Delay-Oriented Resource Allocation for OFDMA Real-Time Mobile Broadband Services

https://doi.org/10.3991/ijim.v17i06.35885

Nadim K. M. Madi¹([⊠]), Alaa Omer Najim² ¹ Department of Information Systems, University of Nizwa, Nizwa, Oman ² Department of Surveying Techniques Engineering, Northern Technical University, Kirkuk, Iraq

nadim@unizwa.edu.om, alaa.omer@ntu.edu.iq

Abstract-Radio Resource Management (RRM) is the key component that influences the Quality of Service (QoS) in emerging mobile systems. The harsh requirements of recent multimedia applications press on the behavior of the resource allocation model at the Medium Access Control (MAC). This, under certain conditions, deteriorates awareness of the scheduler in satisfying QoS characteristics for the diverse attending traffic flows. The trade-off between QoS metrics may be a viable remedy to this problem. However, if the tradeoff is not meticulously addressed, it definitely increases the delay beyond the allowed threshold and plummets the data rates of the involved flows. In this article, a Delay-oriented Resource Allocation (DoRA) scheduler is proposed for the downlink channel to transmit multimedia applications. The main aim of DoRA is to schedule flows of different volumes with guaranteed low delay values and acceptable throughput under harsh network conditions (i.e., high traffic load and mobility). An efficient priority weight function is formulated based on a delay-oriented rule considering the buffer delay of the diverse flows. Eventually, a greedy algorithm assigns channel resources to the prioritized flows with high weights. DoRA is compared to reference schedulers using system-level simulation over two scenarios. The performance results indicate that DoRA maintains a low end-to-end delay with robust throughput and efficient spectrum efficiency for different flow volumes over congested network states.

Keywords—LTE, downlink recourse allocation, MAC scheduler, QoS, greedy algorithm, delay, fairness

1 Introduction

Mobile communication technologies have received wide attention due to the rapid escalation of rich broadband content. From 2011 up to the meantime, the advancements of mobile systems led by Long Term Evolution (LTE) networks have been axial in generating 5.5 billion new smartphone connections worldwide. A recent traffic forecast report by Ericsson indicated that such a tremendous reliance on mobile data led to exponential mobile data growth by the year 2021 which dominated about 43% of the total Internet traffic technologies [1]. This share is highly expected to be

further expanded from 78 exabytes (EB) in 2021 shooting up to 370 EB by 2027. The modern multimedia services are considered the QoS dominant traffic theme and the technology driver in evolved mobile systems [2], [3]. Therein, data rates of several Gbps with low fractions of latency under high-speed user mobility and dense scenarios are attainable for real-time applications. The user applications in this pool are proliferating faster than ever before to flow over the downlink stream. This is due to the nature of such services being bandwidth-hungry with burst and heterogeneous contents. In that, the prominent features of Orthogonal Frequency Division Multiple Access (OFDMA) technology are being leveraged at the downlink (DL) channel. This enables a high system capacity with ideal QoS gains by simultaneously transmitting traffic of different users who are multiplexed in time and frequency domains [4], [33].

The resource allocation process takes place at the evolved Node B (eNB), hence the OFDMA downlink channel is modeled in Frequency Domain (FD) and Time Domain (TD). The OFDMA channel allows allocating the pairs of radio Resource Blocks (RBs) from the same spectrum among multiple User Equipment (UEs) to provision the QoS of MBB services. Each RB has a capacity of 180 kHz and spans over 12 symmetrically spaced sub-carriers in FD. In TD, RBs are allocated to UEs each Time Transmission Interval (TTI), where each TTI spans over 1 ms [5]. For burst traffic, the transport block of data may prolong beyond a TTI. In this case, the LTE frame (equivalent to 10 consecutive TTIs) is utilized during the scheduling process to manage the transmission of such heavy contents by invoking packet fragmentation procedures at the downlink MAC scheduler in eNB.

Based on the intensive review of the related works in Section II, it is observed that the downlink resource allocation still has some limitations retarding it from guaranteeing a low delay for real-time contents when processing traffic with different QoS characteristics. For example, a related LTE downlink scheduler, namely Delay-based Weighted Proportional Fairness (DBWPF) [6], provides a good level of service fairness though for a specific LTE system model scenario. In finite traffic models, although a decent level of delay fairness is attainable by DBWPF, it notably deteriorates when variable bit rate (VBR) traffic is considered in the shared transmission channel for several UEs. In fact, DBWPF leverages a tunable QoS parameter to control the data window size for the attended flows. The QoS parameter, however, is defined statically by the user as a fixed value without any recommended range. This, certainly, restrains the scheduler's responsiveness to maintain the ideal QoS for variable bit rate traffic. Moreover, setting the parameter to a relatively high value (as applied in the referred model [6]) provides an odd scheduler decision that relies more on accepting flows based on their historical delay values (from previous scheduling intervals) rather than the current actual delay. This, eventually, leaves the current flows to starve more delay budget. Types of sensitive traffic that have variable flow volumes will severely be affected.

In this article, a delay-oriented recourse allocation algorithm for OFDMA-based systems, namely DoRA, is proposed to address the scheduling problem for VBR traffic. In details, the proposed work adopts a greedy and priority-based approach that utilizes content information such as actual buffer delay to influence the scheduling decision. DoRA effectively schedules flows of different data volumes whose delay budgets are relatively high and may tend to expire. The significance of the proposed algorithm dwells in the self-adaptable behavior to tune delay during high stages of traffic loads by an exponential utility function. This returns a persistent and ideal level of QoS for diverse traffic classes, especially, during congested states of the network communication scenario.

After the Introduction, the remaining contents are organized as follows: Section II describes the related works of downlink scheduling. In Section III, the proposed delay-based scheduler is presented. In Section IV, performance evaluation using simulation experiment scenarios is described and enclosed with a detailed discussion of the results. Finally, the conclusion of the proposed work is stated in Section V.

2 Related works

The radio resource management (RRM) concept in the evolving LTE mobile generations has been given serious attention by research and development domains. RRM is known as a set of processes and procedures over the comprising layers in the LTE protocol stack. In that, MAC and Physical layers accommodate the most influencing processes to the QoS provisioning. This calls for classifying the involved procedures as latency-oriented or throughput-based, based on the involved user applications [7]. In the downlink MAC scheduler, flows selection and resource allocation are substantial processes that occur in regular intervals (i.e., 1 ms) to deliver application contents to the designated users under QoS constraints. Bearing that in mind, the following details discuss the state-of-art and existing downlink resource allocation and scheduling efforts to improve QoS for certain user applications.

Most of the state-of-art resource allocation methods are classified into channelunaware and channel-aware [8], [9]. Channel-unaware scheduling assumes a timeinvariant and error-free transmission media. For instance, in [9], the authors discussed the principles of straightforward scheduling algorithms such as First In First Out (FIFO), Round Robin (RR), Blind Equal Throughput (BET), and Weighted Fair Queuing (WFQ) in LTE network. In FIFO, the users are served according to a chronological order using flows' arrival time. RR algorithm instead focuses on timesharing systems to realize pure fairness in the amount of time the channel is utilized by users. However, allocating an equal amount of time to users with different requirements is not good for intense-contents applications in terms of data rate. Considering the simplicity of RR, it was extended by [10] in a heuristic resource allocation algorithm. In that, the channel state information (CSI) is added to the RR principle aiming to boost the system capacity. On the other hand, the known principle of Earliest Deadline First (EDF) is widely used as a viable solution to the resource allocation problem. EDF counts for a delay weight on every active flow as it approaches the maximum delay bound. As an enhancement to this rationale, the authors in [11] proposed QoS-unsatisfied sessions by an Earliest Deadline First (QEDF) allocation algorithm for integrated WiFi and cellular system traffic offloading and congestion avoidance. QEDF assigns the channel resources to the session that does not meet the QoS requirements but is approaching the deadline. This algorithm works side by side with another method which allocates more bandwidth to sessions that fulfill the QoS requirements.

Channel-aware techniques benefit from the reported Channel Quality Indicator (CQI) feedbacks from the users during the scheduling process. The simplest form is the Maximum Throughput (MT) principle. Resource Blocks (RBs) in MT are assigned to the users with high CQIs. MT has been extended in many studies, for example in [12], the MT rule was involved in a utility-based resource allocation function named Hybrid Weighted Exponential Logarithm-Maximum Throughput (HWEL-MT). The function aims to increase the system throughput and limit the data loss ratio. Notably, HWEL-MT improves the system throughput and minimizes the delay, however, only for flows with burst traffic and at the expense of the other flow types.

Due to the radical biasness of MT to the system capacity, fairness is considered a major issue. Here, another legacy version of MT is defined as proportional fairness (PF) to rely on the scheduling decision on the channel throughput bounded by the historical average throughput. PF is widely accepted by many resource allocation solutions and thus has been diversely extended. In [13], a scheduling policy named Channel Aware Optimized Proportional Fair (CAOPF) was introduced to optimize the channel behavior that shares traffic considering the measured CQI values. The introduced scheduler relies on static user pre-defined parameters for channel and QoS weights which returns a short-range QoS performance guarantee though. Authors in [14] modified the conventional behavior of PF by integrating the buffer size of the flow while in [31] root mean square (RMS) method is considered to build the cost function of PF. The modified PF-based utility resulted in a mild performance between system throughput and fairness. However, it does not reveal a significant improvement with the QoS guarantee for real-time data.

PF was extended in [6] to handle delay-sensitive traffic by introducing DBWPF. DBWPF adopts a delay weight value that is calculated from historical information of buffer status to prioritize flows to improve delay fairness. Nonetheless, this mechanism results in a high compromise on throughput and deteriorated level of delay for all involved users. Authors in [15] utilized both MT and RR to develop a scheduler that aims at balancing throughput and fairness. Based on this approach, throughput is ensured high over the time domain scheduling. However, the fairness is maintained tight at the expense of more delay, especially on flows with low CQI.

The literature also reveals some comprehensive analysis of the different MAC scheduling procedures where holistic resource assignment models are developed for flows' selection and classification as seen in [16, 17, 18]. These models show an outstanding performance and yet efficient full utilization of the channel resources. The models are devoted to a specific function such as admission control [18], flows QoS analysis and RBs assignment mechanisms [17], or spectral efficiency boosting [16]. In this case, although QoS is provisioned in multi-dimensions, a wise concern should be drawn in terms of computational complexity costs that hinder implementing large-scale systems. A promising solution to this dilemma can be, for example, by considering energy-efficient resource allocation where the power level of the involved sub-channels is adaptively set based on the offered load as in [19].

The problem of resource allocation in the MAC scheduler is also addressed via solutions from different categories. In [26], jitter is assumed to be the main challenge to the real-time streaming video quality. A solution using a greedy algorithm is proposed for the MAC scheduler. The algorithm prioritizes different video flows based on a scheduling formula that counts for an exponential behavior of the buffer as well as the jitter estimation in the frames. Authors in [27] believe that integrating time and frequency domains returns an improved results for real-time and non-real-time applications. For that, the authors developed a channel-based frequency domain method as a top layer in the scheduler to improve the throughput. In the lower layer, a timedomain scheduling method is adopted hence a dynamic class-based establishment algorithm is used to solve the optimization problem for queue-based transmission. In [28], another viable solution for the problem of resource allocation is by using fuzzy logic. A multi-user content-aware priority-based scheduling algorithm is proposed for quantitatively classifying the flows based on the video content and channel conditions. The authors in [29] introduced a deep Reinforcement Learning (RL) approach to solve the resource allocation problem in time and frequency domains. The approach optimizes a set of objective functions considering pre-station channel quality and traffic information as input. The merits of neural networks applications attracted many researchers. For example, in [30] a radio resource allocation using a genetic algorithm (GA) in heterogeneous networks is proposed to optimize user association and resource block allocation and realize multi-QoS objectives. Some resource allocation solutions are especially designed for vehicular scenarios, for example [32], where interference level of neighboring vehicles is counted in the scheduling function.

From the demonstrated works above, it is observed that the delay and throughput should be two key criteria in developing a scheduling algorithm for the downlink channel. Considering network scenarios with rich bandwidth-hungry multimedia traffic of different volumes and desperate QoS profiles, the majority of the discussed scheduling algorithms impose limitations in guaranteeing an efficient level of data flows with a high rate and low latency, especially at high loads of network states. Thereupon, throughout this work, we introduce a delay-oriented scheduling algorithm for the downlink LTE MAC layer that allocates channel resources to different flows by employing a dynamic delay threshold to achieve low latency and high throughput.

3 DoRA: The proposed resource allocation algorithm

DoRA is introduced as an FD scheduling algorithm for the downlink channel to allocate resources to different UE flows in each TTI minding the adaptive QoS profiles. In this manifest, we assume that data flows are transmitted to respective UEs coming from the eNB (where DoRA is adopted) each TTI (1 ms). In general, eNB imposes a certain decision on the admitting flows to be transmitted depending on the reported CQI from a particular UE through the uplink signaling channel. The eNB then exploits the CQI to define the perfect modulation and coding scheme (MCS) to transmit the data for that UE with the least amount of channel errors. With high reported CQI

values, most user data is guaranteed to be successfully received on time with minor delay.

$$i = \arg\max_{i}(m_{i,m}(\mathbf{t})) \tag{1}$$

The process of radio resource allocation is adaptively recurring every TTI in the MAC scheduler to carry the various types of users' data through a shared channel. The resource allocation is accomplished based on the evaluation of flows' metrics weights on the available grid of RBs. At the t^{th} TTI, given the measured metric $M_{i,m}(t)$, the m^{th} RB is allocated to the i^{th} user if *i* satisfies,

To facilitate understanding the syntax and structure of the proposed DoRA algorithm, Table 1 depicts a description of the utilized parameters and notations. Also, the flow chart in Figure 1 demonstrates the rationale of DoRA's involved procedures. Without the loss of generality, DoRA is implemented at the MAC scheduler to enhance three properties of QoS: throughput, fairness, and delay-awareness. In the following sub-sections, we comprehensively describe the logic that accompanies each property and the impact of these integrated properties on the scheduling decision of DoRA. It is noteworthy to remark that DoRA is developed as a combination of all the above-mentioned properties. Therefore, DoRA scheduling metric $M_{i,m}(t)$ is presented by integrating the parameters of each property. The method of how DoRA is realizing each property is discussed phase by phase.

Notation	Description
i	UE index such that $i \in \{1, 2, 3,, n\}$
t	current Time Transmission Interval (TTI) in ms. (t = 1 ms)
$Hol_i(t)$	The measured delay of the UE i flow
τ	The maximum delay bound. It is pre-defined by the user.
$r_i(t)$	The actual data rate of the UE i channel
$\overline{R}_i(t)$	The average transmission rate of UE
$D_i(t)$	The delay weight function for flow i
$M_{i,m}(t)$	The priority metric of UE i on RB m on t TTI
x_i	Delay biasness parameter
\mathcal{Y}_i	Delay-throughput balancing parameter
Z _i	Slope coefficient of the delay function

Table 1. Notations description used in DoRA

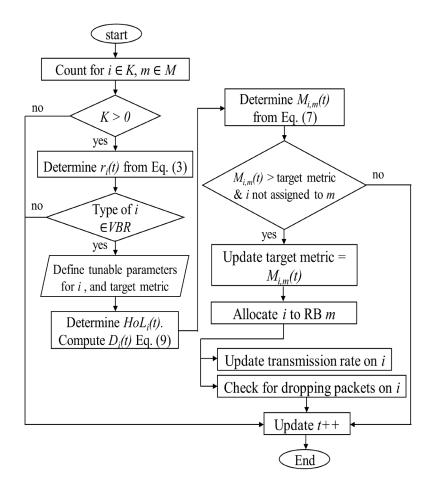


Fig. 1. Flowchart of the proposed resource allocation algorithm (DoRA)

3.1 Channel-awareness property in DoRA

In DoRA, the scheduler benefits from the reported CQI to deliberately measure the channel gain using the well-known Shannon channel capacity principle [20]. Therein, $r_i(t)$ is calculated from the Signal-Interference-to-Noise-Ratio (SINR). Each TTI, SINR is derived from the reported CQI via the control channel. In that, the simple channel-aware scheduling decision is expressed as,

$$M_{i,m}(\mathbf{t}) = r_i(\mathbf{t}) \tag{2}$$

$$r_i(t) = \beta \log_2(1 + \tau_i(t)) \tag{3}$$

β refers to the sub-channel bandwidth (equal to 180 kHz according to the LTE physical layer configurations). $τ_i(t)$ indicates the SINR that is the measured signal strength at UE *i* and it is obtained from the CQI value as follows,

$$\tau_i(\mathbf{t}) = \frac{p_i(\mathbf{t}) \mathbf{g}_i(\mathbf{t})}{Ipw_i(\mathbf{t}) + N pw_0}$$
(4)

where $p_i(t)$, $g_i(t)$, and $Ipw_i(t)$ are the power transmission rate of *i*, channel gain, and the returned interference amount by UE bearer *j* in the t^{th} TTI, respectively. Also, Npw_0 implies the Additive White Gaussian Noise (AWGN) power.

Equation (3) implies that the scheduling decision purely emphasizes throughput maximization. Hence, flows from UE with high CQI values are mostly prioritized for scheduling in each TTI. With this linear behavior of channel-aware resource allocation, the scheduling decision badly suffers from unfair service guarantee as more users with different channel conditions are engaged in the scenario. Technically, this means that cell-edge UEs are relatively denied from gaining radio resources to transmit their data.

3.2 Fairness property in DoRA

For DoRA scheduling decision to consider a sense of fair resource allocation among the involved UEs, we count on the historical transmission rate of a particular UE. Thus, a sort of fairness is established in the scheduling decision to guarantee service to those UEs with a low level of previous transmission rates; i.e, UEs with poor CQI values. The priority metric $M_{i,m}(t)$ with fair channel-aware can be expressed as,

$$M_{i,m}(\mathbf{t}) = \frac{r_i(\mathbf{t})}{\overline{R}_i(\mathbf{t})}$$
⁽⁵⁾

hence, $\overline{R}_{i}(t)$ is calculated as,

$$R_i(t) = \alpha R_i(t-1) + (1-\alpha)\mathbf{r}_i(t)$$
(6)

where α is a parameter that balances throughput with fairness hence $0 < \alpha < 1$. The priority metric value in (5) increases up to the limit of the historical transmission rate $\overline{R}_i(t)$ as the channel data rate of an instantaneous user grows. This condition allows a chance for the majority of the flows to be scheduled in a specific TTI as long as RBs are available. To this end, the scheduling rule based on (5) is considered the fundamental principle to guarantee a good system capacity for the LTE downlink network scenario.

3.3 Delay property in DoRA

As hinted above, the main aim of DoRA is to improve the QoS performance of the RT flows that are transmitted over a shared channel capacity. Therefore, the property of delay-awareness is reckoned as an integral part of the designed resource allocation algorithm. As a matter of fact, the key-design principle in DoRA is to significantly satisfy the QoS of data traffic with different buffer volumes (i.e., burst RT Video or light-weight VoIP data). This is realized by striking low latency and eminent level of throughput over the different traffic loads.

To guarantee QoS for delay-sensitive traffic, delay-driven parameters should be considered to ensure data transmission with low latency. In DoRA, we introduce a delay term that has an effective rule to control the queues of different users' data flows at eNB. The delay term $D_i(t)$ is integrated with the channel-aware principle in (5) to obtain a delay-based scheduling behavior as,

$$M_{i,m}(\mathbf{t}) = D_i(\mathbf{t}) \frac{r_i(\mathbf{t})}{\overline{R}_i(\mathbf{t})}$$
(7)

 $D_i(t)$ is deliberately developed to consider the buffer delay as it has a major impact on the overall end-to-end delay at the network. Therefore, $HoL_i(t)$ is adopted in the delay term as a straightforward measure of the buffer delay for a particular UE flow. For instance, in flow *i*, data load arrives contiguously in form of packets to its queue. Here, for each incoming packet to the queue, its arrival time is captured and stored. As illustrated in Figure 2, $HoL_i(t)$ is defined as the difference between the recent packet serving time and the stamped time for the arriving packet to the service queue. $HoL_i(t)$ is expressed as follows,

$$HoL_{i}(t) = T_{crt}(t) - T_{stm}(t)$$
(8)

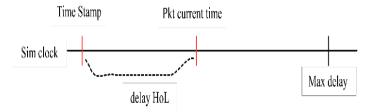


Fig. 2. Illustration of queue delay

where T_{crt} is the current incremented time of the packet in the flow buffer since its arrival. T_{stm} is the arrival time of the packet to the flow buffer.

Bearing in mind the above rationale of the flow buffer, the delay term in (7) can be decomposed to delay-awareness and fair throughput details by adopting the delay weight for each flow i (*HoL_i(t)*) based on the following formula,

$$D_{i}(t) = \frac{x_{i} \operatorname{HoL}_{i}(t)}{\exp\left(\sqrt{1 + \frac{y_{i}}{\operatorname{HoL}_{i}(t)}}\right) + z_{i}}$$
(9)

 x_i , y_i , and z_i are tunable QoS parameters to ideally adjust the behavior of $D_i(t)$ function such that low delay is maintained while still achieving high QoS performance (i.e., throughput and fairness). Adherence to the selection of these parameters' values highly affects the behavior of the overall obtained metric. In particular, they influence the scheduling metric function in the biasness degree towards delay-awareness which somehow establishes a trade-off relation between the service fairness and the throughput.

The central assumption in the proposed resource allocation scheme (DoRA) dictates that flows are imposed on the MAC scheduler in different volumes. This means that the scheduling decision should be conscious of the various $HoL_i(t)$ values. From (9), it is clearly demonstrated that the proposed delay term ($D_i(t)$) shows more consistent behavior in handling flows with different flows' volumes through the overload states of the scenario. This can be interpreted as a good robust level of the scheduling decision against the dynamic changes in the network scenario, and therefore, enable the scheduling flow of different volumes to satisfy their QoS profiles.

In (9), the influence of $D_i(t)$ to scheduling metric raises as long as $HoL_i(t)$ of flow *i* increases up to the value of the exponential term in the denominator part. Also, x_i contributes to $D_i(t)$ values as it is assigned a value greater than or equal to 1. In an arbitrary situation, if two flows have similar channel quality but each maintains a different buffer volume, the calculated $D_i(t)$ will be relatively higher for the flow with a large data buffer. In this case, the scheduler ensures the prioritization of this flow (with a heavier buffer) and transmits its data with the minimum delay budget. Small traffic flows though are still granted an opportunity of scheduling relies on the load distribution of the incoming traffic to the MAC layer. In addition, the other term of the channel-aware rule in (7) ensures that a level of fairness will always be maintained between different flows. Algorithm 1 and Figure 2 provide a further understanding of the resource allocation procedures invoked by DoRA scheduling algorithm for each TTI in the downlink channel.

Paper-Delay-Oriented Resource Allocation for OFDMA Real-Time Mobile Broadband Services

Algorithm 1: Resource Allocation Process in DoRA1 Initialization: 2 Define K as list of flows to be scheduled; 3 Define M as list of available RB; 4 Set TTIcounter = 0; targetMetric = 0; RBAllocation = false; 5 Event On TTI counter do 6 if $K > 0$ then 7 Continue; 8 else 9 Return; 10 end 11 for $m \leftarrow 1$ to M do 12 for $i \leftarrow 1$ to K do 13 Determine $r_i(t)$. based on Eq. (3); 14 if Type of flow $i \in VBR$ then 15 Invoke values of tunable parameters for $i;$ 16 Determine HoL delay on $i;$ 17 Compute $D_i(t)$ using Eq. (9); 18 Compute $D_i(t)$ using Eq. (7); 19 end 20 end 21 end 22 for $m \leftarrow 1$ to M do 23 for $i \leftarrow 1$ to K do 24 If $M_{i,m}(t) > targetMetric And i is notassigned to m then25 targetMetric = M_{i,m}(t);26 Allocate flow i to RB m;30 end$		
1Initialization:2Define K as list of flows to be scheduled;3Define M as list of available RB ;4Set TTIcounter = 0; targetMetric = 0;RBAllocation = false;5Event On TTIcounter do6if $K > 0$ then7 Continue;8else9 Return;10end11for $m \leftarrow 1$ to M do12 for $i \leftarrow 1$ to K do13 Determine $r_i(t)$. based on Eq. (3);14 if Type of flow $i \in VBR$ then15 Invoke values of tunable parameters for i ;16 Determine HoL delay on i ;17 Compute $D_i(t)$ using Eq. (9);18 Compute $M_{i,m}(t)$ using Eq. (7);19 end20if $M_{i,m}(t) > targetMetric And i is notassigned to m then23 for i \leftarrow 1 to K do24 if M_{i,m}(t) > targetMetric And i is notassigned to m then25 RBAllocation = true;27 end28end29 Allocate flow i to RB m;$		
2 Define K as list of flows to be scheduled; 3 Define M as list of available RB ; 4 Set TTIcounter = 0; targetMetric = 0; RBAllocation = false; 5 Event On TTIcounter do 6 if $K > 0$ then 7 Continue; 8 else 9 Return; 10 end 11 for $m \leftarrow 1$ to M do 12 for $i \leftarrow 1$ to K do 13 Determine $r_i(t)$. based on Eq. (3); 14 if Type of flow $i \in VBR$ then 15 Invoke values of tunable parameters for i ; 16 Determine HoL delay on i ; 17 Compute $D_i(t)$ using Eq. (9); 18 Compute $M_{i,m}(t)$ using Eq. (7); 19 end 20 end 21 end 22 for $m \leftarrow 1$ to M do 23 for $i \leftarrow 1$ to K do 24 if $M_{i,m}(t) > targetMetric And i is not 25 assigned to m then 26 targetMetric = M_{i,m}(t);26 Allocate flow i to RB m;$		
$\begin{array}{c c c c c c c c c c c c c c c c c c c $		
4 Set TTIcounter = 0; targetMetric = 0; RBAllocation = false; 5 Event On TTIcounter do 6 if $K > 0$ then 7 Continue; 8 else 9 Return; 10 end 11 for $m \leftarrow 1$ to M do 12 for $i \leftarrow 1$ to K do 13 Determine $r_i(t)$. based on Eq. (3); 14 if Type of flow $i \in VBR$ then 15 Invoke values of tunable parameters for i ; 16 Determine HoL delay on i ; 17 Compute $D_i(t)$ using Eq. (9); 18 Compute $M_{i,m}(t)$ using Eq. (7); 19 end 20 end 21 end 22 for $m \leftarrow 1$ to K do 23 for $i \leftarrow 1$ to K do 24 if $M_{i,m}(t) > targetMetric And i is not assigned to m then 25 end 26 RBAllocation = true; 27 end 28 end 29 Allocate flow i to RB m;$		
$\begin{array}{c c c c c c c c c c c c c c c c c c c $		
$\begin{array}{c c c c c c c c c c c c c c c c c c c $		
$ \begin{array}{c c c c c c c c c c c c c c c c c c c $		
$\begin{array}{c c c c c c c c c c c c c c c c c c c $		
9 Return;10end11for $m \leftarrow 1$ to M do12 for $i \leftarrow 1$ to K do13 Determine $r_i(t)$. based on Eq. (3);14 Invoke values of tunable15 Invoke values of tunable16 Determine HoL delay on i ;17 Compute $D_i(t)$ using Eq. (9);18 end20 end21end22for $m \leftarrow 1$ to M do23 for $i \leftarrow 1$ to K do24 if $M_{i,m}(t) > targetMetric And i is not assigned to m then25 largetMetric = M_{i,m}(t);26 RBAllocation = true;27 end28 end29 Allocate flow i to RB m;$		
10end11for $m \leftarrow 1$ to M do12for $i \leftarrow 1$ to K do13Determine $r_i(t)$. based on Eq. (3);14if Type of flow $i \in VBR$ then15Invoke values of tunable parameters for i ;16Determine HoL delay on i ;17Compute $D_i(t)$ using Eq. (9);18Compute $M_{i,m}(t)$ using Eq. (7);19end20end21end22for $m \leftarrow 1$ to M do23for $i \leftarrow 1$ to K do24if $M_{i,m}(t) > target Metric And i is notassigned to m then25larget Metric = M_{i,m}(t);RBAllocation = true;27end28end29Allocate flow i to RB m;$		
11for $m \leftarrow 1$ to M do12for $i \leftarrow 1$ to K do13Determine $r_i(t)$. based on Eq. (3);14if Type of flow $i \in VBR$ then15Invoke values of tunable parameters for i ;16Determine HoL delay on i ;17Compute $D_i(t)$ using Eq. (9);18Compute $M_{i,m}(t)$ using Eq. (7);19end20end21end22for $m \leftarrow 1$ to M do23for $i \leftarrow 1$ to K do24if $M_{i,m}(t) > targetMetric And i is notassigned to m then25largetMetric = M_{i,m}(t);RBAllocation = true;27end28end29Allocate flow i to RB m;$		
12for $i \leftarrow 1$ to K do13Determine $r_i(t)$. based on Eq. (3);14if Type of flow $i \in VBR$ then15Invoke values of tunable parameters for i ;16Determine HoL delay on i ;17Compute $D_i(t)$ using Eq. (9);18Compute $M_{i,m}(t)$ using Eq. (7);19end20end21end22for $m \leftarrow 1$ to M do23if $M_{i,m}(t) > targetMetric And i is notassigned to m then25targetMetric = M_{i,m}(t);26targetMetric = M_{i,m}(t);27end28end29Allocate flow i to RB m;$		
13 Determine $r_i(t)$. based on Eq. (3); 14 if Type of flow $i \in VBR$ then 15 Invoke values of tunable 15 Determine HoL delay on i ; 16 Determine HoL delay on i ; 17 Compute $D_i(t)$ using Eq. (9); 18 Compute $M_{i,m}(t)$ using Eq. (7); 19 end 20 end 21 end 22 for $m \leftarrow 1$ to M do 23 for $i \leftarrow 1$ to K do 24 if $M_{i,m}(t) > targetMetric And i is not assigned to m then 25 RBAllocation = true; 27 end 28 end 29 Allocate flow i to RB m; $		
14if Type of flow $i \in VBR$ then15Invoke values of tunable parameters for i ;16Determine HoL delay on i ;17Compute $D_i(t)$ using Eq. (9);18Compute $M_{i,m}(t)$ using Eq. (7);19end20end21end22for $m \leftarrow 1$ to M do23for $i \leftarrow 1$ to K do24if $M_{i,m}(t) > targetMetric And i is notassigned to m then25ItargetMetric = M_{i,m}(t);26RBAllocation = true;27end28end29Allocate flow i to RB m;$		
15Invoke values of tunable parameters for i ; Determine HoL delay on i ; Compute $D_i(t)$ using Eq. (9); Compute $M_{i,m}(t)$ using Eq. (7);17end20end21end22for $m \leftarrow 1$ to M do23for $i \leftarrow 1$ to K do24if $M_{i,m}(t) > targetMetric And i is notassigned to m then25 26end27end28end29Allocate flow i to RB m;$		
16parameters for i ; Determine HoL delay on i ; Compute $D_i(t)$ using Eq. (9); Compute $M_{i,m}(t)$ using Eq. (7);19end20end21end22for $m \leftarrow 1$ to M do23for $i \leftarrow 1$ to K do24if $M_{i,m}(t) > targetMetric And i is notassigned to m then25 26 27 28endAllocate flow i to RB m;$		
16Determine HoL delay on i ; Compute $D_i(t)$ using Eq. (9); Compute $M_{i,m}(t)$ using Eq. (7);19end20end21end22for $m \leftarrow 1$ to M do23for $i \leftarrow 1$ to K do24if $M_{i,m}(t) > targetMetric And i is notassigned to m then25end26RBAllocation = true;27end28end29Allocate flow i to RB m;$		
17Compute $D_i(t)$ using Eq. (9);18Compute $M_{i,m}(t)$ using Eq. (7);19end20end21end22for $m \leftarrow 1$ to M do23for $i \leftarrow 1$ to K do24if $M_{i,m}(t) > targetMetric And i is not assigned to m then25 26 27 end28end29 Allocate flow i to RB m;$		
18 Compute $M_{i,m}(t)$ using Eq. (7); 19 end 20 end 21 end 22 for $m \leftarrow 1$ to M do 23 for $i \leftarrow 1$ to K do 24 if $M_{i,m}(t) > targetMetric And i is not assigned to m then 25 targetMetric = M_{i,m}(t); 26 RBAllocation = true; 27 end 28 end 29 Allocate flow i to RB m; $		
19end20end21end22for $m \leftarrow 1$ to M do23for $i \leftarrow 1$ to K do24if $M_{i,m}(t) > targetMetric And i is not assigned to m then25 26 27 28end29 Allocate flow i to RB m;$		
20end21end22for $m \leftarrow 1$ to M do23for $i \leftarrow 1$ to K do24if $M_{i,m}(t) > targetMetric And i is not$ 25l26l27end28end29Allocate flow i to RB m ;		
21end22for $m \leftarrow 1$ to M do23 for $i \leftarrow 1$ to K do24 if $M_{i,m}(t) > targetMetric And i is not assigned to m then25 26 27 RBAllocation = true;28end29 Allocate flow i to RB m;$		
22for $m \leftarrow 1$ to M do23for $i \leftarrow 1$ to K do24if $M_{i,m}(t) > targetMetric And i is not25assigned to m then25targetMetric = M_{i,m}(t);26RBAllocation = true;27end28end29Allocate flow i to RB m;$		
23for $i \leftarrow 1$ to K do24if $M_{i,m}(t) > targetMetric And i is not25if M_{i,m}(t) > targetMetric = M_{i,m}(t);26 targetMetric = M_{i,m}(t);27end28end29Allocate flow i to RB m;$		
24if $M_{i,m}(t) > targetMetric And i is not$ assigned to m then25 26 27 28end29 Allocate flow i to RB m;		
25 $assigned to m then$ $targetMetric = M_{i,m}(t);RBAllocation = true;27end28end29Allocate flow i to RB m;$		
25 $ $ targetMetric = $M_{i,m}(t)$; 26 $ $ RBAllocation = true; 27 end 28 end 29 Allocate flow i to RB m ;		
26RBAllocation = true;27end28end29Allocate flow i to RB m ;		
27 end 28 end 29 Allocate flow <i>i</i> to RB <i>m</i> ;		
$\begin{array}{c c} 28 & \mathbf{end} \\ 29 & \text{Allocate flow } i \text{ to RB } m; \end{array}$		
29 Allocate flow i to RB m ;		
end		
Update Transmission rate on flow i ;		
Check for dropping packets on flow i ;		
TTI counter + +;		
34 end		

3.4 QoS control parameters

The QoS-based parameters x_i , y_i , and z_i have an effective influence in controlling the behavior of the scheduling metric as more flows admit into the scheduler's classifier. x_i ensures a certain level of delay guarantee by influencing the scheduling principle to rely more on prioritizing flows with a high HoD delay value. Ironically, y_i is

adopted in the denominator of the delay term $D_i(t)$ to strike a balancing behavior for the scheduler and allocate RBs based on the channel experience. z_i is a co-efficient parameter that tunes the slope of the delay term. As shown in Figure 3, different settings of the delay-awareness control parameters result in various values of the delay term. As x_i value is set to a value that is greater than y_i , $D_i(t)$ value expands; thus, the scheduler treats flow *i* in a way to decrease delay rather than guarantee fair service with other flows. Inversely, setting x_i to a smaller value than y_i decreases the chance of scheduling flow *i* based on delay meanwhile it ensures fair service.

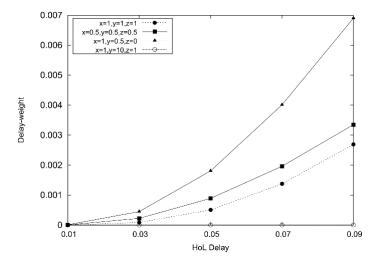


Fig. 3. The impact of the buffer size on the delay weight with different settings of the QoS control parameters

4 Performance evaluation

This section reveals the performance evaluation details of the proposed scheduling scheme which is carried out via extensive system-level simulations experiments. The results are analyzed via four QoS metrics: End-to-End delay, aggregated throughput, Packet Loss Ratio (PLR), and spectral efficiency. To realize the performance significance, the evaluation results of the proposed scheme (DoRA) are discussed against a related reference scheme, namely, DBWPF [6] which is implemented in the same scenario. The system-level simulation experiments were carried out using the event-driven and open-source simulation tool, LTE-Sim [21]. The simulations are conducted based on a defined system model as shown in Figure 4 to evaluate the performance of the proposed algorithm. The following details elaborate on the applied scenario, the performance metrics definitions, and the numerical simulation results.

4.1 Performance metrics

In this work, we count the end-to-end delay as a major criterion to demonstrate the significance of DoRA. End-to-end delay is expressed below,

$$Delay_{e2e}(t) = \frac{1}{N} \sum_{i=1}^{N} (\mathbf{R}_i - G_i)$$
(10)

where G_i is the generation time of packet *i* at the source, R_i is the receiving time of a packet at the MAC layer of the receiver, and *N* is the total number of packets for all flows in the system. End-to-end delay $Delay_{e2e}$ is reckoned as the main performance metric in this study hence it measures all the possible sorts of time-consuming sources in the system: propagation, transmission, queuing, and processing delays. Achieving low values of $Delay_{e2e}$ when transmitting flows of different sizes is the ultimate goal of DoRA scheme to enhance the QoS of MBB applications in LTE systems.

The fairness Index indicates the share level of the users' flows in using the channel resources. It measures the equity level of throughput provided to each user data flow for traffic within the same class. In this study, we utilize Jain's Fairness Index [22] which is expressed as,

$$J_{fairness} = \frac{\left(\sum_{i=1}^{l} r_i(t)\right)^2}{I\sum_{i=1}^{l} r_i(t)^2}$$
(11)

where $I_{fairness}$ represents the number of active users in the system, and $r_i(t)$ is the throughput for the i^{th} connection (user channel). Fairness in this scope means providing a sufficient channel of resources for each flow to transmit its data.

Throughput is the rate of successful packets that the communication channel can forward from a sender to a receiver. System throughput is calculated by,

$$Throughput = \frac{\sum_{i=0}^{N} size_i(t)}{T}$$
(12)

In this work, we measure the aggregated throughput. It indicates the cell throughput contributed by all the involved traffic classes (flow types). The high value of aggregated throughput means that the eNB operates in high capacity but a minor portion of data may be compromised.

The PLR is defined as the ratio of data that does not successfully reach the destination due to different reasons. These reasons include, for instance, violated delay at eNB MAC, exceeding the number of maximum re-transmissions, or arriving at the destination with lost fragments. PLR can be calculated as the inverse of delivery ratio, where pkt_{TX} is the total number of transmitted packets, and pkt_{RX} reflects the successfully received packets,

$$PLR = 1 - \frac{pkt_{TX}}{pkt_{RX}}$$
(13)

Spectral efficiency (SE) is a measure of the efficient use of spectrum resources. In general, spectral efficiency is defined as the information rate that can be transmitted over a given bandwidth. Assuming M is the total number of RBs in the system where $m = \{1, 2, 3, ..., M\} \in M$, and $r_i(t)$ is the data rate of user flow i, the spectral efficiency can be expressed as,

$$SE = 1 - \frac{\sum_{i=1}^{l} r_i(\mathbf{t})}{\sum_{m=1}^{M} \beta}$$
(14)

Maintaining a relatively high value of spectrum efficiency implies that the scheduler is able to boost the system capacity by making the best allocation mapping of the RBs to flows that have a high channel quality experience. Nonetheless, it is noteworthy to mark that absolute emphasis on spectral efficiency may constrain ensuring a low end-to-end delay, especially, during the overload states of the network when resources become scares.

4.2 Simulation scenario

The applied network topology for the simulation is depicted in Figure 4. Therein, the system consists of a single eNB and a number of 100 UEs that are uniformly distributed within the transmission range of the eNB. The eNB is operating with a transmission radius of 1 km, and each UE travels using the Random Way-point mobility model. In the experiments' setup, we simulate two UEs speed scenarios: an urban speed of 30 km/h and a highway speed of 120 km/h. The traffic source is an application server that provides a VBR traffic type. In this study, we assume that each UE establishes one connection (physical channel) with the eNB and carries 2 flows of traffic in a given time. The inbound traffic is prioritized based on the delay-weighted scheduling algorithm. The flow with the highest value is prioritized depending on the value of QoS-based parameters. All the QoS tunable parameters (x_i , y_i , and z_i) are set to 1 according to [23] which maintains the ideal balancing between delay and fairness. In the experiment, the effect of weighted delay in the scheduler depends on tunable parameters and $HoL_i(t)$ delay. The specifications of the physical air interface follow the default 3GPP 4G standard specifications as in [24], [25].

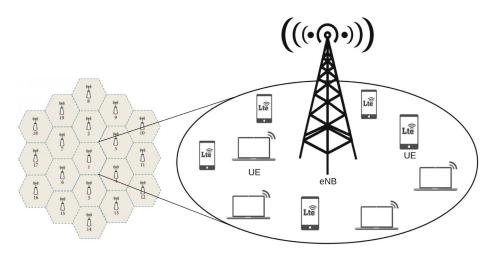


Fig. 4. The simulated network topology

This study focuses on the downlink transmission of the LTE system where users simultaneously run different traffic volumes. We mainly follow the same experimental setup described in the benchmark [6] with a minor modification as highlighted here. In this simulation scenario, the eNB delivers a diverse volume of VBR traffic sources to I numbers of UE. By convention, each UE is linked with a dedicated VBR traffic source which generates packets of variable intervals that are determined using a uniform random function. This pattern enables the user application to offer variable data rates between 0.2 - 0.6 Mbps for the pre-defined packet size. An incremented volume of (20-100) UEs is located at the central cell; other cell sites though operate as interference sources to mimic the real-life scenario. Each UE is uniformly distributed between 500 and 1000 m, with a speed considered in two cases (30 and 120 Km/h). A cluster of four cells is formed and regularly replicated to cover the entire service area to employ the frequency reuse concept. Each eNB is assumed to be equipped with two antennas and operate on a transmission power of 43 dBm. A total number of 50 available sub-channels is configured based on the bandwidth capacity to transmit different transport block sizes (TBS). The utilized TBS size depends on the selected Modulation and Coding Scheme (MCS) at the physical layer. The normal mode of Cycle Prefix (CP) and the OFDM symbols of the Physical Downlink Control Channel (PDCCH) are assumed as recommended by the standardized specifications in [25], [26]. Moreover, a periodic CQI is reported by UE every two TTI is assumed over the full bandwidth capacity. The corresponding UE's CQI feedback is then sent to the eNB through the uplink control channel of the UEs. The overall simulation parameters are summarized in Table 2.

Parameter	Description
Simulation time	10 ⁶ ms
Carrier frequency	2 GHz
Bandwidth	10 MHz (50 subchannels per TTI) [25]
Number of UEs	100 UE
UE speed	30 and 120 km/h
Max delay	100 ms
Mobility model	Random waypoint
Number of cells	19 macro-cells
Cell radius	1000 m
Traffic volume	Uniformly distributed in {0.2, 0.6} Mbps for each UE
eNB TX power	43 dBm
Path loss	128.1+37 log10
Thermal Noise density	-174 dBm/Hz
Noise Figure	7 dB
Carrier frequency	2 GHz

Table 2. Description of the Overall Simulation Parameters

4.3 Results and discussions

The experienced average delay against the increased number of UE over different user mobility levels is presented in Figure 5 and Figure 6. When UE speed is 30 km/h, the eNB MAC scheduler transmits most of the offered network load because the channel variations are not high. Therefore, our proposed scheduling algorithm DoRA significantly minimizes the experienced delay of the attended UE flows to be less than the 3GPP standardized threshold (100 ms) defined for multimedia traffic [24].

This achievement is attributed to the adoption of a delay-dependent scheduling decision by DoRA that prioritizes flows with high $HoL_i(t)$ delay values. Therefore, all UE flows are transmitted with low delay and also improve the measured channel data rate. Remarkably, the improvement in delay by DoRA remains in a steady pattern compared with DBWPF when UE mobility is high. In fact, DBWPF mainly emphasizes on delay fairness rather than reducing the experienced delay; hence, the scheduler decision is restricted to the buffer size and delay of the UE buffer in previous TTI, which induces DBWPF to ensure low delay as traffic volume varies.

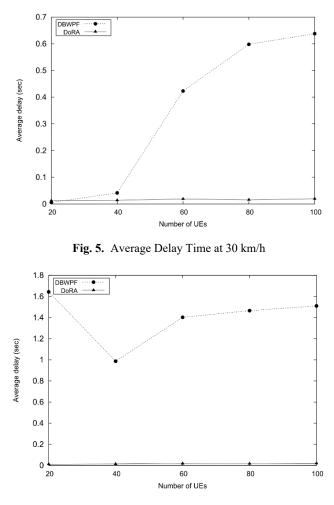


Fig. 6. Average Delay Time at 120 km/h

Figures 7 and 8 present the results of flows Jain's fairness index. Generally speaking, as the MAC scheduler reckons for channel awareness property, the fairness pattern decreases as the number of UE involved in the system increases. The proposed scheduling algorithm, DoRA, outperforms the referenced scheme, DBWPF, over different user mobility levels. DoRA sustains ideal throughput fairness from two aspects that complement each other. On one hand, it limits the user channel data rate to its recorded historical transmission rate. This allows users with low channel states to have a selected for being assigned to RBs and transmit their payloads. Besides, in the implemented delay function, fairness is ensured by defining the QoS tunable parameters (i.e., y_i , and x_i). In that, the scheduling decision is applied by prioritizing the metric as long as the delay weight of the flow is relatively high (based on the $HoL_i(t)$) and limited to the exponential term of its delay as expressed in (9). In the case of DBWPF, UEs experience the worst throughput fairness level. Basically, the principle

of DBWPF scheduling suggests ensuring the fairness level of experienced delay over the attended flows rather than scheduling flows with high throughput. This condition eventually results in a low throughput fairness trend especially when channel variability is high.

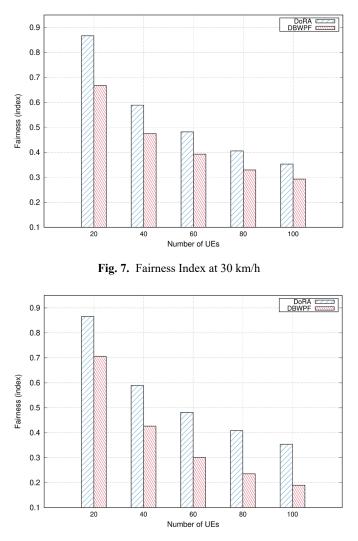


Fig. 8. Fairness Index at 120 km/h

The aggregated system throughput with respect to the number of UEs over the two modes of speed is illustrated in Figure 9 and Figure 10. According to the presented results, DoRA achieves a prominent improvement in the system throughput concerning the referenced algorithm. DoRA can maintain a high level of throughput due to the configurable QoS parameters. Wherein, x_i and y_i are set to proper values such that

the upper part of $D_i(t)$ in (9) is restricted to the limit of the exponential function in the lower part. Consequently, this condition influences the scheduler's decision to prioritize based on throughput if the HoL delays are relatively low while increasing the delay dependency decision when HoL values increase.

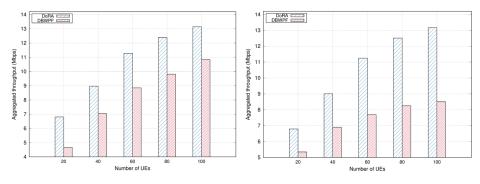


Fig. 9. Aggregated Throughput at 30 km/h

Fig. 10. Aggregated Throughput at 120 km/h

The results of PLR with the increasing flows volume are presented in Figure 11 and Figure 12. Despite the low number of UE flows offered at the light load network state, DBWPF draws the worst behavior. The PLR is reduced up to 35% compared with DBWPF when 20 UE are connected to the network with high mobility (120 km/h). This is due to the reason that improving the delivery rate is not the core aim of DBWPF procedures. This in turn talks to a high amount of data loss.

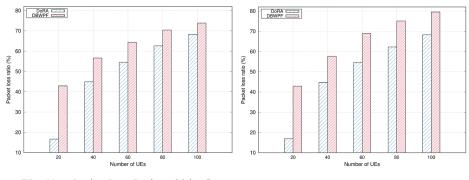


Fig. 11. Packet Loss Ratio at 30 km/h

Fig. 12. Packet Loss Ratio at 120 km/h

Regarding the proposed scheduler, it greatly keeps a low base of the resultant PLR despite the slight increase in the case of high UE mobility and network load. When UEs are moving at the speed of 30 km/h and the cell is loaded with 50 UEs, the packet loss falls to 50% compared to the level in DBWPF. This is interpreted for the same reason as mentioned in the discussions of the throughput results. The ability of DoRA to schedule flows to achieve high throughput explains the low behavior of PLR. In fact, the three QoS tunable parameters enable DoRA scheduling rule to return good configuration flexibility for the scheduler. This helps the flows to obtain relatively

closed priority weight values. With that, the scheduler can guarantee that the majority of flows are scheduled with less data loss.

The results of spectral efficiency over different UE speeds are illustrated in Figure 13 and Figure 14. Normally, spectral efficiency increases when the gained system throughput is high. Therefore, DoRA is expected to achieve the highest cell spectral efficiency with respect to the reference scheduler. At the low network load, DoRA maintains a high spectral efficiency as more UEs connect the cell. With these results, DoRA is seen to utilize channel resources more wisely; hence, most of the generated traffic processed by the base station is successfully delivered to their respective UEs. Also, DoRA scheduler considers close attention to the UEs channels with high SINR. By the rule of thumb, assigning RBs to flows belonging to high SINR channels will indeed improve resource utilization and thus increases the system SI.

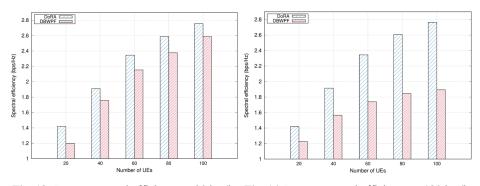


Fig. 13. System spectral efficiency at 30 km/h Fig. 14.System spectral efficiency at 120 km/h

On the other hand, DBWPF demonstrates the worst case of spectral efficiency during high UE speed. This result leads to the conclusion that relying on the principle of scheduling different flows to balance delay results in severely deteriorating the achieved cell capacity. UE flows with high CQI values are not always scheduled in certain TTI because of the impact of the delay term in DBWPF. Consequently, the majority of channel resources are utilized with low modulation schemes which return low spectral efficiency.

5 Conclusion

In this article, DoRA algorithm is proposed as a promising solution to enhance the behavior of resource allocation for multimedia applications. The analysis of the related work demonstrates a limitation in that the downlink MAC scheduler function is insufficient to guarantee a prominent level of data rate and low latency, especially, at high loads of network states. The existing solutions, ironically, emphasize a single dimension QoS guarantee (throughput, or delay) uncontrolled compromise on other QoS metrics as seen by the reference method, namely DBWPF. For that, DoRA was introduced with a major aim to work for modern mobile systems networks and handle diverse QoS profiles for RT variable bit rate traffic aiming to decrease delay and sus-

tain long-term data rate fairness. To do that, DoRA scheduler adopts a greedy algorithm hence a novel scheduling rule is formulated by exploiting the flows' HoL delay, dynamic delay threshold, and adaptive throughput formulation to achieve low latency and consistent channel gain for different traffic buffer volumes. The performance evaluation is carried out using system-level simulations over the user mobility scenarios of 30 Km/h and 120 Km/h. The results are discussed comparatively with the related scheduling scheme (DBWPF) in terms of end-to-end delay, fairness, aggregated throughput, PLR, and SE. The outcomes indicate a tangible decrease in the end-to-end delay and PLR even with a high network load. Besides, throughput fairness, average data rate, and spectral efficiency are maintained at eminent levels in the case of adopting DoRA as a MAC scheduler. The obtained results of DoRA substantiates a reliable implementation of the eNB downlink MAC scheduler which can be deployed in modern mobile broadband networks and guarantee a long-term QoS for the user services. A future direction of this work is to extend the algorithm to consider energy efficiency for emerging 5G indoor communication scenarios.

6 Acknowledgment

The authors would like to acknowledge the University of Nizwa with its IS Environment research cluster that functions under the department of Information Systems (DIS) for their valuable support.

7 References

- Ericsson, "Ericsson Mobility Report: Mobile data traffic increased almost 300-fold over 10 years," *Telefonaktiebolaget LM Ericsson 1994-2022*, vol. 2021, 2021. Telefonaktiebolaget LM Ericsson.
- [2] A. Mamane, M. Fattah, M. El Ghazi, and M. El Bekkali, "5G Enhanced Mobile Broadband (eMBB): Evaluation of Scheduling Algorithms Performances for Time-Division Duplex Mode," *iJIM*, vol. 16, no. 01, p. 121, 2022. <u>https://doi.org/10.3991/ijim.v16i01.25941</u>
- [3] H. Kim, Design and optimization for 5G wireless communications. John Wiley & Sons, 2020. <u>https://doi.org/10.1002/9781119494492</u>
- [4] M. Vaezi, Z. Ding, and H. V. Poor, Multiple access techniques for 5G wireless networks and beyond. Springer, 2019. <u>https://doi.org/10.1007/978-3-319-92090-0</u>
- [5] H. Mousavi, I. S. Amiri, M. Mostafavi, and C. Choon, "LTE physical layer: Performance analysis and evaluation," *Applied computing and informatics*, vol. 15, no. 1, pp. 34-44, 2019. <u>https://doi.org/10.1016/j.aci.2017.09.008</u>
- [6] S. Liu, C. Zhang, Y. Zhou, and Y. Zhang, "Delay-based weighted proportional fair algorithm for lte downlink packet scheduling," *Wireless Personal Communications*, vol. 82, no. 3, pp. 1955-1965, 2015. <u>https://doi.org/10.1007/s11277-015-2324-3</u>
- [7] S. Radhakrishnan, S. Neduncheliyan, and K. Thyagharajan, "A review of downlink packet scheduling algorithms for real time traffic in LTE-advanced networks," *Indian Journal of Science and technology*, vol. 9, no. 4, pp. 1-5, 2016. <u>https://doi.org/10.17485/ijst/2016/ v9i4/84061</u>

- [8] U. Nwawelu, C. Ani, and M. Ahaneku, "Comparative analysis of the performance of resource allocation algorithms in long term evolution networks," *Nigerian Journal of Technology*, vol. 36, no. 1, pp. 163-171, 2017. <u>https://doi.org/10.4314/njt.v36i1.21</u>
- [9] 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 sur*veys & tutorials, vol. 15, no. 2, pp. 678-700, 2012. <u>https://doi.org/10.1109/SURV.2012.</u> 060912.00100
- [10] R. Sianipar, A. Fahmi, and R. Munadi, "Physical Resource Block (prb) Allocation Using Modified Round Robin Scheduling Algorithm In Downlink Lte Network," *eProceedings* of Engineering, vol. 5, no. 3, 2018.
- [11] C.-S. Lin and S.-I. Sou, "QoS-aware dynamic bandwidth reallocation with deadline assurance for multipath data offloading," *Computer Networks*, vol. 153, pp. 103-112, 2019. <u>https://doi.org/10.1016/j.comnet.2019.02.014</u>
- [12] F. Asadollahi and R. Dehdasht-Heydari, "Introduction of a novel hybrid weighted exponential logarithm-maximum throughput (HWEL-MT) scheduler for QoS improvement of LTE/4G cellular networks," *Wireless Personal Communications*, vol. 98, no. 1, pp. 91-104, 2018. <u>https://doi.org/10.1007/s11277-017-4857-0</u>
- [13] B. S. Monikandan, A. Sivasubramanian, S. Babu, G. Prasanna Venkatesan, and C. Arunachalaperumal, "Channel aware optimized proportional fair scheduler for LTE downlink," *Peer-to-Peer Networking and Applications*, vol. 13, no. 6, pp. 2135-2144, 2020. <u>https://doi.org/10.1007/s12083-019-00826-z</u>
- [14] A. Mamane, M. Fattah, M. El Ghazi, Y. Balboul, M. El Bekkali, and S. Mazer, "Proportional fair buffer scheduling algorithm for 5G enhanced mobile broadband," *International Journal of Electrical & Computer Engineering (2088-8708)*, vol. 11, no. 5, 2021. <u>https://doi.org/10.11591/ijece.v11i5.pp4165-4173</u>
- [15] M. Asvial, G. Dewandaru, and A. N. Rachman, "Modification of round robin and best CQI scheduling method for 3GPP LTE downlink," *International Journal of Technology*, vol. 6, no. 2, pp. 130-138, 2015. <u>https://doi.org/10.14716/ijtech.v6i2.964</u>
- [16] Y.-T. Mai and C.-C. Hu, "Design of dynamic resource allocation scheme for real-time traffic in the LTE network," *EURASIP Journal on Wireless Communications and Networking*, vol. 2022, no. 1, pp. 1-24, 2022. <u>https://doi.org/10.1186/s13638-022-02095-6</u>
- [17] N. K. Madi, Z. M. Hanapi, M. Othman, and S. K. Subramaniam, "Delay-based and QoSaware packet scheduling for RT and NRT multimedia services in LTE downlink systems," *EURASIP Journal on Wireless Communications and Networking*, vol. 2018, no. 1, pp. 1-21, 2018. <u>https://doi.org/10.1186/s13638-018-1185-3</u>
- [18] N. K. Madi, Z. B. M. Hanapi, M. Othman, and S. Subramaniam, "Two-level QoS-aware frame-based downlink resources allocation for RT/NRT services fairness in LTE networks," *Telecommunication Systems*, vol. 66, no. 3, pp. 357-375, 2017. <u>https://doi.org/ 10.1007/s11235-017-0289-0</u>
- [19] N. K. Madi, Z. M. Hanapi, M. Othman, and S. K. Subramaniam, "Energy-aware Subchannels Power Allocation for Downlink Transmissions in OFDMA Systems," in 2018 IEEE 4th International Symposium on Telecommunication Technologies (ISTT), 2018, pp. 1-6: IEEE. <u>https://doi.org/10.1109/ISTT.2018.8701723</u>
- [20] U. Madhow, Fundamentals of digital communication. Cambridge university press, 2008. <u>https://doi.org/10.1017/CB09780511807046</u>
- [21] 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, 2011. <u>https://doi.org/10.1109/TMM.2011.</u> 2152381

- [22] R. K. Jain, D.-M. W. Chiu, and W. R. Hawe, "A quantitative measure of fairness and discrimination," *Eastern Research Laboratory, Digital Equipment Corporation, Hudson, MA*, vol. 21, 1984.
- [23] B. Liu, H. Tian, and L. Xu, "An efficient downlink packet scheduling algorithm for real time traffics in LTE systems," in 2013 IEEE 10th Consumer Communications and Networking Conference (CCNC), 2013, pp. 364-369: IEEE. <u>https://doi.org/10.1109/CCNC. 2013.6488471</u>
- [24] G. T. 25.951, "Technical Specification Group Radio Access Network; FDD Base Station (BS) classification (3GPP TR 25.951 V16.0.0)," *ETSI*, pp. 1-63, 2020. ETSI.
- [25] ETSI TS 136 104 V8.10.0, "LTE; Evolved Universal Terrestrial Radio Access (E-UTRA); Base Station (BS) radio transmission and reception", *ETSI 3GPP TS 36.104 version* 8.10.0, 2010.
- [26] A. A. Masli, F. Y. Ahmed, and A. M. Mansoor, "QoS-Aware Scheduling Algorithm Enabling Video Services in LTE Networks," *Computers*, vol. 11, no. 5, p. 77, 2022. <u>https://doi.org/10.3390/computers11050077</u>
- [27] L. M. Tuan et al., "ITFDS: channel-aware integrated time and frequency-based downlink LTE scheduling in MANET," Sensors, vol. 20, no. 12, p. 3394, 2020. <u>https://doi.org/ 10.3390/s20123394</u>
- [28] P. E. Omiyi, M. M. Nasralla, I. U. Rehman, N. Khan, and M. G. Martini, "An intelligent fuzzy logic-based content and channel aware downlink scheduler for scalable video over OFDMA wireless systems," *Electronics*, vol. 9, no. 7, p. 1071, 2020. <u>https://doi.org/ 10.3390/electronics9071071</u>
- [29] R. Balakrishnan, K. Sankhe, V. S. Somayazulu, R. Vannithamby, and J. Sydir, "Deep reinforcement learning based traffic-and channel-aware OFDMA resource allocation," in 2019 IEEE Global Communications Conference (GLOBECOM), 2019, pp. 1-6: IEEE. https://doi.org/10.1109/GLOBECOM38437.2019.9014270
- [30] S. J. Kao, K. C. Hsiao, and F. M. Chang, "Radio resource allocation using genetic algorithm in heterogeneous networks," *International Journal of Communication Systems*, vol. 32, no. 15, p. e4089, 2019. <u>https://doi.org/10.1002/dac.4089</u>
- [31] N. M. Elshennawy, "Modified Proportional Fair Scheduling Algorithm for Heterogeneous LTE-A Networks," *International Journal of Interactive Mobile Technologies*, vol. 14, no. 10, 2020. https://doi.org/10.3991/ijim.v14i10.14389
- [32] H. Mahdi, B. Al-Bander, M. H. Alwan, M. S. Abood, and M. M. Hamdi, "Vehicular Networks Performance Evaluation Based on Downlink Scheduling Algorithms for High-Speed Long Term Evolution--Vehicle," *International Journal of Interactive Mobile Technologies*, vol. 15, no. 21, 2021. https://doi.org/10.3991/ijim.v15i21.22475
- [33] A. Mamane, M. Fattah, M. El Ghazi, and M. El Bekkali, "5G Enhanced Mobile Broadband (eMBB): Evaluation of Scheduling Algorithms Performances for Time-Division Duplex Mode," *nternational Journal of Interactive Mobile Technologies—iJIM*, vol. 16, no. 01, p. 121, 2022. <u>https://doi.org/10.3991/ijim.v16i01.25941</u>

8 Authors

Nadim K. M. Madi is currently an assistant professor with the department of Information Systems (DIS), CEMIS, in the University of Nizwa (UoN). He is the head of Information Systems Environment research cluster at DIS in UoN. He received his B.Sc. degree in Information Technology, major of Computer Networks from Universiti Utara Malaysia, in 2010. He finished the MSc degree of Computer Science, ma-

joring of Distributed Computing from University Putra Malaysia, in 2012. He has been awarded the MS.c. in First-Class with honors. In 2018, He accomplished the Ph.D. in field of Radio Resource Management and QoS enhancements in mobile networks from the same institution. His research experience started in early 2011. His research themes include IoT technologies, emerging and ubiquitous mobile systems, mobile radio resource management, network QoS performance analysis, network modeling, and algorithm analysis. Dr. Nadim is an active researcher with a recognized record of publication in many reputed international journals. He has been awarded the best conference paper presentation. Also, he is actively involved as a reviewer for some merit international peer-reviewed journals from IEEE, and Springer. His publication contributed to several fields such as IoT technologies, Emerging mobile systems, radio resource management, Quality of Service analysis, Software-defined networks, and cloud computing security.

Alaa Omer Najim is currently an assistant lecturer with the department of Survey Engineering in Technical college of Kirkuk. She is the coordinator of postgraduate programs and the officer of public relations and scholarships at the same institute. She received her B.Sc. degree in software engineering from technical college of Kirkuk in 2004. She finished the MSc degree of computer science and information technology, majoring of computer network from University Putra Malaysia, in 2018. Her research experience started in early 2015. Her research themes include mobile radio resource management, network QoS performance analysis, network modeling and algorithm analysis.

Article submitted 2022-10-08. Resubmitted 2022-12-30. Final acceptance 2023-02-06. Final version published as submitted by the authors.