the amount of transmitted data, and the sum of interferences from the other cells, which is modeled as a Gaussian noise (as the number of interfering cells is high enough). In both cases it is assumed that the number of users in each interfering cell is high enough to generate interference over the whole bandwidth all the time (i.e. full load condition). In order to ease the numerical evaluation, we have considered that the signal strength of the Gaussian interference is only affected by path loss, which is modeled using the Hata Model [13]; while the signal received by users from the simulated coordinated set not only suffers from path loss but also from small scale fading, which is Rayleigh-distributed.

TABLE I
PARAMETER SETTINGS

LTE Parameter	Value/Mode
Number of cells	61
Cell radius	500 m
Number of antennas	Single antenna at eNBs and UEs
Transmit power of eNBs	49 dBm
Path loss model	Hata model [13]
Multipath model	Rayleigh fading
Mobile Speed	4km/h (pedestrian)
Channel bandwidth	20 MHz
OFDM symbols per TTI	14
PRB size	12 subcarriers
Carrier frequency	2.5 GHz
Modulation schemes	QPSK, 16QAM and 64QAM
Target BLER	10%
Number of users	10
Simulation length	20s
Type of traffic	M2M

Specifically, 10 users randomly distributed have been simulated in the observed cell. Each of them receives traffic from a video surveillance IP camera source model [14], which corresponds to a delay-sensitive remote control M2M application. Once users are located in the observed cell, it is necessary to define the cell-edge and cell-center area: a user is considered to be at the cell-edge area whenever the difference between the received power from the observed cell and from the adjacent cells is lower than 5 dB. This value has been chosen in order to ensure that cooperation is only applied to users that have very poor channel conditions.

In PFR technique, the size of B_{edge} is determined proportionally to the number of cell-edge users. In JT technique, this process is not necessary since no bandwidth division is done.

IV. SIMULATION RESULTS

Results presented in this section are associated to three cases of study: 1) no technique applied, i.e. the cellular network described in Section III with a reuse 1 factor (named as Baseline), 2) non-precoded JT technique and 3) PFR technique.

Figure 3 depicts the mean SINR associated to cell-edge users (red lines) and cell-center users (blue lines) in the system for the three cases of study. SINR levels are presented for different transmission power increment values at the eNB. The parameter transmission power increment is defined as

$$\Delta P_{Tx} = P_{Tx} - P_n - Lp_{\max} \ [dB] \quad (1)$$

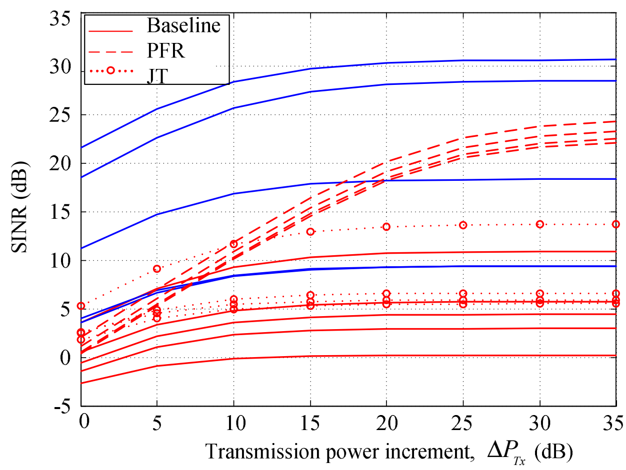


Fig. 3. Mean SINR per UE (cell-edge users in red color)

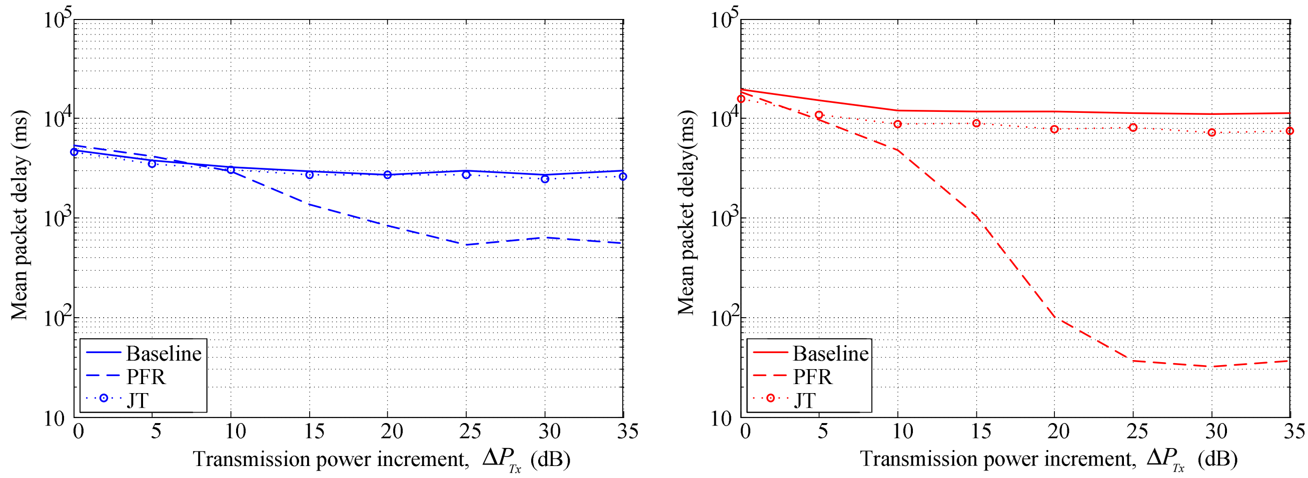


Figure 4. a) Mean packet delay in the observed cell b) Mean packet delay of cell-edge users

where $Lp_{\max}$ corresponds with path loss at the cell boundary, P_n is the noise power and P_{Tx} is the transmission power. It can be seen that for low values of transmission power, SINR is dominated by the noise power. As ΔP_{Tx} gets higher (above 20 dB), SINR is getting dominated by interference. Finally, mean SINR values saturate because the effect of increasing the desired signal power compensates the interference signal power increase.

It should be noted that only the mean SINR associated to cell-edge users is changed when any of the techniques is applied. Indeed, the increase of mean SINR of cell-edge users in non-precoded JT is due to the fact that the signal power of the worst users is increased whereas in PFR it is so because no interferences are received from cells of reuse distance of 3. As it can be seen, the increment in SINR achieved with PFR is higher than that associated to non-precoded JT. This is because in non-precoded JT the increment in SINR depends on both the distance to the cell boundary and the distance to the adjacent cells while in PFR the increment is achieved since the amount of interference is divided by a factor of 3.

Figures 4a) and 4b) show results for mean packet delay in the observed cell and mean packet delay associated to cell-edge users, respectively, in the three cases of study considered. On one hand, it can be seen that the observed cell mean packet delay is essentially the same for both the Baseline configuration and the non-precoded JT technique, while mean packet delay of cell-edge users is slightly reduced. On the other hand, PFR improves delay results for most values of ΔP_{Tx}. Only for the lowest values of the transmission power increment, results are quite similar because at this point the SINR of users at the cell-edge is quite similar. As the improvement in packet delay results is due to the increase in the SINR, it was expected that the delay reduction achieved by PFR technique will be more significant than that associated to non-precoded JT.

To end this section, Figure 5a) and 5b) present a comparison of average user throughput (i.e. mean throughput per user in the observed cell considering the whole cell) and average cell-edge users' throughput (i.e. mean throughput per user considering only cell-edge region in the observed cell), respectively. It is worth noticing that the average user throughput is slightly worse for the non-precoded JT technique than for the Baseline configuration. This is so because the adjacent cell has fewer resource elements to share because of the use of the JT. Regarding throughput results, it is only advisable the simultaneous transmission from more than one cell when the user throughput increases more than double. This is actually a more demanding decision criterion than a power threshold of 5 dB. Nevertheless, the interest of this technique comes from the performance improvement of cell-edge users. As it can be seen from Figure 5b), the average cell-edge users' throughput is increased significantly. In particular, the average cell-edge users' throughput gain at 20 dB is 64% whereas the relative loss in average user throughput is 5%. Furthermore, average user throughput results for the PFR technique outperforms Baseline and JT results for medium and high values of power transmissions increment. Moreover, it can be seen that the improvement in cell-edge users' throughput achieved with PRF is the highest of the three scenarios considered. It should be noted that the use of PFR has a trade-off between fairness among users and system performance, in terms of throughput. This is because an amount of resources are dedicated to cell-edge users who exploit them in a less efficient way than cell-center users. Nevertheless, in our case the use of PFR not only improves average cell-edge users' throughput but also average user throughput. This is due to the type of traffic considered. Indeed, the M2M traffic model has a relative low traffic rate so that cell-center users have enough resources to achieve the source rate using only the B_{center} bandwidth.

V. CONCLUSIONS

This paper presents a performance comparison between two of the most studied techniques used to increase the system

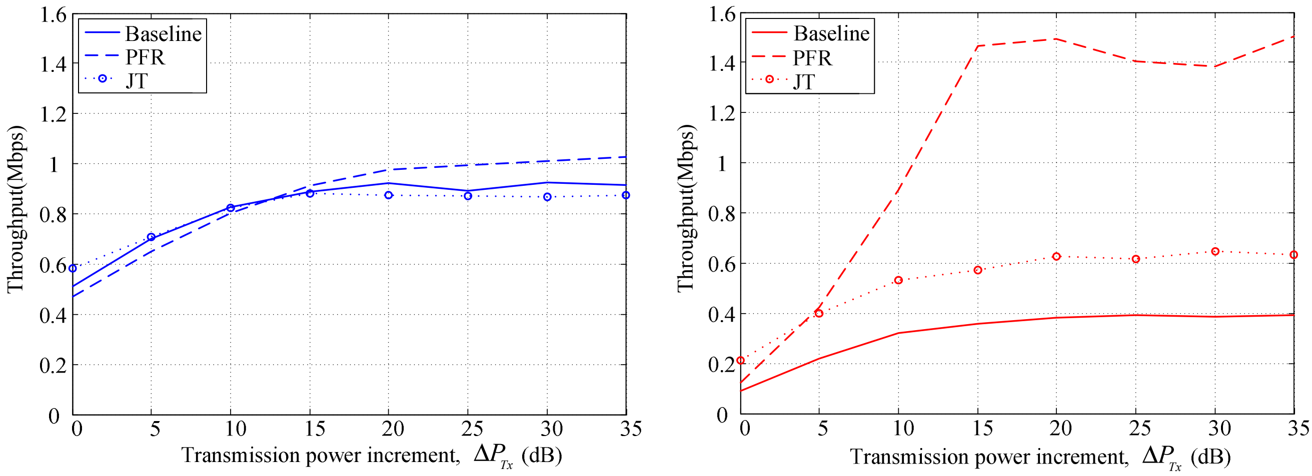


Figure 5. a) Average user throughput b) Average throughput of cell-edge users

throughput, especially at cell edge areas: non-precoded JT and PRF.

Simulations results show that both techniques improve the average cell-edge users' throughput compared to the Baseline configuration. Furthermore, PFR technique presents better performance in terms of delay and average user throughput for the traffic type considered. Indeed, PFR significantly improves delay results for both the observed cell and users located at the cell-edge for an important range of transmission power increments. Taking the above results into account, it can be concluded that the use of non-precoded JT is not adequate when M2M traffic is considered.

ACKNOWLEDGMENTS

The research leading to these results has received funding from the European Community's Seventh Framework Programme (FP7) under the LOLA project (Achieving Low-Latency in Wireless Communications) grant agreement Nº 248993. This work has also been performed in the framework of the project TEC2010-18451 funded by the Spanish Government. Finally, this work was partially supported by the University of Malaga – Campus de Excelencia Internacional Andalucía Tech.

REFERENCES

[1] 3GPP TR 36.913 V11.0.0 (2012), "Requirements for further advancements for Evolved Universal Terrestrial Radio Access (E-UTRA) (LTE-Advanced)"

[2] M. Sawahashi, Y. Kishiyama, A. Morimoto, D. Nishkawa, M. Tanno, "Coordinated multipoint transmission/reception techniques for LTE-Advanced", IEEE Wireless Commun. 17(3), 26-34 (2010)

[3] J. Zhang, R. Chen, J.G. Andrews, A. Ghosh, R.W. Heath, "Networked MIMO with clustered linear precoding", IEEE Trans. Wireless Commun. 8(4), 1910–1921 (2009)

[4] H. Dahrouj, W. Yu, "Coordinated beamforming for the multicell multiantenna wireless system", IEEE Trans Wireless Commun. 9(5), 1748–1759 (2010)

[5] Abdel-Baset Hamza et Al. "Inter-Cell Interference Coordination", 4G++ Project: Advanced Performance Boosting Techniques in 4th Generation Wireless Systems, Deliverable D4.1, 2011

[6] Haipeng Lei, Lei Zhang, Xin Zhang, Dacheng Yang, "A Novel Multi-Cell OFDMA System Structure using Fractional Frequency Reuse", IEEE 18th International Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC), pp. 1 – 5, 2007

[7] A. Orrevad, "M2M Traffic Characteristics (When machines participate in communication)", in Information and Communication Technology 2009, KTH Sweden: Stockholm. p. 56.

[8] N. Nikaein, S. Krco, "Latency for Real-Time Machine-to-Machine Communication in LTE-Based System Architecture", European Wireless 2011, Austria.

[9] 3GPP TR 36.912 V11.0.0 (2012), "Feasibility study for Further Advancements for E-UTRA (LTE-Advanced)"

[10] 3GPP TSG-RAN meeting #53 RP-111373 "3GPP™ Work Description: RAN overload control for Machine-Type Communications" 13 – 16 September, 2011

[11] 3GPP TSG-RAN meeting #53 RP-111112 "3GPP™ Work Description: Provision of low-cost MTC UEs based on LTE" 13 – 16 September, 2011

[12] 3GPP TR 36.814 V9.0.0 (2010-03), "Further advancements for E-UTRA physical layer aspects (Release 9)"

[13] P. Godlewski, M. Maqbool, M. Coupechoux, "Analytical evaluation of various frequency reuse schemes in cellular OFDMA networks," in perso.telecom-paristech.fr, 2008.

[14] A. Alsawah, I. Fialkow, "Optimal frequency-reuse partitioning for ubiquitous coverage in cellular systems," in 16th European Signal Processing Conference - publi-etis.ensea.fr, 2008.

[15] European Cooperative in the Field of Science and Technical Research EURO-COST231, "Urban transmission loss models for mobile radio in the 900- and 1800 MHz bands (Revision 2)," COST 231 TD(90)119 Rev. 2, The Hague, The Netherlands, September 1991, available at http://www.lx.it.pt/cost231/final_report.htm

[16] EU FP7 Project LOLA (Achieving Low-Latency in Wireless Communications), Project nº 248993, D3.2 Network related analyses of M2M and online-gaming traffic in HSPA, v1.0, June 2010.

EVALUATION OF LATENCY-AWARE SCHEDULING TECHNIQUES FOR M2M TRAFFIC OVER LTE

I.M. Delgado-Luque, F. Blánquez-Casado, M. Garcia Fuertes, G. Gomez, M.C. Aguayo-Torres, J.T. Entrambasaguas, J. Baños

AT4 wireless, S.A., Málaga, Spain
University of Málaga, Málaga, Spain

ABSTRACT

Machine to Machine (M2M) communications are expected to grow dramatically in next years. Scheduling techniques are determinant to achieve high spectral efficiency in wireless systems and to provide QoS guarantees to system users. In this work, several scheduling algorithms are evaluated in order to accommodate delay limited M2M communications over an LTE system. Simulation results show a reduction of mean and 95th percentile packet delay.

Index Terms— Scheduling, LTE, Machine to Machine, delay

1. INTRODUCTION

Some estimations foresee the world could have a trillion communicating devices (including sensor and RFID networks) in a decade [1]. It is expected that most of them will be wireless and a high percentage will not be operated directly by humans. This Machine to Machine (M2M) communications have certain characteristics which makes them hard to accommodate in mobile networks. In particular, certain flows are very sensitive to delay.

There are M2M applications, such as vehicle collision detection and avoidance, sensor-based alarms and remote control, etc., that require extremely low latency values. In remote video surveillance type applications, for instance, Unmanned Ground Vehicles (UGV) and Unmanned Aerial Vehicles (UAV), devices carrying video cameras (usually robots) are remotely controlled by a human, or by an intelligent control system, based on the video information captured by the camera and transferred to the control point. In terms of latency the video flow is more critical than the control signal. Assuming the use of optimized video codecs for real-time video transmission, the latency introduced by the network, and in particular by the Medium Access Control (MAC) and physical (PHY) layers [2], becomes important in the latency budget.

3GPP is working to improve its Long Term Evolution (LTE) standard performance for M2M applications in next LTE-Advanced releases [3]. Efforts are being put on the provision of enhancements for RAN overload control for Machine-Type Communications (MTC) [4] and on the provision of low-cost MTC User Equipment based on LTE [5]. However, there is little or no work identified on the provision of enhancements for latency-constrained M2M applications. The use of latency-aware and optimized MAC schedulers is therefore crucial so that these latency-constrained applications are allocated with the necessary physical resources to ensure proper operation.

In a multiuser system, an optimized MAC layer should allocate radio resources to users according to several parameters, including traffic source characteristics, Quality of Service (QoS) needs [6] and the frequency, time and space diversity of the mobile radio channel.

When traffic sources are variable-rate and their transmission resources requirements fluctuate asynchronously for different users, exploiting the source multiuser diversity (statistical gain) allows more users to be accommodated on the system. A MAC design adaptable to the changing traffic and channel characteristics and to the specific QoS requirements, i.e. a cross-layer design, improves the system performance by exploiting the radio resources more efficiently [7].

Several scheduling algorithms over LTE have been proposed in the literature [8]. Literature is also extensive on heuristic approaches which take into account the source behaviour (see [9] and [10]). The chosen techniques should be sufficiently flexible to accommodate traditional traffic sources as well and the existence of other sources with different QoS requirements [6]. In LTE, the eNodeB is responsible for implementing the scheduling policies both in the downlink and uplink scenarios and the UEs are informed about the resource allocation decisions on a subframe basis, through control channels.

In this paper, we focus on the evaluation of resource allocation algorithms for M2M communications over LTE. Work is partially result of EU FP7 Project LOLA

(Achieving Low-Latency in Wireless Communications), whose traffic analysis for M2M has been presented in [11]. Advanced scheduling algorithms have been chosen considering different aspects like the instantaneous channel conditions, latency requirements or pending retransmissions in order to adequately allocate transmission turns to users.

The remainder of this paper is organized as follows. Section II briefly describes the proposed resource allocation algorithms. In Section III the link-based simulator used to evaluate these algorithms is introduced. Scheduler performance is then presented in Section IV. Finally, some concluding remarks are given in Section V.

2. DESCRIPTION OF THE PROPOSED SCHEDULERS

Three scheduling algorithms chosen from the literature have been analyzed for delay-dependent traffic coming from a M2M traffic source: Opportunistic hard priority [12][13], Channel Dependent Earliest Deadline Due (CD-EDD) [14] and CD-EDD with postponed EDD term [15].

2.1. Opportunistic Hard Priority

This algorithm applies opportunistic priority to the transmissions of delay-sensitive flows if a maximum packet delay threshold is exceeded.

The algorithm sets the same priority to all packets as long as packet delay is below the threshold and sets high priority to packets exceeding the threshold until they are served. A delay threshold D_t and a delay budget D_b is assigned to each delay-sensitive flow. LTE defines a delay budget for the radio interface of 80 ms for VoIP traffic [16]. The delay threshold must be chosen to be lower than the delay budget with a sufficient margin so that the scheduler can serve packets exceeding this threshold before violating the delay budget. Packets exceeding the delay budget are discarded.

Initially, users are cyclically ordered based on arrival time to their transmission queues. The waiting time of the head of line (HOL) packet in each queue is continuously monitored. If the HOL packet delay of a user exceeds the delay threshold the Opportunistic Hard Priority scheduler will give priority to the transmission of such packet until it gets served (strict prioritization).In contrast, Proportional Fair (PF) policy can be seen as a kind of soft prioritization scheduling (assuming similar source traffic characteristics): users with worse average channel conditions are expected to suffer higher delay; to compensate it, PF policy indirectly prioritizes them through the inverse of the average throughput applied to the instantaneous potential rate.

The algorithm implemented by the Opportunistic Hard Priority scheduler is as follows:

1. Set the delay parameters for each data flow: delay budget and delay threshold
2. Set the priorities for each data flow:
   a) If (D^b – HOL packet delay) < 0 → the packet is discarded.
   b) else if (D^b – HOL packet delay) < D^t → priority = 1
   c) else priority = 0
3. Sort the list of data flows from highest to lowest priority.
   a) Allocate a set of free Physical Resource Blocks (PRB) to the flow with the highest priority in the list if, and only if, the following conditions are met:
      - There are non-assigned PRBs.
      - There is data waiting to be served.
      - There is at least one free HARQ channel.
      - The data flow has not received the maximum allowable number of assignments in the current Transmission Time Interval (TTI).

      If there is more than one data flow with the same priority, a Round Robin (RR) scheme is applied.
   b) Remove the served data flow from the list.
   c) If it is possible to continue, go to step 3.a).

2.2. Channel Dependent Earliest Deadline Due (CD-EDD)

The priority assigned by this scheme depends on two components: the delay-aware component (EDD) and the channel-aware component (PF). The EDD term works in such a way that users are prioritized as their HOL packet delay gets close to the delay budget. The PF term favours terminals with temporarily good channel conditions. The scheduling tag assigned to user k is therefore calculated according to the following formula:

$$\frac{T_k[n]}{R_k[n]} \cdot \frac{W_k[n]}{D_k^b - W_k[n]} \tag{1}$$

where:

- $T_k[n]$ is the throughput of user k at TTI n
- $R_k[n]$ is the average throughput of user k up to TTI n
- $W_k[n]$ is the waiting time of the HOL-packet in the queue of user k (expressed in TTIs)
- D_k^b is the maximum allowable delay (or delay budget) of user k (in TTIs)

The PF and EDD terms are well differentiated in the formula. The term associated to the PF algorithm is the first quotient ($\frac{T_k[n]}{R_k[n]}$) and the EDD term is the second quotient ($\frac{W_k[n]}{D_k^b - W_k[n]}$). As the delay of the HOL packet gets close to D_k^b, the EDD term quickly dominates the scheduling tag.

For low HOL packets delays, the PF term dominates the scheduling tag. As for the Opportunistic Hard Priority algorithm, packets exceeding the delay budget are discarded.

2.3. CD-EDD with postponed EDD term

This scheme is a modification of the previous algorithm, affecting the delay-aware term. The PF term will now dominate the scheduling tag as long as HOL packet delays are far from exceeding the delay budget. Such approach is based on a simple utility function:

$$\frac{T_k[n]}{R_k[n]} \cdot \left(\frac{\max(0, W_k[n] - D_k^t)}{D_k^b - W_k[n]} + 1 \right) \tag{2}$$

where D_k^t is the minimum delay (or delay threshold) associated to user k. The delay-aware term will provide priority to users whose HOL packets are greater than this threshold (D_k^t). This approach takes advantage of the flow delay tolerance in order to increase the system capacity. Again, packets violating the delay budget are discarded.

The objective of these three algorithms is to ensure that the instantaneous packet delay is kept below a certain value. If a packet waiting in the queue has exceeded the delay budget, the system discards such packet.

3. LINK-BASED SIMULATION ENVIROMENT

A proprietary simulation environment oriented to model and simulate complex MIMO-OFDM wireless systems has been used. The simulation environment is composed of a number of UEs connected to a Base Station (BS) through a frequency-selective Rayleigh fading channel. The high-level architecture for the direct link is depicted in Figure 1. The simulation environment has been adapted to implement and evaluate the proposed algorithms over an LTE-like system.

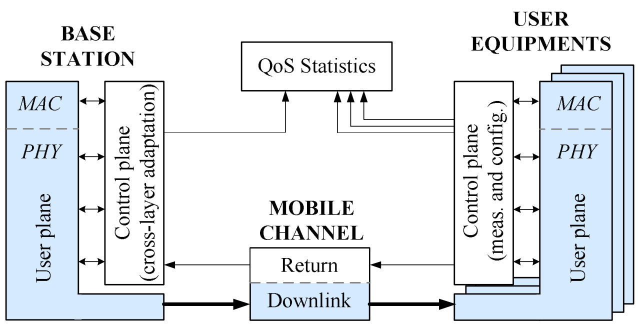


Figure 1. Simplified DL link-based simulation environment architecture

Baseline simulations have been done in order to evaluate the latency reduction achieved with different scheduling techniques [17]. A PF algorithm has been chosen as the baseline scheduling algorithm to compare latency results. A summary of the simulation parameters is shown in Table 1.

Table 1. Parameter settings

LTE Parameter	Value/Mode
Channel model	Extended pedestrian A (TS 36.803)
Mobile Speed	4 km/h (pedestrian)
Channel Bandwidth	20 MHz
OFDM symbols per TTI	14
PRB size	12 subcarriers
Carrier Frequency	2.5 GHz
Modulation schemes	QPSK, 16QAM and 64QAM
Target BLER	10%
ACK feedback delay	8 ms
CQI delay	2 ms
Nº of CQI bits	4
Max. Number of HARQ retransmissions	8
Number of parallel HARQ processes	8
MIMO mode	2x2 1 layer spatial multiplexing (Beamforming)
Codebook	TS 36.211
Channel Estimation	Non-ideal Zhao
MIMO detection	ZF
Noise power estimation	Error based
Signalling overhead	2/21
Number of users	10
Simulation length	20s
Type of traffic	M2M

The M2M traffic source used in this work is an IP video surveillance camera [11] transmitting a video flow for a delay-sensitive remote control M2M application. This traffic application can be mapped to the standardized QoS Class Identifier (QCI) number 7 defined in 3GPP specifications [16]. For this QCI, the maximum packet delay allowed is 100 ms. The following delay parameters have been associated for the algorithms to be analyzed:

Table 2. Parameters of the proposed schedulers

Opportunistic Hard Priority		CD-EDD		CD-EDD with postponed EDD term	
D^b	100 ms	D^b	100ms	D^b	100 ms
D^t	50 ms			D^t	50 ms

4. RESULTS

Results of mean and 95th percentile packet delay, packet loss rate and throughput per user are shown in Figure 2, Figure 3, Figure 4 and Figure 5, respectively, for the proposed scheduling algorithms. These results are discussed in following subsections.

When the HOL packet delay exceeds the delay budget, the three algorithms studied in this paper discard such packet instead of incrementing the delay of the remaining packets to be served (which could be more damaging for the QoS).

Discarding packets is more likely to happen when the mean SNR is low (0-5 dB). In these conditions, the throughput per user is very low (see Figure 5) due to the high outage probability and the need of using robust coding schemes that ensure the target BLER. As a consequence, the probability of a delay budget violation increases and thus, a high number of packets have to be discarded so that delay results are kept below the delay budget value (100 ms), which implies a reduction near to 90% compared to the baseline results. However, for these simulation conditions, the discarding packet process does not imply a reduction of the average throughput compared with the baseline results (see Figure 5). This is because the achievable throughput in such conditions is lower than the load of the traffic source, so there will always be packets queued.

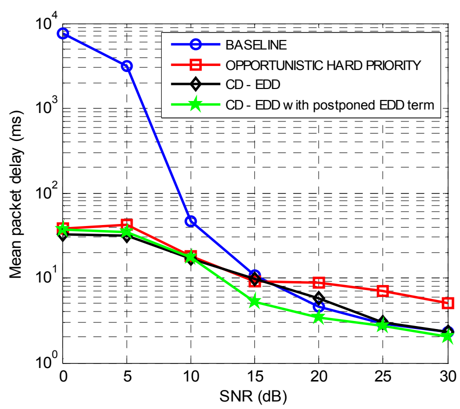


Figure 2. Mean packet delay results

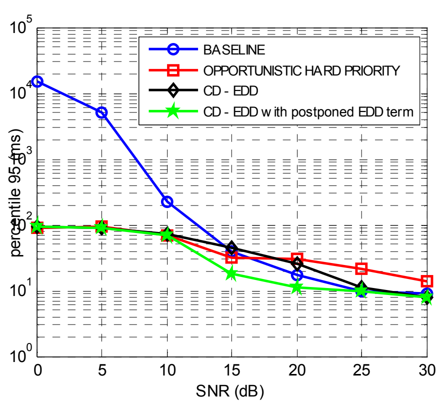


Figure 3. Percentile 95th of mean packet delay

When the mean SNR is increased, the packet loss rate decreases reaching 0 at a level of 15 dB (see Figure 4). Then, packet delay results are mostly influenced by the utility function of each algorithm. Also, for the three proposed algorithms, the average throughput is a bit lower than for the baseline (see Figure 5), as these algorithms favour users experiencing high packet delays. These users have in general worse channel conditions, so their throughput is consequently lower.

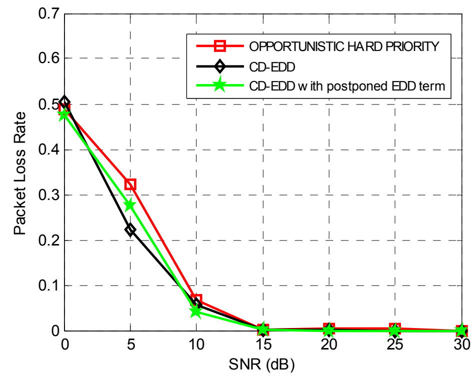


Figure 4. Packet loss rate

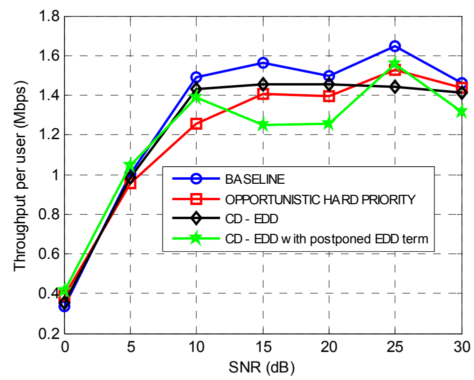


Figure 5. Throughput per user

4.1. Opportunistic Hard Priority

When mean SNR is increased, the performance of the opportunistic hard priority is similar to a RR algorithm: there are a lower number of packets whose waiting time exceeds the delay threshold; therefore, the algorithm will set the same priority to almost all data flows, which will be allocated in a cyclic order. This is the reason why the baseline configuration (based on a PF scheme) achieves a better performance for high SNR values.

4.2. Channel Dependent Earliest Deadline Due (CD-EDD)

For high SNR levels, mean packet delay associated to the CD-EDD technique gets slightly higher values than those associated to the baseline configuration. This is because HOL packet delays are low in such scenario and the EDD-term gives a rather low priority, thus degrading the performance compared to the case in which the PF-term dominates the scheduling decision. However, the CD-EDD improves the opportunistic hard priority algorithm.

4.3. CD-EDD with postponed EDD term

Packet delay results of the CD-EDD with postponed EDD term are also shown in Figure 2 and Figure 3. This algorithm improves packet delay results for the whole range of SNR values, in contrast to the results obtained with previous algorithms. As the value of the SNR is increased, HOL packet delays decrease, so the probability of exceeding the D^t value is also decreased. Thus, the CD-EDD algorithm works as a PF algorithm. When the HOL packet delay of a user is above the D^t the EDD term gives higher priority to that user, which decreases the average packet delay results (especially for medium SNR values). Therefore, SNR values do not lead to lower priorities, as it occurs for the CD-EDD.

Regarding the configuration values of D^b and D^t parameters, a reduction of these values would improve delay results as higher priority would be given to users experiencing high packet delays (generally due to worse channel conditions) at the expense of a system throughput reduction and an increase of the packet loss rate.

5. CONCLUSSIONS

This paper presents a performance comparison between three delay-aware scheduling algorithms for M2M traffic over an LTE system.

Simulation results show that, for low SNR values (between 0 and 15 dB), the three delay-aware algorithms are able to reduce considerably the mean and 95th percentile packet delay at the expense of discarding those packets exceeding the allowable delay. For high SNR values (from 15 to 30 dB) packet delay results are mostly influenced by the utility function of each algorithm. In that sense, the algorithm called CD-EDD with postponed EDD term achieves the best performance in terms of delay, as the influence of the HOL packet delay is not always affecting the scheduling decision.

6. ACKNOWLEDGEMENTS

The research leading to these results has received funding from the European Community's Seventh Framework Programme (FP7) under the LOLA project (Achieving Low-Latency in Wireless Communications) grant agreement Nº 248993. This work has also been performed in the framework of the Junta de Andalucía (Proyecto de Excelencia P07-TIC-03226) and the Spanish Government under the project TEC2010-18451.

7. REFERENCES

[1] IDC Estimates, "The Expanding Digital Universe", available at http://www.emc.com/collateral/analyst-reports/expanding-digital-idc-white-paper.pdf

[2] EU FP7 Project LOLA (Achieving Low-Latency in Wireless Communications), Project nº 248993, D2.1 Target application scenarios, v1.0, May 2010.

[3] 3GPP TR 36.912 V10.0.0 (2011-03), "3rd Generation Partnership Project; Technical Specification Group Radio Access Network; Feasibility study for Further Advancements for E-UTRA (LTE-Advanced) (Release 10)".

[4] 3GPP TSG-RAN Meeting #53, RP-111373. "3GPP™ Work Item Description: RAN overload control for Machine-Type Communications". Fukuoka (Japan), 13 – 16 September, 2011.

[5] 3GPP TSG-RAN Meeting #53, RP-111112. "3GPP™ Work Item Description: Provision of low-cost MTC UEs based on LTE". Fukuoka (Japan), 13 – 16 September, 2011.

[6] 3GPP TS 23.107 V8.2.0 (2011-12), "3rd Generation Partnership Project; Technical Specification Group Services and System Aspects; Quality of Service (QoS) concept and architecture (Release 8)".

[7] J.T. Entrambasaguas, M.C. Aguayo-Torres, G. Gómez and J.F. Paris, "Multiuser capacity and fairness evaluation of channel/QoS-aware multiplexing algorithms," IEEE Network, pp. 24-30, May-June 2007.

[8] H. Luo, S. Ci, D. Wu, J. Wu, H. Tang, "Quality-driven cross-layer optimized video delivery over LTE". IEEE Communications Magazine, vol. 48, no. 2, pp. 102-109, February, 2010.

[9] B. Sadiq, R. Madan, A. Sampath, "Downlink Scheduling for Multiclass Traffic in LTE". EURASIP Journal on Wireless Communications and Networking, vol. 2009.

[10] Y. Qian, C. Ren, S. Tang, M. Chen, "Multi-service QoS guaranteed based downlink cross-layer resource block allocation algorithm in LTE systems". International Conference on Wireless Communications & Signal Processing, pp. 1-4, November, 2009.

[11] EU FP7 Project LOLA (Achieving Low-Latency in Wireless Communications), Project nº 248993, D3.2 Network related analysis of M2M and online-gaming traffic in HSPA, v1.0, June 2010.

[12] S. Choi, K. Jun, Y. Shin, S. Kang and B. Choi, "MAC Scheduling Scheme for VoIP Traffic Service in 3G LTE". Proceedings of IEEE Vehicular Technology Conference (VTC-Fall). Baltimore (USA), October 2007.

[13] M. Andreozzi, G. Stea, A. Cacioccola and R. Rossi, "Flexible scheduling for real-time services in High-Speed Packet Access cellular networks". European Wireless 2009, Aalborg (Denmark), May 2009.

[14] A. K. F. Khattab and K. M. F. Elsayed, "Channel-quality Dependent Earliest Deadline Due Fair Scheduling Schemes for Wireless Multimedia Networks". Proceedings of MSWiM 2004, Venice (Italy), October 2004.

[15] G. Barriac and J. Holtzman, "Introducing Delay Sensitivity into Proportional Fair Algorithm for CDMA Downlink Scheduling". Proceedings of ISSTA 2002, Vol. 3, pp. 652-656, Parsippany (USA), September 2002.

[16] 3GPP TS 23.203 V11.3.0 (2011-09), "3rd Generation Partnership Project; Technical Specification Group Services and System Aspects; Policy and charging control architecture (Release 11)".

[17] EU FP7 Project LOLA (Achieving Low-Latency in Wireless Communications), Project nº 248993, D4.5 Scheduling Policies for M2M and Gaming Traffic, v2.0, January 2012.