

QoS and Energy Efficient Resource Allocation in Downlink OFDMA systems

Dionysia Triantafyllopoulou, and Klaus Moessner
 Institute for Communication Systems (ICS)
 University of Surrey
 Guildford, Surrey, GU2 7XH, United Kingdom
 email: {d.triantafyllopoulou, k.moessner}@surrey.ac.uk

Abstract—In this paper we present and evaluate the performance of a resource allocation algorithm to enhance the Quality of Service (QoS) provision and energy efficiency of downlink Orthogonal Frequency Division Multiple Access (OFDMA) systems. The proposed algorithm performs resource allocation using information on the downlink packet delay, the average delay and data rate of past allocations, as well as the downlink users' buffer status in order to minimize packet segmentation. Based on simulation results, the proposed algorithm achieves significant performance improvement in terms of packet timeout rate, goodput, fairness, and average delay. Moreover, the effect of poor QoS provision on energy efficiency is demonstrated through the evaluation of the performance in terms of energy consumption per successfully received bit.

Index Terms— Energy efficiency, Orthogonal Frequency Division Multiple Access (OFDMA), Quality of Service (QoS), resource allocation.

I. INTRODUCTION

One of the major challenges in future mobile communication networks is the need for increased capacity, Quality of Service (QoS) provision and energy efficiency. Among the most prominent approaches to achieve these goals is the design and employment of efficient resource allocation schemes. To this end, significant research progress has been made in the area of downlink resource allocation in Long Term Evolution (LTE), and Orthogonal Frequency Division Multiple Access (OFDMA) networks in general [1].

In [2]-[10], emphasis is given on resource allocation schemes that aim at enhanced QoS provision. Specifically, in [2], the authors propose and compare the performance of two scheduling algorithms, i.e., a maximum-rate scheduler, which prioritizes the users based on the highest supported bit rate according to their channel quality, and a proportional-fair (PF) scheduler, which improves fairness among users. Two of the most notable resource allocation rules, namely the exponential (EXP) rule and the log rule are described in [3] and [4], respectively. In [5], the performance of the well-known maximum-rate, round robin, PF, EXP/PF [6], and maximum-largest weighted delay first (M-LWDF) [7] resource allocation

algorithms is compared, with M-LWDF achieving higher system throughput and fairness.

In [8], a two-level scheduling algorithm is described. In the upper level, the amount of data of each source in order to satisfy its delay constraint is calculated. At the lower level the PF scheduler is used. A weighted round-robin resource allocation algorithm is proposed in [9] to improve system throughput, guarantee application layer QoS, in terms of video distortion, and ensure fairness, taking into consideration the channel quality, the packet delay constraints and the average data rate of each user. In [10], a resource allocation algorithm that operates in three phases is presented. In the first phase, the resource blocks are allocated to the User Equipment (UE) devices based on their Channel Quality Information (CQI). Then, the packet delays are predicted. Finally, the transmission order is rearranged and the packets that cannot meet their delay requirements are discarded.

Very useful conclusions regarding the trade-off between the energy efficiency (EE), i.e., the ratio of throughput over the total power consumption expressed in bits per Joule, and QoS in OFDMA networks are drawn in the recent bibliography. Specifically, the relation between energy efficiency and spectral efficiency, i.e., the ratio of throughput over the bandwidth expressed in bits per second per Hertz, in a single-cell OFDMA network is shown to be a quasi-concave function [11]. In [12], the authors propose a method for energy efficient resource allocation and an algorithm for trading bandwidth for energy efficiency during low load periods. A scheduling technique that, combined with the use of an energy efficient scheduler, allows users whose energy consumption is dominated by control channel overhead to reduce their overall energy expenditure by reducing the number of resource blocks allocated to them is presented in [13]. In [14], an optimization problem for the energy efficiency maximization subject to users' minimum transmission rate requirements and maximum transmission power constraints is described.

According to the related literature, downlink resource allocation is usually performed per resource block, i.e., in each subframe a resource block is allocated to the user who maximizes a specific metric. The overall performance of the system highly depends on the parameters that comprise this metric. However, this approach does not take into

consideration the effect of packet segmentation on the overall system performance. Specifically, in case the allocated resource blocks are not enough to transmit a user's packet as a whole, the need for packet segmentation introduces overhead, due to the introduction of separate protocol headers in each packet segment. Furthermore, in the case of real-time applications, where Automatic Repeat reQuest (ARQ) techniques are not employed, the loss of one packet segment results in the need to discard all the already received segments of the same packet and the waste of the respective resources used for their transmission. Moreover, as shown in Table I, the proposals that focus on QoS provision to real time applications do not consider energy efficiency [2]-[10], while the ones that mainly focus on energy efficiency do not explicitly support the strict constraints of realistic real-time applications in terms of packet delay [12]-[14].

Motivated by the above, in this paper we propose a QoS-oriented and energy efficient resource allocation algorithm for downlink LTE systems. Resource allocation is performed taking into consideration the packet delays in the downlink direction, the average delay and data rate of allocations in the past, as well as downlink buffer status in terms of packet segmentation. The main contributions of this paper with respect to the reviewed literature are summarized as follows:

- 1) Consideration of the effect of packet segmentation, performed at the Radio Link Control (RLC) layer, on the resource allocation. The proposed algorithm i) aims at allocating enough resources in each subframe to each user in order to transmit their packets as a whole and ii) prioritizes users whose pending packets are already segmented in order to transmit packet segments as soon as possible and avoid the discarding of already received packet segments due to expiration.
- 2) Demonstration of the effect of QoS on energy efficiency through the evaluation of the system performance in terms of the total energy consumption per successfully received bit.

This paper is organized as follows. Section II introduces the system model. Section III describes in detail the proposed downlink resource allocation algorithm, whose performance is evaluated through simulations in section IV. Finally, section V contains conclusions and discusses on plans for future work.

II. SYSTEM MODEL

The system model consists of a single LTE macro cell and a number of UE devices, randomly deployed in the macro cell coverage area. For the remainder of this document the terms user and UE are used interchangeably. Each user has an active real-time video connection on the downlink and the eNodeB is responsible to allocate the available resources in a fair, QoS and energy efficient manner, employing the proposed resource allocation algorithm. Table II summarizes the parameters used for the formulation and performance evaluation of the proposed algorithm.

In the time domain, downlink LTE transmissions are organized into radio frames, each of which consists of two half-frames. A half-frame consists of five equally sized subframes of length T_{sf} each. Each subframe consists of two equally sized slots. Each slot consists of N_{symp}^{DL} OFDM symbols, including cyclic prefix. The exact value of N_{symp}^{DL} depends on the cyclic prefix length, which is configured by the higher layers. The resource grid describing the downlink transmitted signals in each slot consists of $N_{RB}^{DL} \times N_{SC}^{RB}$ subcarriers and N_{symp}^{DL} OFDM symbols. The smallest physical resource in LTE is a resource element, consisting of one subcarrier during one OFDM symbol. Resource elements are grouped into resource blocks, where each resource block consists of N_{SC}^{RB} consecutive subcarriers in the frequency domain and one slot consisting of N_{symp}^{DL} OFDM symbols in the time domain [15]. A scheduling block consists of two consecutive resource blocks, spanning a subframe of length equal to T_{sf} , and is the minimum amount of resources that can be allocated to a user in a subframe.

III. THE PROPOSED ALGORITHM

As a first step, the set of active users U is sorted in descending order of $m_i^{DL}(t)$. This is a metric that aims to provide higher resource allocation priority to users with increased waiting time with respect to the delay threshold, high average delay and low average data rate of their allocations in the past, as well as segments of already transmitted packets. To this end, $m_i^{DL}(t)$ is defined as

$$m_i^{DL}(t) = \delta_i \frac{d_i^{DL}(t)}{d_{th,i}} \exp\left(\frac{\bar{d}_i^{DL}(t)}{\bar{r}_i^{DL}(t)}\right) \quad (1)$$

In order to avoid wasting of resources and transmission power spent for already transmitted packet segments, δ_i is a

TABLE I.
QoS AND ENERGY CONSUMPTION PARAMETERS CONSIDERED IN DOWNLINK RESOURCE ALLOCATION

Reference	Traffic model	QoS parameters	Energy efficiency parameters
PF-Multiuser [2]	Not explicitly specified	Average bit rate maximization, fairness	-
EXP-Rule [3]	Mutually independent ergodic Markov chains with countable state spaces	Throughput	-
Log Rule [4]	i.i.d. Bernoulli process, 1kb packet size	Average packet delay, 99 th percentile delay	-
EXP/PF, M-LWDF [5]	128 kb/s video streaming service	System throughput, packet loss rate, fairness	-
[8]	H.264 video, voice over Internet Protocol (VoIP), infinite buffer (best effort)	Packet loss rate minimization, average goodput maximization, fairness, peak signal to noise ratio (PSNR)	-
[9]	H.264 video	PSNR	-
[10]	H.264 video	Invalid packet rate, goodput, packet average delay, packet dropping rates	-
TCoM [12], [13]	Constant size buffer model	Data rate	RF energy consumption gain
QA-ERS [14]	Not explicitly specified, minimum rate requirement per user 128 kb/s	QoS satisfaction index	Energy efficiency

TABLE II.
DEFINITION OF SYSTEM MODEL PARAMETERS

Parameter	Definition
$m_i^{DL}(t)$	Downlink scheduling metric of user i
δ_i	Segmentation bias
$d_i^{DL}(t)$	Downlink queuing delay of user i (s)
$d_{th,i}$	Queuing delay threshold of user i (s)
$\bar{D}_i^{DL}(t)$	Average downlink delay of user i (s)
$\bar{R}_i^{DL}(t)$	Average downlink rate of user i (b/s)
β	Average delay and rate calculation factor
$r_i^{DL}(t)$	Instantaneous downlink rate of user i (b/s)
L_i	Length of HoL data packet of user i (b)
M_i	Modulation of user i (b/symbol)
C_i	Coding rate of user i
$NSB_{req_i}(M_i, C_i)$	Number of required scheduling blocks of user i with Modulation and Coding Scheme (MCS) (M_i, C_i)
L_{SB}^{DL}	Number of data carrying resource elements in a downlink scheduling block
$\Phi_{i,(M_i,C_i)}$	Set of available scheduling blocks for user i with MCS (M_i, C_i)
$L_{i,(M_i,C_i)}$	Number of bits that can be allocated to user i with MCS (M_i, C_i)
$\gamma_{i,j}$	SNR of user i on scheduling block j
$P_{i,j}$	Transmission power of user i on scheduling block j
α_i	Acceptable packet loss rate of user i
N_{RB}	Total number of resource blocks per slot
N_{syml}^{DL}	Number of OFDM symbols per downlink resource block
N_{SC}^{RB}	Number of subcarriers per resource block
$N_{RB,SB}$	Number of resource blocks per scheduling block
T_{sf}	Subframe duration (s)
U	Set of users
Φ	Set of available scheduling blocks
G_i	Set of allocated scheduling blocks to user i

parameter that gives resource allocation priority to users whose head-of-line (HoL) packet is segmented. The aim of this is to transmit segments of the same packet as closely as possible to each other in order to avoid their expiration and the discarding of the already successfully received segments. δ_i is formulated as follows:

$$\delta_i = \begin{cases} \delta_{bias}, & \text{HoL packet of user } i \text{ is segmented} \\ 1 - \delta_{bias}, & \text{else} \end{cases}, \quad (2)$$

where $0 \leq \delta_{bias} \leq 1$.

$d_i^{DL}(t)$ is the delay of the HoL packet of user i , while $d_{th,i}$ is the delay threshold, beyond which the packet is no longer considered usable and is discarded by the user's buffer.

$\bar{D}_i^{DL}(t)$ and $\bar{R}_i^{DL}(t)$ are the average delay and data rate, respectively, experienced by user i in the past, and are calculated using a weighted moving average formula:

$$\bar{D}_i^{DL}(t) = \beta d_i^{DL}(t) + (1 - \beta) \bar{D}_i^{DL}(t - 1) \quad \text{and} \quad (3)$$

$$\bar{R}_i^{DL}(t) = \beta r_i^{DL}(t) + (1 - \beta) \bar{R}_i^{DL}(t - 1), \quad (4)$$

where $r_i^{DL}(t)$ is the instantaneous downlink data rate of user i and $0 \leq \beta \leq 1$. The incorporation of $\bar{D}_i^{DL}(t)$ and $\bar{R}_i^{DL}(t)$ in $m_i^{DL}(t)$ allows the prioritization of users that were served with high average delay and low average data rate in the past, thus increasing the fairness of the proposed solution.

In order to avoid packet segmentation, in each subframe the proposed algorithm aims at allocating to each user enough

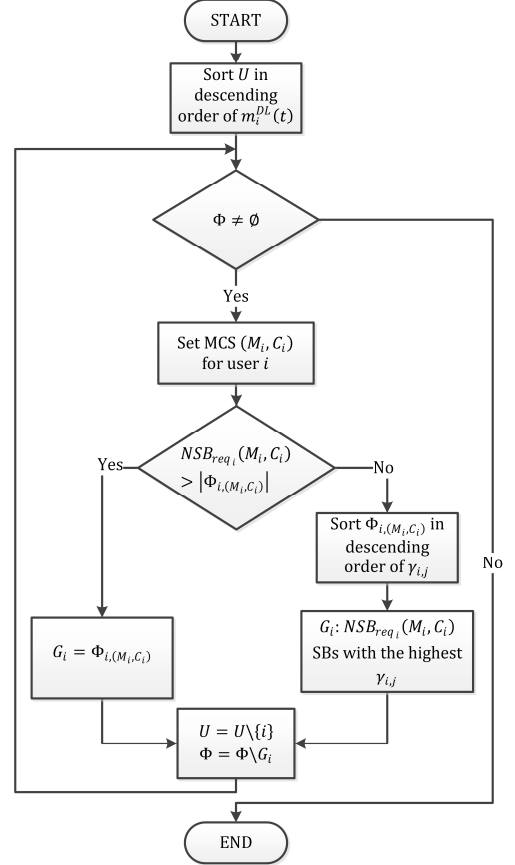


Fig. 1. Flowchart of the proposed algorithm

scheduling blocks in order to transmit its HoL packet as a whole. The required amount of scheduling blocks, $NSB_{req_i}(M_i, C_i)$, depends on the HoL packet size L_i , the modulation M_i and coding rate C_i , and the number of data carrying resource elements per downlink scheduling block L_{SB}^{DL} as follows:

$$NSB_{req_i}(M_i, C_i) = \left\lceil \frac{L_i}{C_i \log_2 M_i L_{SB}^{DL}} \right\rceil. \quad (5)$$

This number is calculated for all MCSs. Then, the MCS of user i is determined in (6), shown at the bottom of the page, where $|\Phi_{i,(M_i,C_i)}|$ is the number of available scheduling blocks of user i with MCS (M_i, C_i) . A scheduling block j is considered available to user i with MCS (M_i, C_i) if its Signal-to-Noise Ratio (SNR) $\gamma_{i,j}$ exceeds a specific threshold γ_{th} .

Therefore, in case there are enough scheduling blocks to accommodate the user's HoL packet as a whole, the MCS selected is the one that results in the need for the minimum number of scheduling blocks. Otherwise, the MCS is the one that results in the maximum number of bits that can be allocated to this user, i.e., $L_{i,(M_i,C_i)} = |\Phi_{i,(M_i,C_i)}| L_{SB}^{DL} C_i \log_2 M_i$.

A flowchart of the proposed resource allocation performed in each subframe is shown in Fig. 1. For each user $i \in U$, in descending order of $m_i^{DL}(t)$, the proposed downlink resource

$$(M_i^*, C_i^*) = \begin{cases} \arg \min_{(M_i, C_i)} \{NSB_{req_i}(M_i, C_i)\}, & \exists (M_i, C_i): NSB_{req_i}(M_i, C_i) \leq |\Phi_{i,(M_i,C_i)}| \\ \arg \max_{(M_i, C_i)} \{L_{i,(M_i,C_i)}\}, & \forall (M_i, C_i): NSB_{req_i}(M_i, C_i) > |\Phi_{i,(M_i,C_i)}| \end{cases} \quad (6)$$

allocation algorithm performs the following steps, if there are scheduling blocks available for allocation, i.e., set Φ is non-empty:

- 1) Firstly, the user's MCS is determined, based on (5) and (6).
- 2) If there are enough scheduling blocks to accommodate the user's HoL packet as a whole, i.e., $NSB_{req_i}(M_i, C_i) \leq |\Phi_{i,(M_i,C_i)}|$, the set $\Phi_{i,(M_i,C_i)}$ is ordered in descending order of SNR $\gamma_{i,j}$, and the $NSB_{req_i}(M_i, C_i)$ scheduling blocks of this set with the highest SNR, which are not necessarily contiguous, are the ones that comprise G_i , i.e., the set of all scheduling blocks allocated to user i in this subframe.
- 3) Otherwise, if the available scheduling blocks of set $\Phi_{i,(M_i,C_i)}$ are not enough to accommodate the packet as a whole, i.e., $NSB_{req_i}(M_i, C_i) > |\Phi_{i,(M_i,C_i)}|$, all of them are allocated to user i and the packet needs to be segmented.
- 4) When the resource allocation for user i is finalized, the user is removed from U and all its allocated scheduling blocks, i.e., belonging to G_i , are removed from the set Φ of available scheduling blocks.

If $\Phi \neq \emptyset$ and $U \neq \emptyset$, the resource allocation algorithm proceeds to the next user, otherwise the resource allocation for this subframe is complete and the algorithm terminates.

IV. PERFORMANCE EVALUATION

To evaluate the performance of the proposed resource allocation algorithm, a simulation model was built in MATLAB. The performance of the system employing the proposed algorithm is compared to three legacy systems that employ the PF, EE, and M-LWDF algorithms, respectively. Specifically, according to the PF algorithm, a scheduling block is allocated to the user who maximizes the ratio of instantaneous data rate to the average data rate, i.e., $m_{i,PF}^{DL}(t) = (r_i^{DL}(t)/\bar{R}_i^{DL}(t))$. The EE algorithm allocates a scheduling block to the user who maximizes the ratio of instantaneous data rate to the transmission power required, i.e., $m_{i,EE}^{DL}(t) = (r_i^{DL}(t)/P_{i,j})$, where $P_{i,j}$ is the transmission power of user i on scheduling block j . However, it has to be noted that no power control is performed in downlink LTE systems, therefore all scheduling blocks have the same transmission power level. Therefore, in this case the EE algorithm is similar to the maximum-rate algorithm. Finally, the M-LWDF algorithm allocates a scheduling block to the user who maximizes the product $m_{i,M-LWDF}^{DL}(t) = (-\log \alpha_i/d_{th,i})d_i^{DL}(t)(r_i^{DL}(t)/\bar{R}_i^{DL}(t))$, where α_i is the acceptable packet loss rate of user i .

The simulation environment consists of a single LTE cell and a variable number of UE devices within the cell's coverage area. The individual subsystems of the simulation model employed are as follows:

The *traffic generator* uses the Joint Scalable Video Model (JSVM) reference software [16] in order to generate variable-length video traffic frames for each UE, starting at a random instance within the first 33 ms of a simulation run. The video sequence used is the well-known "Highway" video sequence

Parameter	Value
Physical layer parameters	Channel bandwidth: 10MHz, Subframe duration (T_{sf}): 1ms, Number of RBs (N_{RB}): 50
Resource block format	Number of subcarriers per RB (N_{SC}^{RB}): 12, Number of symbols per RB (N_{Symb}^{DL}): 7, Subcarrier spacing: 15kHz
Reference Signal transmissions	4 Reference Signals per RB
Downlink control region size	2 OFDM symbols per subframe
TDD configuration	Configuration 1, DL:UL 3:2
Modulation and Coding Schemes	QPSK 1/2, 16-QAM 1/2, 64-QAM 3/4
Inter-eNodeB distance	500m
Path loss model	$128.1+37.6\log_{10} d$, d : distance from the eNodeB (km)
Transmitter antenna gain	18dBi
Receiver antenna gain	0dBi
Cable loss	2dB
Receiver Noise Floor	-95dBm
Interference margin	3dB
Control channel overhead	1dB
Shadowing	Log normal, $\sigma=8$ dB
Fading	Rayleigh
Maximum transmission power	20W
Maximum tolerable delay ($d_{th,i}$)	20ms
RLC mode	Unacknowledged mode (UM)
Traffic model	H264 video traffic QCIF 176x144
Protocol header sizes	RTP/UDP/IP with ROCH Compression: 3 bytes, PDCP: 2bytes, RLC: 3 bytes, MAC: 2 bytes, CRC: 3bytes
Segmentation bias (δ_{bias})	0.9
Average delay and rate calculation factor (β)	0.2

[17], with a rate of 30 frames per second (fps). The created video traffic frames are provided to the resource allocator.

The *channel model* simulates the physical layer channel conditions by providing path loss, shadowing, and short-term fading. It produces bit errors randomly for each connection, based on the allocated scheduling blocks and the MCS per user. Perfect channel knowledge is assumed for the purposes of Adaptive Modulation and Coding (AMC). The link budget parameters are summarized in Table III [18].

The *resource allocator* is the entity that is responsible for allocating the downlink resources to the different UE devices following either the proposed algorithm, or the legacy approaches.

The simulation scenario considers an increasing number of users, each one with one downlink video connection. The systems' performance is evaluated in terms of packet timeout rate, delay, goodput, fairness, and energy efficiency of successfully received bits. All simulation model parameters are summarized in Table III. In order to achieve statistical accuracy, 50 simulation runs were executed.

Fig. 2 depicts the packet timeout rate, i.e., the number of packets that are discarded due to expiration in the unit of time, with respect to an increasing number of users. In the case of the PF and EE systems, the packet timeout rate follows a sharp increase with the increase of the number of users due to the fact that the increased congestion results in excessive packet delays and packet expirations that cannot be avoided, since delay is not considered in their resource allocation process. In the case of the M-LWDF algorithm, the packet delay is taken

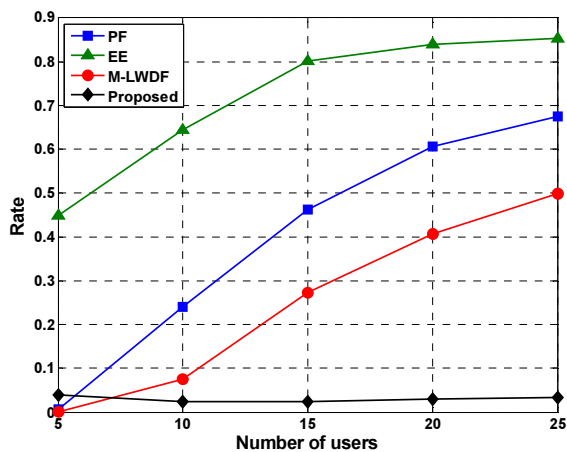


Fig. 2. Average packet timeout rate versus the number of users.

into consideration, however, the packets are not prioritized based on their segmentation status, therefore the excessive delay of at least one packet segment results in the expiration and discarding of the whole packet. Therefore, the system employing the proposed algorithm significantly outperforms the three legacy systems in terms of packet timeout rate. This is a result of the prioritization of users based on their packet delays with respect to their delay threshold as well as their segmentation status, therefore significantly reducing the packet expirations.

Fig. 3 depicts the average packet delay with respect to an increasing number of users. In the PF and EE systems, the average packet delay increases very quickly with the increase of the number of users. This is a result of the fact that these algorithms do not take into consideration the packet delay in the resource allocation process. As already described above, even though the M-LWDF algorithm takes into consideration the packet delay it does not prioritize the packets based on their segmentation status, therefore the increased delay of at least one packet segment results in the delayed packet reassembly at the receiving side. Therefore, the system employing the proposed algorithm, which prioritizes users with segmented packets and increased delay with respect to their delay threshold, significantly outperforms the three legacy systems in terms of average packet delay.

Fig. 4 depicts the fairness of the four systems. Fairness is evaluated using the Jain Index of Fairness, i.e., $FI = (\sum_{i \in UE} T_i(t))^2 / (|UE| \cdot \sum_{i \in UE} T_i^2(t))$, where $T_i(t)$ is the throughput of user i [19]. As expected, the PF and M-LWDF systems that take into consideration the average data rate of past allocations achieve higher fairness, compared to the EE system, which only considers the users' instantaneous data rates in the resource allocation. The system that employs the proposed algorithm achieves improved fairness compared to all the legacy systems. This is a result of the fact that the proposed algorithm also takes into consideration the average packet delay $\bar{D}_i^{UL}(t)$ in the user prioritization, favoring users that have experienced high average delay in past allocations.

Fig. 5 depicts the average goodput, i.e., the number of useful bits that reach the application layer in the unit of time. As it can be seen, the goodput follows a declining course with

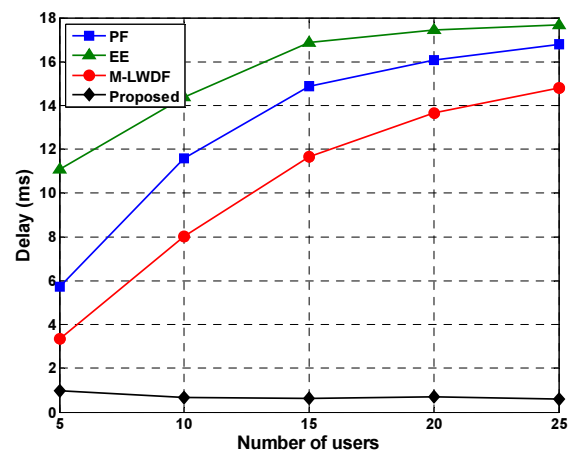


Fig. 3. Average delay versus the number of users

the increase of the number of users, as a result of the increasing congestion, which leads to excessive packet delays and timeouts. However, all the legacy systems experience a rapid deterioration of the goodput with the increase of the number of users. On the contrary, the system employing the proposed algorithm achieves a significantly improved goodput, even in the cases of increased number of users.

In order to highlight the interdependency of the energy efficiency and QoS provision, Fig. 6 depicts the systems' performance in terms of energy efficiency of successfully received bits. This is defined as the amount of data successfully concatenated at the receiver's RLC layer (in Mb) for a given amount of transmission energy (in J) and represents the average energy consumption per successfully received bit. As it can be seen, in the proposed system the energy efficiency of received bits is more than 6-times improved compared to that of the legacy systems. This is a result of the fact that, due to packet segmentation performed at the RLC layer, a packet segment loss may be unrecoverable at the receiving side, therefore leading to the waste of already received packet segments, whose transmission consumed energy. This could be partly mitigated by efficient ARQ schemes. However, these are not appropriate for real-time applications, since the required retransmissions induce additional delays that may result in a packet having expired

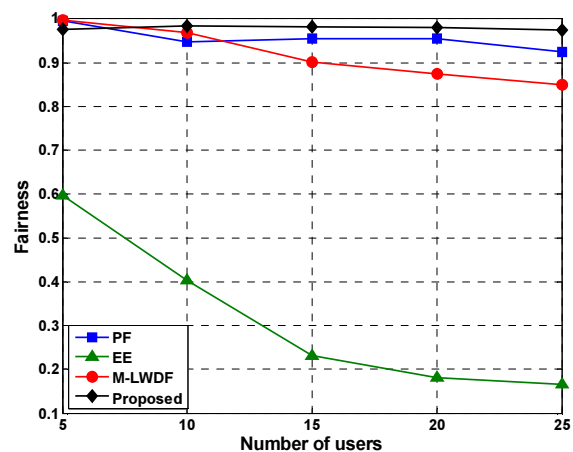


Fig. 4. Fairness versus the number of users.

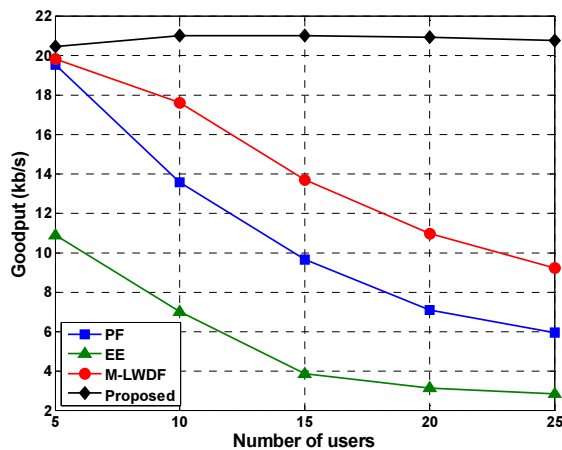


Fig. 5. Average goodput versus the number of users.

before being reassembled at the receiving side. This result highlights the effect that enhanced QoS provision has on energy efficiency, since the lower packet loss rate and the prioritization of segmented packets of the proposed system results in lower waste of already transmitted packet segments, and a larger amount of packets successfully being reassembled by the receiver RLC layer.

V. CONCLUSION

In this paper we introduced a downlink resource allocation algorithm for OFDMA systems, which focuses on QoS provision in real-time applications and energy efficiency. The proposed algorithm prioritizes users based on their estimated packet delay, the average delay and data rate of past allocations, as well as their buffer status with regards to packet segmentation. Simulation results highlight the considerable performance improvement achieved by the proposed algorithm compared to the PF, EE, and M-LWDF algorithms in terms of packet timeout rate, goodput, fairness, and average delay. In order to emphasize on the negative effect of poor QoS provision on energy efficiency, the system was also evaluated in terms of energy consumption per successfully received bit. Our plans for future work include the extension of the proposed downlink resource allocation to a multicell scenario, and the enhancement of its functionality with interference coordination and avoidance features.

ACKNOWLEDGMENT

We would like to acknowledge the support of the University of Surrey 5GIC (<http://www.surrey.ac.uk/5gic>) members for this work.

REFERENCES

- [1] F. Capozzi, G. Piro, L.A. Grieco, G. Boggia, and P. Camarda, "Downlink Packet Scheduling in LTE Cellular Networks: Key Design Issues and a Survey", *IEEE Communications Surveys & Tutorials*, vol.15, no.2, pp.678-700, Second Quarter 2013.
- [2] R. Kwan, C. Leung, and J. Zhang, "Proportional Fair Multiuser Scheduling in LTE", *IEEE Signal Processing Letters*, vol.16, no.6, pp.461-464, June 2009.
- [3] S. Shakkottai, and A.L. Stolyar, "Scheduling for multiple flows sharing a time-varying channel: The exponential rule", *Translations of the American Mathematical Society-Series 2*, vol.207, pp.185-202, 2002.

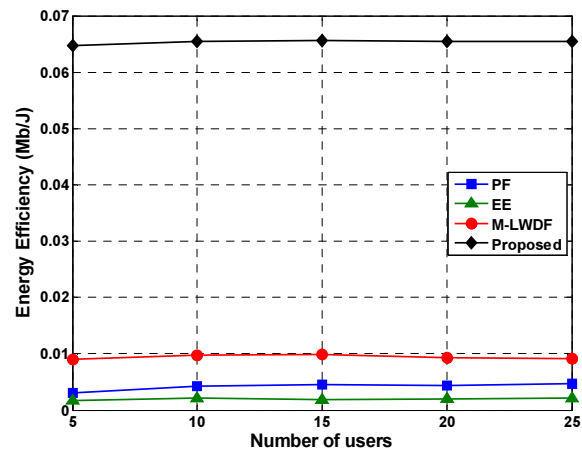


Fig. 6. Energy efficiency of successfully received bits versus the number of users.

- [4] B. Sadiq, S.J. Baek, and G. De Veciana, "Delay-optimal opportunistic scheduling and approximations: the log rule", *IEEE/ACM Transactions on Networking*, vol.19, no.2, pp.405-418, April 2011.
- [5] H.A.M. Ramli, R. Basukala, K. Sandrasegaran, and R. Patachaianand, "Performance of well known packet scheduling algorithms in the downlink 3GPP LTE system", in *IEEE 9th Malaysia International Conference on Communications (MICC)*, Kuala Lumpur, Malaysia, pp.815-820, Dec. 2009.
- [6] J.-H. Rhee, J.M. Holtzman, and D.-K. Kim, "Performance analysis of the adaptive EXP/PF channel scheduler in an AMC/TDM system", *IEEE Communications Letters*, vol.8, no.8, pp.497-499, Aug. 2004.
- [7] M. Andrews, K. Kumaran, K. Ramanan, A. Stolyar, P. Whiting, and R. Vijayakumar, "Providing quality of service over a shared wireless link", *IEEE Communications Magazine*, vol.39, no.2, pp.150-154, Feb 2001.
- [8] G. Piro, L.A. Grieco, G. Boggia, R. Fortuna, and P. Camarda, "Two-Level Downlink Scheduling for Real-Time Multimedia Services in LTE Networks", *IEEE Transactions on Multimedia*, vol.13, no.5, pp.1052-1065, Oct. 2011.
- [9] H. Luo, S. Ci, D. Wu, J. Wu, and H. Tang, "Quality-driven cross-layer optimized video delivery over LTE", *IEEE Communications Magazine*, vol.48, no.2, pp.102-109, Feb. 2010.
- [10] W.K. Lai, and C.-L. Tang, "QoS-aware downlink packet scheduling for LTE networks", *Elsevier Computer Networks*, vol.57, no.9, pp.1689-1698, May 2013.
- [11] C. Xiong, G.Y. Li, S. Zhang, Y. Chen, and S. Xu, "Energy- and Spectral-Efficiency Tradeoff in Downlink OFDMA Networks", *IEEE Transactions on Wireless Communications*, vol.10, no.11, pp.3874-3886, Nov. 2011.
- [12] S. Videv, and H. Haas, "Energy-Efficient Scheduling and Bandwidth-Energy Efficiency Trade-Off with Low Load", in *IEEE International Conference on Communications (ICC)*, Kyoto, Japan, pp.1-5, June 2011.
- [13] S. Videv, H. Haas, J.S. Thompson, and P.M. Grant, "Energy efficient resource allocation in wireless systems with control channel overhead", in *IEEE Wireless Communications and Networking Conference Workshops (WCNCW)*, Paris, France, pp.64-68, April 2012.
- [14] X. Xiao, X. Tao, and J. Lu, "QoS-Aware Energy-Efficient Radio Resource Scheduling in Multi-User OFDMA Systems", *IEEE Communications Letters*, vol.17, no.1, pp.75-78, Jan. 2013.
- [15] 3GPP TS 36.211 V12.3.0 (2014-09): 3rd Generation Partnership Project; Technical Specification Group Radio Access Network; Evolved Universal Terrestrial Radio Access (E-UTRA); Physical channels and modulation (Release 12).
- [16] Joint Scalable Video Model (JSVM) reference software, <http://www.hhi.fraunhofer.de/de/kompetenzfelder/image-processing/research-groups/image-video-coding/svc-extension-of-h264avc/jsvm-reference-software.html>
- [17] YUV Video Sequences, <http://trace.eas.asu.edu/yuv/>
- [18] H. Holma, and A. Toskala, *LTE for UMTS: Evolution to LTE-Advanced*, Chichester, U.K., John Wiley & Sons Ltd, 2011.
- [19] R. Jain, D. Chiu, and W. Hawe, "A Quantitative Measure of Fairness and Discrimination for Resource Allocation in Shared Computer Systems", DEC Research Report TR-301, Sept. 1984.