Efficient Allocation of LTE Downlink Spectral Resource to Improve Fairness and Throughput

You-Chiun Wang and Dai-Rong Jhong

Abstract—For the current generation of cellular communication systems, *long term evolution (LTE)* has been the major protocol to support high-speed data transmission. It is critical to allocate downlink spectral resource in LTE, namely *resource blocks (RBs)*, but the issue is not well addressed in the standard. Therefore, the paper develops an efficient RB allocation algorithm with four mechanisms to improve both fairness and throughput in LTE. For fairness concern, our RB allocation algorithm employs a resource-reservation mechanism to prevent cell-edge *user equipments (UEs)* from starvation, and a credit-driven mechanism to keep track of the amount of resource given to each UE. For throughput concern, it adopts both weight-assignment and RB-matching mechanisms to allocate each RB to a packet according to its flow type and length. Through simulations, we demonstrate that the proposed RB allocation algorithm can significantly increase both throughput and fairness while reducing packet dropping and delays of real-time flows, as compared with previous methods.

Index Terms—downlink transmission, long term evolution (LTE), network throughput, resource allocation, system fairness.

1 Introduction

NOWADAYS, mobile phones are in widespread use for people to enjoy wireless service anytime, anywhere. Moreover, various large-demand downlink applications, for example, video downloads and multimedia streaming, have been dominating network traffic in the Internet [1]. Consequently, the 3rd Generation Partnership Project (3GPP) keeps working out the standard of *long term evolution (LTE)* to support high-speed wireless access for the current (and next) generation of communication systems.

In the downlink communication, LTE adopts orthogonal frequency division multiple access (OFDMA), which realizes data transmission by assigning subsets of subcarriers to individual receivers. A *resource block (RB)* is the basic unit to allocate the spectral resource to each *user equipment (UE)*. Each RB is able to carry different amount of information through a different modulation and coding scheme. How to allocate RBs to UEs according to network condition and traffic demands is called the *LTE downlink scheduling problem*. This problem substantially affects system performance and user experience. Nevertheless, 3GPP does not cope with the problem but leaves it to LTE implementers.

There have been several classic methods used to solve the LTE downlink scheduling problem. Specifically, the max-CQI method [2] takes a greedy policy by assigning each RB to the UE with the best *channel quality indication (CQI)*, which is an indicator of the current channel condition. It improves the overall throughput but may starve the UEs encountering worse channel quality. For fairness concern, the *proportional fair (PF)* method [3] uses a criterion $P_i = r_i/r_i^{\text{avg}}$ to determine RB allocation, where r_i and r_i^{avg} are the current and past data rates of each UE, respectively. However, PF ignores the delay constraint of packets. Thus, the *modified largest weighted delay first (M-LWDF)* method [4] applies a weight w_i and the head-of-line (HOL) packet delay d_i to PF, where it picks the UE with the largest value of $(w_i \cdot d_i \cdot P_i)$ to receive each RB. In addition, the *exponential proportional fair (EXP/PF)* method [5] further enhances M-LWDF by adding the average HOL packet delay d_{RT} of all real-time flows. Specifically, it selects for each RB the UE with the highest value of $(e^{\sigma} \cdot P_i)$, where e is Euler's number and $\sigma = (w_i d_i - d_{\text{RT}})/(1 + \sqrt{d_{\text{RT}}})$.

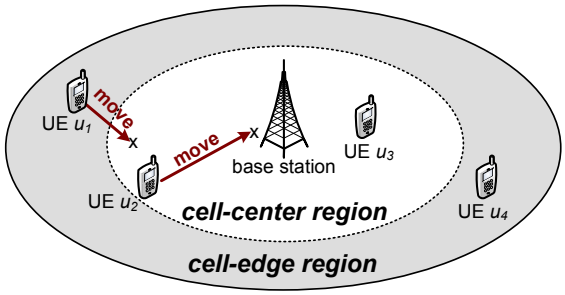


Fig. 1: Two examples to exhibit unfair transmission by the PF method and its variations, where '×' indicates the stopping positions of UEs u_1 and u_2.

In the above methods, a number of issues are arisen. First of all, many methods are enhanced from PF in essence, which take sides with the UEs whose channel quality improves (in other words, P_i increases). Nevertheless, PF does not have a *global view* to adaptively adjust the amount of resource allocated to UEs in order to maintain fairness. Fig. 1 shows an example, where both UEs u_1 and u_2 stay in their original positions for a while and then move toward the *base station (BS)* such that $P_1 < P_2$. In this case, the PF method and its variations will prefer giving resource to u_2, which consequently increases the difference between the amount of data transmission by u_1 and u_2. Thus, the network becomes more unfair. In practice, when the channel quality of both u_1 and u_2 improves, we should give more resource to u_1 to improve fairness, because u_1 received less data than u_2 did in the past. Second, these methods do not give special treatment for *cell-edge UEs*, and thus such UEs may not acquire sufficient resource due to bad channel quality (i.e., starvation), thereby further hurting system fairness. Fig. 1 presents an example, where UEs u_3 and u_4 do not move for a long time. In this case,

The authors are with the Department of Computer Science and Engineering, National Sun Yat-sen University, Kaohsiung, 80424, Taiwan. E-mail: ycwang@cse.nsysu.edu.tw; m023040024@student.nsysu.edu.tw

we have $P_3 \approx P_4 \approx 1$ when the channel condition remains static. Therefore, the PF-based methods will treat both u_3 and u_4 as no difference according to the criterion P_i. In fact, u_4 (i.e., a cell-edge UE) will receive much less data than u_3 (i.e., a cell-center UE) does. Third, some methods such as max-CQI do not *differentiate* flows according to their types. Thus, they would not well support quality of service (QoS) for real-time service. Fourth, no method addresses the *size-relationship* between packets and RBs. In particular, when an RB with a more complex modulation (i.e., more capacity) is used to transmit a short-length packet, the RB is actually wasted.

Based on these discussions, this paper develops an efficient RB allocation algorithm with the objectives of improving both system fairness and network throughput in LTE. It contains four core mechanisms as follows:

- **Resource-reservation mechanism:** To prevent the UEs located in the cell-edge region from starvation, the BS reserves a dynamic portion of spectral resource for them in advance.
- **Credit-driven mechanism:** To provide fair transmission, each UE is associated with a credit value to adaptively control the amount of resource that it can acquire in a TTI.
- **Weight-assignment mechanism:** To support QoS, we differentiate real-time flows from non-real-time ones through different weights (by referring to their HOL packet delays).
- **RB-matching mechanism:** To save RBs' capacity, we match each RB with the HOL packet of a flow according to the the number of bits carried by the RB and the length of the packet.

Our contribution is to develop the RB allocation algorithm which takes care of cell-edge UEs and employs the credit value to keep track of the amount of resource allocated to UEs, so as to provide more fair transmission. Moreover, our RB allocation algorithm can increase network throughput while supporting QoS for real-time flows, with the help of both weight-assignment and RB-matching mechanisms. We show that the proposed RB allocation algorithm incurs less computation and memory complexity, which helps the BS fast determine RB assignment to all flows in a short scheduling period. Furthermore, through simulation, the experimental results demonstrate that our RB allocation algorithm can significantly outperform the aforementioned methods.

We organize the remainder of this paper as follows. Section 2 gives a brief survey of LTE and the problem definition. Section 3 discusses existing work related to LTE downlink resource scheduling. In Section 4, we present our RB allocation algorithm and analyze its computation time and memory consumption. Then, Section 5 evaluates system performance by simulations. Afterwards, we make a conclusion and discuss future work in Section 6.

2 LTE Survey and Its Downlink Scheduling Problem

In LTE, the management of spectral resource is achieved on the basis of each cell. We thus focus the discussion on an LTE macro-cell coordinated by one BS. The spectral resource is materialized by a two-dimensional array of RBs in both time and frequency domains; specifically, each RB has 0.5 ms duration and 180 KHz bandwidth. According to the bandwidth of a downlink channel, the BS can provide various numbers of RBs in a scheduling period, namely *transmission time interval (TTI)*. The duration of each TTI is 1 ms, so it contains two columns of RBs. LTE supports six types of channels whose bandwidths are 1.4, 3, 5, 10, 15, and 20 MHz, where the BS can allocate at most 6, 15, 25, 50, 75, and 100 RBs in each TTI, respectively[1]. When the communication techniques of single-input single-output (SISO) or single-user multiple-input and multiple-output (SU-MIMO) are applied, RBs are viewed as exclusive resource. It means that the BS is not allowed to assign the same RB to multiple UEs in a TTI.

TABLE 1: LTE CQI table for three modulations: QPSK, 16QAM, and 64QAM, where the value of code rate is multiplied by 1024 and the number of bits carried is an average value.

index	modulation	code rate	efficiency	bits carried
0	out of range			
1	QPSK	78	0.1523	12.79
2	QPSK	120	0.2344	19.69
3	QPSK	193	0.3770	31.67
4	QPSK	308	0.6016	50.53
5	QPSK	449	0.8770	73.67
6	QPSK	602	1.1758	98.77
7	16QAM	378	1.4766	124.03
8	16QAM	490	1.9141	160.78
9	16QAM	616	2.4063	202.13
10	64QAM	466	2.7305	229.36
11	64QAM	567	3.3223	279.07
12	64QAM	666	3.9023	327.79
13	64QAM	772	4.5234	379.97
14	64QAM	873	5.1152	429.68
15	64QAM	948	5.5547	466.59

To update the information of channel condition, each UE routinely reports it CQI evaluation to the BS in every TTI. Table 1 gives available CQIs defined in the LTE standard [6]. In particular, a larger CQI index implies that the UE has better channel condition. The BS then selects the modulation and coding scheme used to transmit the UE's downlink data by referring to its CQI. According to Table 1, LTE supports three types of modulation (from simple to complex), each with different coding rates: QPSK (quadrature phase-shift keying, where CQI ≤ 6), 16QAM (quadrature amplitude modulation, where $7 \leq$ CQI ≤ 9), and 64QAM (where CQI ≥ 10). Each RB is able to carry more data bits when it employs more complex modulation (and a higher code rate). However, this fact relies on the good channel condition of the corresponding UE. In other words, when a UE suffers from bad channel quality (e.g., a cell-edge UE), it can use only simple modulation for communication.

Given each UE's CQI and its amount of downlink traffic in a TTI, the LTE downlink scheduling problem determines how to efficiently distribute RBs among UEs to satisfy its transmission requirement, under the SISO/SU-MIMO assumption (i.e., RBs are non-sharable). Theoretically, only when the BS has sufficient resource can we find a feasible solution to the LTE downlink scheduling problem. When this condition cannot be met, our objectives are to increase network throughput, improve system fairness, and support QoS for real-time flows. In particular, we employ Jain's fairness index [7] to evaluate

1. The advanced version of LTE, LTE-A, can integrate multiple channels to obtain much larger bandwidth (up to 100 MHz) through the technique of *carrier aggregation*. It involves the selection of different channels for communication, whose issue is out of the paper's scope. Instead, we aim at resource allocation in one single channel.

the fairness degree as follows:

$$\mathcal{I} = \frac{\left(\sum_{i=1}^{n} \tilde{r}_i\right)^2}{n \sum_{i=1}^{n} \tilde{r}_i^2}, \quad (1)$$

where $\tilde{r}_i$ is the normalized throughput of a flow and n is the number of total flows in the cell. According to Eq. (1), we have $0 < \mathcal{I} \leq 1$, and a larger index means that the BS achieves more fair transmission. Moreover, to evaluate the degree of QoS support for real-time flows, we measure both the dropping ratio and average delay of real-time packets.

3 Related Work

In the literature, many methods have been proposed to solve the LTE downlink scheduling problem, which can be classified into two categories. One aims at supporting QoS for real-time/multimedia flows. The other considers providing fair transmission among flows.

3.1 LTE Scheduling with QoS Consideration

To support QoS, a number of studies try to reduce packet dropping of real-time flows. In particular, [8] proposes a virtual queue to forecast the coming packets, and discards the packets from the queue that will certainly miss their deadlines. Then, it adopts the max-CQI strategy to allocate RBs among UEs. In [9], flows are divided into urgent and non-urgent. Urgent flows are assigned with a high priority to acquire resource first in order to alleviate their packet discarding. Non-urgent flows, including non-real-time flows and real-time flows whose deadlines are not expired yet, have a low priority to get residual resource. Samia et al. [10] use a cooperative game to model the scheduling problem, and adopt the Nucleolus solution [11] to minimize dissatisfaction with the allocation of resource that real-time flows has received. However, these studies do not address the fairness issue.

Our previous work [12] computes the amount of resource given to each UE by max-CQI, and asks non-urgent flows to return a fraction of their allocated RBs by a taxing mechanism. These RBs are then reassigned to the flows being threatened by packet dropping. There are two major differences between the taxing mechanism in [12] and the resource-reservation mechanism in this paper. First, the taxing mechanism is reactive, which means that it will be done only after RB allocation, while the resource-reservation mechanism is proactive, as it reserves a portion of resource in advance before RB allocation. Second, the taxing mechanism aims at reducing packet dropping of real-time flows, while the resource-reservation mechanism avoids starving cell-edge UEs. Moreover, we adopt the credit-driven mechanism to provide fair transmission, which is not addressed in [12]. These features significantly distinguish this paper from [12].

Some work focuses on multimedia transmission in LTE networks. Specifically, [13] discusses how to deliver video streaming in a smooth manner. The BS refers to multiple parameters of each video flow such as data rate, delay limitation, and signal distortion to determine its resource allocation and video coding. The work [14] proposes a double-layer scheduling framework for LTE multimedia communication. One layer calculates the amount of information that a multimedia flow has to send during each period to meet its delay constraint. Afterwards, the other layer allocates RBs to every flow accordingly through the PF strategy. Liu and Chen [15] propose a downlink scheduling method that considers not only packet dropping of video flows but also service degradation due to hard handoff. To do so, the quota of video data is decided by the transmission deadlines of the corresponding packets, and then the BS allocates RBs for sending out these data. Apparently, these research efforts have different objectives with our work.

3.2 LTE Scheduling with Fairness Consideration

A few studies convert the LTE downlink scheduling problem to other theoretical problems to provide fair transmission. For example, [16] indicates that LTE resource allocation is usually modeled as a nonlinear problem, and shows that it can be converted to the linear integer programming model. Iturralde et al. [17] transform the scheduling problem into a bankruptcy-game problem, and apply the Shapley value [18] to allocate resource to flows. Huang et al. [19] apply the Nash Bargaining solution to the scheduling problem, so as to make the result of resource allocation become Pareto-optimal [20]. However, these studies involve relatively complex calculation in resource allocation, which may be applied to only small-scale networks (in particular, their performance evaluation is conducted in a small network with just 30 to 60 UEs). In contrast to them, our work employs a simple credit-driven mechanism to maintain fair transmission among flows, which helps the BS fast compute resource allocation in a short TTI.

On the other hand, [21] adopts the α-fair utility function [22] to deal out resource by referring to each UE's average throughput and a parameter α. Ali et al. [23] also use a utility function to calculate the satisfaction degree of each flow, and make flows compete for resource by their utility values. The work [24] combines the PF method with the earliest-deadline-first (EDF) approach [25], which always selects the packet with the most urgent deadline to transmit. In this way, the hybrid scheme tries to exploit both PF's fairness property and EDF's bounded-delay feature. In [26], UEs are divided into three groups, namely 'very good', 'average', and 'poor', depending on their CQI reports. Assuming that the distribution of UEs in each group is almost equal, [26] picks a fixed portion of UEs in each group and allocates RBs to them, with the goals of maximizing throughput and increasing fairness. However, this assumption may not be necessarily valid in practical scenarios.

Distinguishing from the above work, we seek to improve fairness by not only taking care of cell-edge UEs but also adopting a credit idea. Moreover, through weight-assignment and RB-matching mechanisms, our proposed RB allocation algorithm can increase throughput while alleviating packet dropping and delay of real-time service. Experimental results in Section 5 will also verify its effectiveness.

4 The Proposed Algorithm for RB Allocation

Our RB allocation algorithm is composed of four core mechanisms to manage downlink spectral resource in LTE. Below, we first present the detailed design of each mechanism, and then discuss how our RB allocation algorithm integrates these mechanisms, followed by some discussions on the proposed algorithm.

4.1 Resource-reservation Mechanism

When a UE stays in the cell-edge region (as shown in Fig. 1), its channel quality may become worse because of serious path loss or signal interference caused from the neighboring cells. These UEs are called *cell-edge UEs* and they may be also urgent for spectral resource. Unfortunately, many methods such as max-CQI disfavor cell-edge UEs, or treat them as no difference comparing with other UEs (for example, the PF method does not differentiate between UEs u_3 and u_4 in Fig. 1). The cell-edge UEs will inevitably lose the resource competition, thereby resulting in starvation. To overcome the above situation, this mechanism suggests reserving a small, dynamic portion of resource in advance to be allocated to only cell-edge UEs.

Let α_i be the traffic demand of a UE u_i in the current TTI. In addition, we denote by $\mathcal{U}$ and $\mathcal{U}_E$ the set of all UEs and the set of cell-edge UEs, respectively. Then, the BS needs to reserve a number of ξ RBs to the UEs in $\mathcal{U}_E$ as follows:

$$\xi = \left\lceil \min \left\{ \frac{\sum_{u_i \in \mathcal{U}_E} \alpha_i}{\sum_{u_i \in \mathcal{U}} \alpha_i}, \beta \right\} \times m \right\rceil, \tag{2}$$

where m is the number of available RBs in a TTI. According to Eq. (2), ξ is proportional to the traffic demands of all cell-edge UEs. However, because cell-edge UEs can still compete with others for the $(m - \xi)$ unreserved RBs, it is not a good idea to reserve a large amount of resource for them, otherwise the overall throughput will significantly decrease. That is why we add a small threshold β in Eq. (2) to restrict the number of reserved RBs for cell-edge UEs, where $0 < \beta \leq 0.1$.

In the above mechanism, one question is how to classify UEs. A naive solution is to use the Euclidean distance between each UE and the BS as a reference. Once the distance is larger than a predefined threshold, we add the UE to $\mathcal{U}_E$. However, this solution has two shortcomings. First, the BS has to continually keep track of the location of every UE, which complicates the design. Second, finding the distance threshold is not an easy job. Consequently, we propose a simple but practical solution by using CQI. In particular, when a UE reports a CQI index less than seven, it is viewed as a cell-edge UE. Our solution adds almost no overhead to the original proposal of LTE, because every UE has to periodically report its CQI to the BS following the LTE specification. Moreover, according to Table 1, the BS should choose the simplest but the most robust modulation (i.e., QPSK) to transmit data when CQI ≤ 6. In this case, there is a high possibility that the UE stays in the cell-edge region. Lemmas 1 and 2 give analysis on the amount of computation time and memory required by the resource-reservation mechanism, respectively.

Lemma 1. Given N UEs, the computation complexity of the resource-reservation mechanism is $O(2N)$.

Proof: In the resource-reservation mechanism, the BS has to find UEs in $\mathcal{U}_E$ first. By using the CQI method, it becomes straightforward to classify UEs by checking whether a UE's CQI is larger than seven or not. Since each UE has to report its CQI measurement in a TTI (referring to Section 2) and the BS will check each UE once, it thus takes $O(N)$ time to classify UEs. Then, it also spends $O(N)$ time to calculate ξ by Eq. (2), as we have to calculate the sum of traffic demands of UEs in $\mathcal{U}_E$ and $\mathcal{U}$. Therefore, the overall complexity of the resource-reservation mechanism will be $O(2N)$. □

Lemma 2. Given N UEs, the amount of memory required by the resource-reservation mechanism is $O(N)$.

Proof: It is clear that the resource-reservation mechanism will divide all UEs into two disjointed sets $\mathcal{U}_E$ and $(\mathcal{U} - \mathcal{U}_E)$. Thus, the BS should maintain two lists of UEs accordingly. In this case, the amount of memory required by the mechanism will be obviously $O(N)$. □

4.2 Credit-driven Mechanism

Conventional *weighted fair queuing* can well support fair transmission among flows from a global view [27]. Its idea is to keep track of the (weighted) difference between the amount of data transmission of flows, and seek to minimize the difference. Inspired by this idea, we expect that each UE u_i will obtain a constant amount τ_i of resource in every TTI, where τ_i depends on u_i's traffic demand. In theory, if all UEs each exactly receives $k\tau_i$ amount of downlink data during k TTIs, for any $k \geq 1$, the transmission is said to be fair; in other words, we have $\mathcal{I} = 1$ in Eq. (1). Nevertheless, because the channel condition will vary and flows may have different lengths of packets, it is infeasible to ask the BS to transmit exact τ_i amount of data to each UE u_i in every TTI. Therefore, we use a variable A_i to record the *accumulative difference* between the amount of downlink data actually received by each UE u_i and the amount of resource that it expects to obtain. Specifically, let $r_{i,k}$ be the amount of downlink data transmitted to u_i in the kth TTI. Then, we can calculate the accumulative difference as follows:

$$A_i = \sum_k r_{i,k} - \tau_i. \tag{3}$$

From Eq. (3), $A_i > 0$ indicates that UE u_i consumes more resource than expectation, so the BS should give its resource to other UEs in order to maintain system fairness. On the contrary, $A_i < 0$ implies that u_i does not receive sufficient data, so it is better to allocate more resource to u_i in the next TTI. In case of $D_i = 0$, it means that u_i has gotten expected resource in the current TTI.

However, the variation of A_i could be quite large, especially when some UEs have much better channel quality but others do not. Therefore, by taking the minimum and maximum values of accumulative differences of all UEs (respectively denoted by $A_{\min}$ and $A_{\max}$), we can convert A_i to a *normalized credit*:

$$\hat{\mathbf{C}}_i = 2 - \frac{A_i - A_{\min}}{A_{\max} - A_{\min}}. \tag{4}$$

In this way, we can restrict the credit value $\hat{\mathbf{C}}_i$ between 1 and 2. In particular, a larger credit value implies that UE u_i has a higher priority to acquire resource, because it receives less data than expectation, and vice versa.

We present an example with three UEs, where τ_i = 40Kb/TTI for $i = 1..3$. In a TTI, suppose that UEs u_1, u_2, and u_3 actually receive an amount of 40, 60, and 20 Kb downlink data, respectively. Then, we can derive that $A_1 = 0$, $A_2 = 20$, and $A_3 = -20$. Because $A_{\min} = -20$ and $A_{\max} = 20$, the normalized credits will be $\hat{\mathbf{C}}_1 = 1.5$, $\hat{\mathbf{C}}_2 = 1$, and $\hat{\mathbf{C}}_3 = 2$. In this case, these three UEs have priorities of $u_3 > u_1 > u_2$ to acquire the network resource. Consequently, UEs u_3 and u_2 will be given more and less resource in the next TTI, respectively, so as to maintain their fairness. In Lemmas 3 and 4, we analyze computation time and memory consumption of the credit-driven mechanism, respectively.

Lemma 3. Given N UEs, the computation complexity of the credit-driven mechanism is $O(3N)$.

TABLE 2: QCI table defined in LTE (for real-time flows).

QCI	delay budget	loss rate	representative applications
1	100 ms	10^{-2}	conversational voice (e.g., VoIP)
2	150 ms	10^{-3}	conversational video (e.g., live streaming)
3	50 ms	10^{-3}	real-time gaming
4	300 ms	10^{-6}	non-conversational video (e.g, buffered streaming)
65	75 ms	10^{-2}	mission critical user plane push to talk voice (MCPTT)
66	100 ms	10^{-2}	non-MCPTT

Proof: Given $r_{i,k}$ in a TTI, the BS has to update the accumulative difference for each UE by Eq. (3), which spends $O(N)$ computation time. Moreover, to convert the accumulative difference A_i to the normalized credit $\hat{\mathbf{C}}_i$, we has to find the values of both $A_{\min}$ and $A_{\max}$. This operation consumes $O(N)$ time because we need to search all A_i values once. Beside, it takes $O(N)$ time to do the above conversion. Therefore, the overall time complexity of the credit-driven mechanism is $O(N) + O(N) + O(N) = O(3N)$. □

Lemma 4. Given N UEs, the amount of memory used by the credit-driven mechanism is $O(2N)$.

Proof: In the credit-driven mechanism, the BS has to keep track of the accumulative difference A_i for each UE and convert it to the corresponding credit $\hat{\mathbf{C}}_i$. These variables are reused in every TTI. Thus, the amount of memory consumed by the credit-driven mechanism will be $O(2N)$. □

4.3 Weight-assignment Mechanism

Generally speaking, each UE can possess multiple flows that share the resource acquired by the UE. However, real-time flows are usually characterized by strict delay requirement, and we should reduce their packet dropping due to exceeding deadlines. Specifically, for each real-time flow $f_{i,j}$, we expect that

$$\text{Prob}\{D_{i,j} > \delta_{i,j}\} \leq p_{i,j}, \tag{5}$$

where $D_{i,j}$ is a random variable denoting the steady-state packet delay of flow $f_{i,j}$, and $\delta_{i,j}$ is the delay threshold (i.e., deadline) of $f_{i,j}$'s packets. In Eq. (5), $p_{i,j}$ is the maximum *tolerable* probability of packet loss due to out of deadline, which depends on the application's requirement. In practice, LTE has specified *QoS class identifier (QCI)* for different types of flows in order to support QoS, which defines both 'delay budget' and 'loss rate' of the packets of each flow. Table 2 presents QCIs of real-time flows. Here, the delay budget limits the packet deadline while the loss rate gives a suggestion for the maximum tolerable probability. In other words, both parameters $\delta_{i,j}$ and $p_{i,j}$ are in fact constants for each real-time flow, and their values can be easily determined in advance by referring to Table 2.

To take the effect of $\delta_{i,j}$ and $p_{i,j}$ into consideration, we modify the idea of M-LWDF to set a *weight* for each real-time flow $f_{i,j}$ as follows:

$$\hat{\mathbf{W}}_{i,j} = -\log p_{i,j} \times \frac{d_{i,j}}{\delta_{i,j}}, \tag{6}$$

where $d_{i,j}$ is the delay of flow $f_{i,j}$'s HOL packet. In particular, if a real-time flow cannot tolerate a high packet loss rate (in other words, it will have a smaller $p_{i,j}$ value), the BS will assign it with a larger weight because of the effect of $(-\log p_{i,j})$, and vice versa. Besides, according to Eq. (6), when the HOL packet of a real-time flow is on the point of expiring (i.e., a larger $d_{i,j}$ value) or the flow has a more stringent delay constraint (i.e., a smaller $\delta_{i,j}$ value), it will be also given with a larger weight for transmission. We then discuss both time complexity and memory usage of the weight-assignment mechanism in Lemmas 5 and 6, respectively.

Lemma 5. The computation complexity of the weight-assignment mechanism is $O(n_R)$, where n_R is the number of real-time flows.

Proof: The weight-assignment mechanism involves in only the weight calculation of each real-time flow by Eq. (6). As mentioned earlier, both parameters $p_{i,j}$ and $\delta_{i,j}$ are fixed for a real-time flow and they can be determined by Table 2. Therefore, we can keep a small table to record the value of $(-\log p_{i,j}/\delta_{i,j})$ for each real-time flow in advance. In this way, Eq. (6) requires just one simple multiplication. Because there are n_R real-time flows, it thus takes $O(n_R)$ time to conduct the weight-assignment mechanism. □

Lemma 6. The amount of memory consumed by the weight-assignment mechanism is $O(2n_R)$, where n_R is the number of real-time flows.

Proof: As discussed in Lemma 5, we use a table to record the value of $(-\log p_{i,j}/\delta_{i,j})$ for each real-time flow, which requires the amount of $O(n_R)$ memory. In addition, for each real-time flow, we have to use a variable $\mathbf{W}_{i,j}$ to store its weight. Thus, the overall memory consumption of the weight-assignment mechanism is $O(2n_R)$. □

4.4 RB-matching Mechanism

Most LTE scheduling methods assign an RB to each flow according to the flow's channel condition, data rate, or HOL packet delay. However, none of them considers whether the RB is really fit for the transmitting packet by its length. For example, when a large-capacity RB is selected to transmit a short-length packet, the RB is apparently wasted. To address this issue, we compute a *fitness degree* $\hat{\mathbf{F}}_{i,j}$ for each flow $f_{i,j}$ in the RB-matching mechanism. In particular, a larger $\hat{\mathbf{F}}_{i,j}$ degree implies that the RB is more suitable to transmit the flow's packet.

Let $l_{i,j}$ denote the length of flow $f_{i,j}$'s HOL packet, and q_i be the capacity of an RB in respect to its UE u_i. In addition, we define $Q_{64\text{QAM}}$ and $Q_{16\text{QAM}}$ to be the minimum capacity of an RB with the 64QAM and 16QAM modulation, respectively. Then, Fig. 2 illustrates the three cases to determine the degree $\hat{\mathbf{F}}_{i,j}$:

1) Case of $q_i \geq Q_{64\text{QAM}}$:
In this case, the RB has relatively large capacity. Therefore, we prefer assigning it to a long-length packet in order to reduce potential wastage. In consequence, we define the fitness degree of flow $f_{i,j}$ by

$$\hat{\mathbf{F}}_{i,j} = \min\{l_{i,j}/q_i, 1\}. \tag{7}$$

According to Eq. (7), the fitness degree depends on the length of flow $f_{i,j}$'s HOL packet. However, when

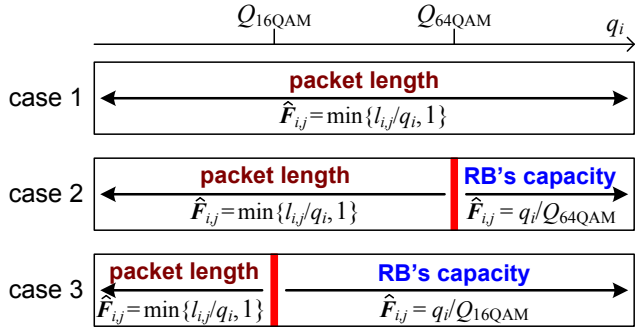


Fig. 2: The concept of RB-matching mechanism.

the HOL packets of multiple flows each has a length larger than the RB's capacity q_i, it does not matter to choose which flow, because we need more than one RB to finish sending out each of these packets. Therefore, we select the minimum value between $l_{i,j}/q_i$ and one in Eq. (7).

2) Case of $Q_{16\text{QAM}} \leq q_i < Q_{64\text{QAM}}$:
We further consider two subcases. In the subcase of $l_{i,j} \geq Q_{64\text{QAM}}$, the packet's length obviously exceeds the RB's capacity. Therefore, the fitness degree should depend on the RB's capacity to allow the RB to send out more data bits of the packet. In the subcase of $l_{i,j} < Q_{64\text{QAM}}$, the fitness degree will depend on the length of the transmitting packet, so as to reduce the wastage in RB's capacity. To sum up, we suggest setting

$$\hat{\mathbf{F}}_{i,j} = \begin{cases} q_i/Q_{64\text{QAM}}, & \text{if } l_{i,j} \geq Q_{64\text{QAM}}, \\ \min\{l_{i,j}/q_i, 1\}, & \text{if } l_{i,j} < Q_{64\text{QAM}}. \end{cases} \quad (8)$$

3) Case of $q_i < Q_{16\text{QAM}}$:
Based on the similar reasons of the previous case, we also suggest setting the fitness degree of flow $f_{i,j}$ as follows:

$$\hat{\mathbf{F}}_{i,j} = \begin{cases} q_i/Q_{16\text{QAM}}, & \text{if } l_{i,j} \geq Q_{16\text{QAM}}, \\ \min\{l_{i,j}/q_i, 1\}, & \text{if } l_{i,j} < Q_{16\text{QAM}}. \end{cases} \quad (9)$$

It is worth noting that the fitness degree $\hat{\mathbf{F}}_{i,j}$ must be ranged within $(0, 1]$ according to Eqs. (7), (8), and (9). We then analyze the amount of computation time and memory usage of the RB-matching mechanism in Lemmas 7 and 8, respectively.

Lemma 7. Let m and n be the number of RBs and flows, respectively. Then, the computation complexity of the RB-matching mechanism is $O(mn)$.

Proof: In the RB-assignment mechanism, the BS checks only the HOL packet for each flow. Consequently, we need to check at most $O(n)$ packets. On the other hand, for each packet, we will use the three exclusive cases to find the corresponding fitness degree related to each RB. Since there are m RBs, the computation complexity will thus be $O(mn)$. □

Lemma 8. Given m RBs and n flows, the amount of memory spent by the RB-matching mechanism is $O(mn)$.

Proof: The objective of the RB-matching mechanism is to compute the fitness degree $\hat{\mathbf{F}}_{i,j}$ for each pair of flow and RB. In other words, the only information that this mechanism has to record is $\hat{\mathbf{F}}_{i,j}$. Since we have m RBs and n flows, the amount of memory spent by the RB-matching mechanism will be $O(mn)$. □

4.5 Algorithm Design and Discussion

Our proposed RB allocation algorithm works based on the aforementioned four mechanisms. Specifically, it repeats the three steps in each TTI to help the BS distribute resource among flows as follows:

- **[Step 1: Discard overdue packets]**
We follow the similar idea in [12] to alleviate unnecessary data transmission in advance. In particular, for each flow, the BS checks whether it has any overdue packet, which will become invalid even though it can be transmitted to the corresponding UE now. Here, the HOL packet of a flow $f_{i,j}$ (belonging to UE u_i) is considered as *overdue* if

$$d_{i,j} + \psi(h_{i,j}) > \delta_{i,j}, \quad (10)$$

where $\psi(h_{i,j})$ denotes the propagation latency required by the physical layer to finish transmitting the HOL packet $h_{i,j}$ to UE u_i, and $\delta_{i,j}$ represents the delay tolerant time of flow $f_{i,j}$. In this case, the HOL packet needs to be discarded, and the BS checks the next packet again in the queue according to Eq. (10), until the new HOL packet is not overdue or the queue becomes empty.
- **[Step 2: Reserve resource for cell-edge UEs]**
With the help of the resource-reservation mechanism, the BS keeps ξ RBs to be allocated to the cell-edge UEs in $\mathcal{U}_E$. Afterwards, all UEs in $\mathcal{U}$ are allowed to compete for $(m - \xi)$ unreserved RBs.
- **[Step 3: Allocate RBs to flows]**
Based on the credit-driven, weight-assignment, and RB-matching mechanisms, for each RB, we calculate a bidding value $b_{i,j}$ for every flow $f_{i,j}$. The flow with the largest bidding value can be allocated with that RB. In particular, if $f_{i,j}$ belongs to a cell-edge UE or it is a non-real-time flow of a cell-center UE, then we set its bidding value as follows:

$$b_{i,j} = \hat{\mathbf{C}}_i \times \hat{\mathbf{F}}_{i,j} \times r_i. \quad (11)$$

Otherwise, $f_{i,j}$ must be a real-time flow of a cell-center UE. In this case, we set its bidding value as follows:

$$b_{i,j} = \hat{\mathbf{C}}_i \times \hat{\mathbf{F}}_{i,j} \times \hat{\mathbf{W}}_{i,j} \times r_i. \quad (12)$$

We then discuss the design rationale of our RB allocation algorithm. First of all, the BS removes those packets that inevitably miss deadlines beforehand. Thus, it can avoid wasting network bandwidth on transmitting overdue (and useless) packets. Second, the BS spends extra ξ RBs to allow cell-edge UEs to have an opportunity to transmit their packets. According to Eq. (2), we have $\xi \leq 0.1m$, so the above reservation will not significantly degrade the overall throughput. Moreover, cell-edge UEs can also compete for unreserved RBs. This is to deal with the two situations when 1) most UEs locate in the cell-edge region, or 2) the cell-edge UEs have large amount of traffic requirements. Third, we apply the weight-assignment mechanism (i.e., $\hat{\mathbf{W}}_{i,j}$) only to the real-time flows of cell-center UEs. Here, because cell-edge UEs can use only the simplest modulation (i.e., QPSK) for transmission, it would not have obvious impact to differentiate flows by their types. Thus, we do not add $\hat{\mathbf{W}}_{i,j}$ to their bidding values in Eq. (11). Finally, the credit value $\hat{\mathbf{C}}_i$ provides a global view for the BS to easily identify those UEs that do not obtain sufficient resource. In this

way, the BS can provide more fair transmission as comparing with the traditional PF policy that uses a local metric r_i/r_i^{avg}.

Our four mechanisms are easy to implement in practice, and they consider practical constraints of LTE. Specifically, for the resource-reservation mechanism, since each UE has to report its CQI measurement to the BS in every TTI (according to the LTE specification), it becomes much simple to classify UEs into cell-edge and cell-center groups by just checking their CQI values. Then, for the credit-driven mechanism, because we deal with the downlink traffic, the BS must have the knowledge of the amount of data transmitted to u_i in a TTI (i.e., $r_{i,k}$). Also, τ_i is a predefined parameter so that the BS has no difficulty in calculating the accumulative difference A_i for each UE by Eq. (3) and converting it to a normalized credit $\hat{\mathbf{C}}_i$ by Eq. (4). On the other hand, the weight-assignment mechanism refers to the idea of M-LWDF, a classic scheduling method, to set a weight for each real-time flow. It is based on the delay threshold $\delta_{i,j}$ and the maximum tolerant probability $p_{i,j}$ of packet loss of each such flow, whose values can be determined in advanced according to the QCI table (i.e., Table 2) defined in the LTE standard. Finally, it is trivial to find the capacity q_i of each RB by referring to the CQI table (i.e., Table 1, which is also defined in the LTE standard). Thus, it becomes easy to match each RB with each packet in the RB-matching mechanism.

In our RB allocation algorithm, we translate these mechanisms into parameters $\hat{\mathbf{C}}_i$, $\hat{\mathbf{F}}_{i,j}$, and $\hat{\mathbf{W}}_{i,j}$. Each flow then can use its bidding value $b_{i,j}$ by Eq. (11) or Eq. (12) to bid for RBs. Afterwards, the BS adopts the *exponential effective SINR mapping (EESM)* method [28] to compute the effective SINR (signal-to-interference-plus-noise ratio) of the allocating RBs to a UE, which determines the actual amount of downlink data sent to that UE. Both Theorem 1 and Theorem 2 analyze the time and memory complexity of our RB allocation algorithm, which also demonstrates its high efficiency in computation and memory usage.

Theorem 1. Suppose that the maximum length of each flow's queue is B. Given m RBs and n flows, the worst-case time complexity of our RB allocation algorithm is $O((B+m)\times n)$.

Proof: In Step 1, the BS iteratively checks the HOL packet of each flow's queue and discards those overdue packets. The worst case occurs when every flow has a full queue and only the last packet in each queue is not overdue. In this case, the BS has to check B packets for each queue. In other words, the time complexity of Step 1 in the worst case will be $O(Bn)$. Then, Step 2 adopts the resource-reservation mechanism to classify UEs, which takes $O(2N)$ time according to Lemma 1, where N is the number of UEs in $\mathcal{U}$. Afterwards, Step 3 is a combination of the other three mechanisms. According to Lemmas 3, 5, and 7, this step will spend time of $O(3N)+O(n_R)+O(mn)$, where n_R is the number of real-time flows. Therefore, the overall time complexity of our RB allocation algorithm will be

$$\begin{aligned} &O(Bn)+O(2N)+O(3N)+O(n_R)+O(mn)\\ &=O(Bn)+O(N)+O(n_R)+O(mn). \end{aligned} \tag{13}$$

It is trivial that $n_R \le n$. Besides, since each UE can have one or more flows, we have $N \le n$. In consequence, Eq. (13) can be simplified to $O(Bn)+O(mn)=O((B+m)\times n)$, thereby proving the theorem. □

Theorem 2. Given m RBs and n flows, the amount of memory spent by our RB allocation algorithm is $O(mn)$.

Proof: Step 1 uses Eq. (10) to check overdue packets, where the variables $d_{i,j}$, $\psi(h_{i,j})$, and $\delta_{i,j}$ can be reused for each packet. Thus, the amount of memory spent by Step 1 is $O(1)$. According to Lemma 2, Step 2 consumes an amount of $O(N)$ memory to divide UEs into two groups of $\mathcal{U}_E$ and $(\mathcal{U}-\mathcal{U}_E)$ through the resource-reservation mechanism. Then, Step 3 uses the credit-driven, weight-assignment, and RB-matching mechanisms to allocate RBs to each flow. According to Lemmas 4, 6, and 8, it requires an amount of $(O(2N)+O(2n_R)+O(mn))$ memory to store the necessary information and data structures. Therefore, the overall memory complexity of our RB allocation algorithm will be

$$\begin{aligned} &O(1)+O(N)+O(2N)+O(2n_R)+O(mn)\\ &=O(N)+O(n_R)+O(mn). \end{aligned} \tag{14}$$

As mentioned earlier, we have both $N \le n$ and $n_R \le n$. Thus, the above equation can be simplified to $O(mn)$, thereby proving the theorem. □

5 Performance Evaluation

In this section, we evaluate the performance of our proposed RB allocation algorithm by using LTE-Sim, which is an open-source network simulator to model LTE behavior [32]. Table 3 presents the simulation parameters, where we set the transmission-related parameters (for loss and fading effect) following the LTE specification. Our simulations consider a 3km-radius LTE macro-cell, inside which 50 to 100 UEs move with a velocity of 3km/h and 120km/h (to simulate the walking and driving situations, respectively). Moreover, we also consider a *very high UE-density* scenario, where there are 300 UEs in the cell. This scenario helps evaluate system performance under a overload situation. In addition, three extra BSs are placed near the cell to generate signal interference to the region of cell edge. Each UE has two real-time flows (VoIP and video streaming) and one non-real-flow (constant-bit-rate, CBR). The deadlines of real-time packets are set to 100 ms.

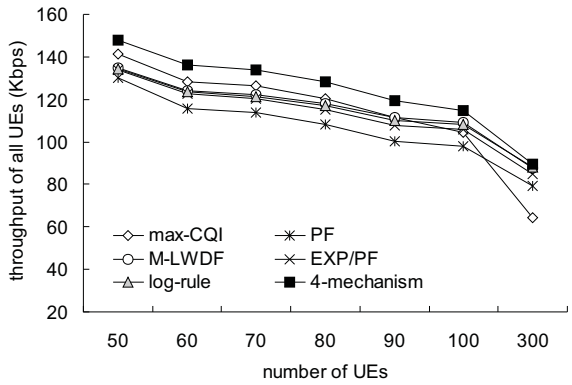
As discussed earlier in Section 2, LTE supports six types of downlink channels with different bandwidth. While many studies consider only narrow-band channels, for example, [9], [10], [16], [19] use 5 MHz channels and [15], [17], [24], [26] adopt 10 MHz channels in their experiments, our simulations employ a channel with the largest 20 MHz bandwidth to test different methods in a larger solution search space. We compare our RB allocation algorithm (denoted by '4-mechanism' in the simulation figures) with the max-CQI, PF, M-LWDF, and EXP/PF methods mentioned in Section 1. Except for these popular scheduling methods, we also compare our RB allocation algorithm with one classic method, called *log-rule* [33]. In particular, for each RB, the log-rule method picks the flow $f_{i,j}$ (belonging to UE u_i) that has the largest value of $(x\varphi_i \times \log(y+zd_{i,j}))$, where x, y, and z are tunable parameters, and φ_i is the spectral efficiency of u_i on the channel. Below, we measure network throughput, system fairness, and QoS support for real-time flows by the above scheduling methods and our RB allocation algorithm.

5.1 Network Throughput

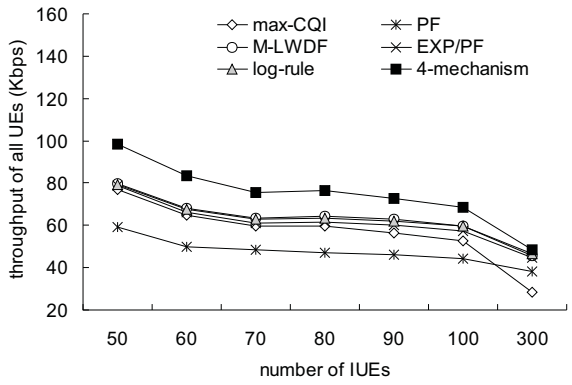
We first measure the average throughput of all UEs in $\mathcal{U}$, as shown in Fig. 3. Generally speaking, since the amount of downlink resource is fixed, the average throughput decreases as the number of UEs increases. Such effect is more obvious

TABLE 3: Experimental parameters used in our simulations.

BS-related parameters:	
type of BS	macro-cell BS with radius of 3 km
downlink channel	one single channel with bandwidth of 20 MHz
number of RBs	100 in each TTI
frame structure	FDD (frequency division duplexing) mode
modulation	QPSK, 16QAM, and 64QAM
UE-related parameters:	
number of UEs	50, 60, 70, 80, 90, 100, and 300 (very high UE-density scenario)
mobility model	random direction [29]
moving velocity	3 km/h (walking) and 120 km/h (driving)
real-time flow	8.4 Kbps VoIP traffic and 242 Kbps video streaming
non-real-time flow	12 Kbps CBR traffic
Transmission-related parameters (based on LTE specification [30]):	
path loss	$128.1 + 37.6 \log L$, where L is measured in km
penetration loss	10 dB
propagation loss	urban macro-cell model
slow/shadowing fading	log-normal distribution whose mean = 0 dB and standard deviation = 8 dB
fast/multipath fading	Jakes fading model [31]
Other parameters:	
simulation time	120 seconds
4-mechanism	τ_i = 2 Mbps and β = 0.05

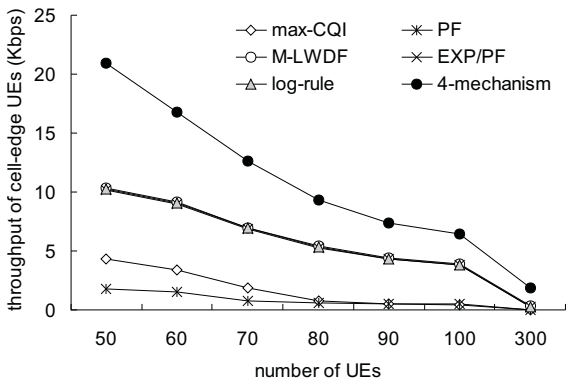


(a) UE velocity: 3km/h

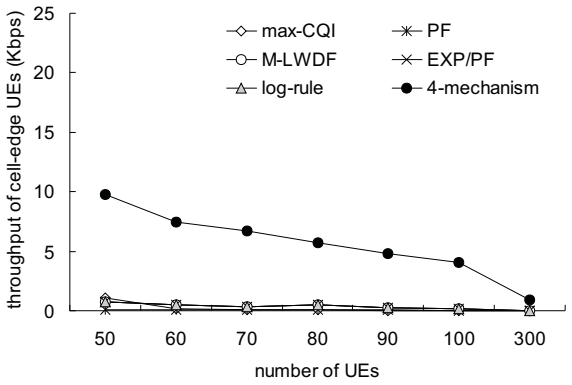


(b) UE velocity: 120km/h

Fig. 3: Comparison on the average throughput of all UEs.



(a) UE velocity: 3km/h

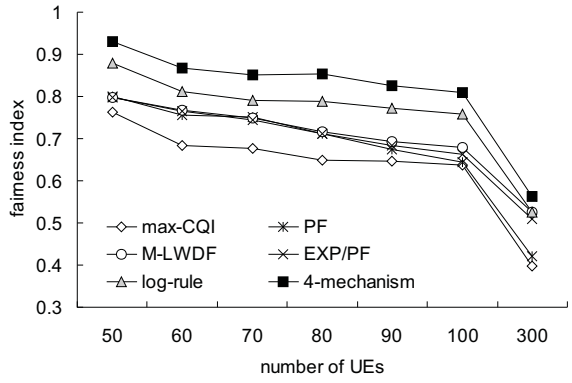


(b) UE velocity: 120km/h

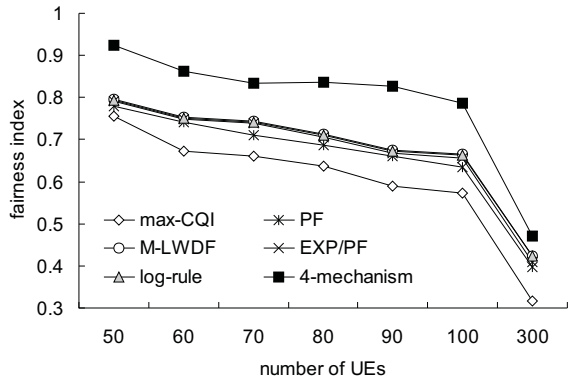
Fig. 4: Comparison on the average throughput of cell-edge UEs.

under the very high UE-density scenario. In addition, when UEs move in a higher velocity, their throughput will degrade because the fading effect becomes more significant.

For the max-CQI method, it greedily assigns RBs to the UEs that have the best CQI, so it results in higher throughput when UEs move slowly (i.e., 3 km/h). However, the channel condition may vary drastically when UEs move in a high velocity (i.e., 120 km/h), so the max-CQI method wins only the PF method in this situation. Interestingly, such a greedy policy may not perform well under the very high UE-density scenario, so the max-CQI method results in the lowest throughput when there are 300 UEs in the cell. On the other hand, since both M-LWDF and EXP/PF methods are the enhancements of the PF method, they will have higher throughput comparing with the PF method. The log-rule method refers to the spectral efficiency of each UE, which also helps improve throughput. Comparing to these methods, our RB allocation algorithm not only discards overdue packets in advance but also saves RBs' capacity by the RB-assignment mechanism. Therefore, it can achieve the highest throughput among all methods. Even under the very high UE-density scenario, our RB allocation algorithm can still have higher throughput than other methods, which demonstrates its effectiveness in terms of network transmission.

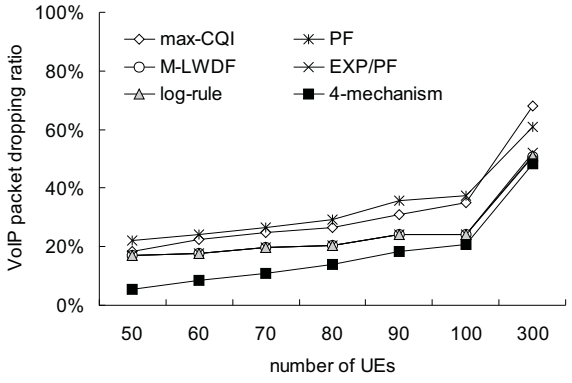


(a) UE velocity: 3km/h

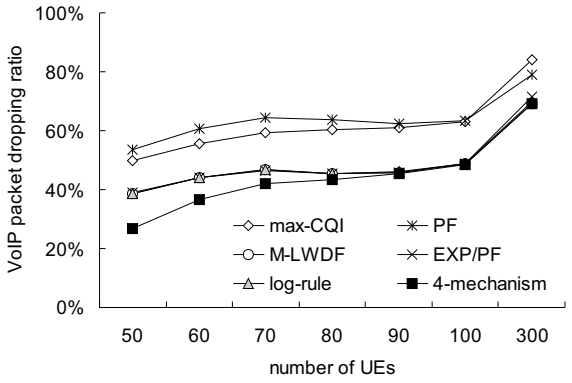


(b) UE velocity: 120km/h

Fig. 5: Comparison on Jain's fairness index.



(a) UE velocity: 3km/h



(b) UE velocity: 120km/h

Fig. 6: Comparison on the average dropping ratio of VoIP packets.

We then measure the average throughput of cell-edge UEs in $\mathcal{U}_E$ by different methods, as illustrated in Fig. 4. When UEs move in 3 km/h velocity, the network condition becomes relatively stable, so cell-edge UEs can obtain some resource in most methods. However, as the number of UEs grows, cell-edge throughput decreases accordingly, because more UEs will compete for the same resource. Under the very high UE-density scenario, cell-edge throughput is almost down to zero in all methods except for ours. On the other hand, when UEs move in 120 km/h velocity, the network condition varies drastically, so all methods (except for ours) give almost nothing to cell-edge UEs. Comparing with the above methods, our RB allocation algorithm keeps a dynamic portion of resource to be allocated to only cell-edge UEs by Eq. (2), which prevents them from starvation. Such a mechanism also helps improve system fairness, which will be discussed in the next section.

5.2 System Fairness

Fig. 5 evaluates the Jain's fairness index of each method calculated by Eq. (1). When the fairness index is closer to one, it means that the LTE network provides more fair transmission among all UEs. By comparing with Fig. 5(a) and Fig. 5(b), we observe that the fairness index decreases when the velocity of UEs increases, because the channel condition changes more drastically. Moreover, the fairness index drops fast under the very high UE-density scenario (in particular, less than 0.57 and 0.48 when UEs move in a velocity of 3 km/h and 120 km/h, respectively), since there will exist more UEs that can receive only little or even no resource from the BS.

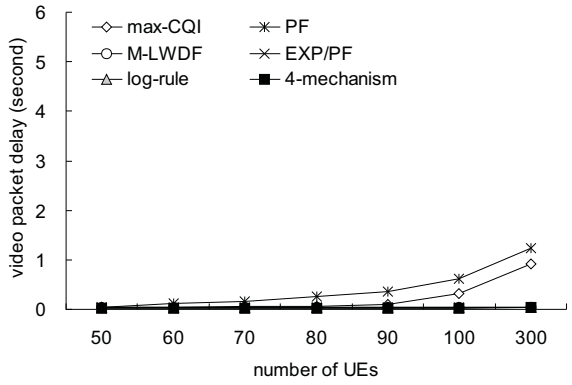
In Fig. 5, the max-CQI method always has the lowest fairness index, because it attempts to increase network throughput at the expense of those UEs with bad channel quality. On the other hand, the PF-based methods (i.e., PF, M-LWDF, and EXP/PF) use the metric of r_i/r_i^{avg} to allocate RBs, which provides a certain level of fairness. The log-rule method takes the spectral efficiency of each UE into consideration, which further improves fairness especially when the network becomes more stable (i.e., UE velocity = 3 km/h).

Comparing with these methods, our RB allocation algorithm adopts the credit-driven mechanism to provide a global view to control the amount of resource given to UEs for fair transmission. It evaluates the difference between expected and actual amount of resource given to each UE, and seeks to reduce such difference. Moreover, the resource-reservation mechanism guarantees that cell-edge UEs can receive some resource, even though they have bad channel quality. These two mechanisms together make our RB allocation algorithm always have the largest fairness index, even under the very high UE-density scenario.

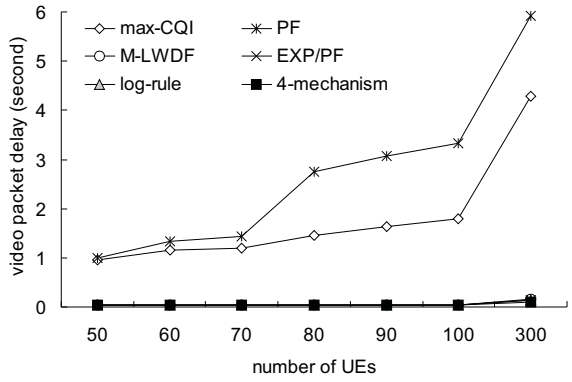
5.3 QoS Support

Next, we evaluate the average packet dropping ratio of VoIP flows by different methods, whose results are shown in Fig. 6. It can be expected that the packet dropping ratio will increase when the number of UEs grows or UEs move in a higher velocity. This phenomenon becomes more obvious under the very high UE-density scenario, where 48.1% $\sim$ 84.2% of VoIP packets will be dropped due to serious resource competition by numerous UEs.

From Fig. 6, both max-CQI and PF methods lead to more VoIP packet dropping, because they do not differentiate real-time flows from non-real-time ones. On the other hand, the M-LWDF, EXP/PF, and log-rule methods take packet delay $d_{i,j}$ into account, so they can alleviate packet dropping. The



(a) UE velocity: 3km/h



(b) UE velocity: 120km/h

Fig. 7: Comparison on the average delay of video packets.

weight-assignment mechanism in our RB allocation algorithm inherits the idea of M-LWDF. Moreover, its RB-assignment mechanism is able to better utilize RBs' capacity and thus improves throughput. Therefore, our RB allocation algorithm can further reduce the average packet dropping ratio of VoIP flows.

Finally, we investigate the average delay of video packets, as presented in Fig. 7. In the implementation of LTE-Sim, both max-CQI and PF methods do not discard the packets that have passed their deadlines. In other words, UEs may receive 'overdue' packets in these two methods. Therefore, even though the packet deadline of a video flow is 100 ms, the average video packet delay by both max-CQI and PF methods will become much larger than 100 ms (especially when UEs move in 120 km/h velocity or under the very high UE-density scenario). On the other hand, the M-LWDF, EXP/PF, log-rule, and our RB allocation methods address delays of real-time packets and drop those overdue packets. Consequently, their average video packet delays can be always kept below 100 ms in this experiment.

6 Conclusion and Future Work

LTE provides high-speed wireless access for 4G communication systems. The LTE downlink scheduling problem plays a critical role in system design but is not well addressed in 3GPP standards. This paper points out some drawbacks of existing solutions, and develops an efficient RB allocation algorithm by adopting the resource-reservation, credit-driven, weight-assignment, and RB-matching mechanisms. The designs of these mechanisms consider the practical constraints of LTE. We also prove that the proposed algorithm is lightweight in respect of computation and requires less memory storage, which assists the BS in quickly dealing out spectral resource among flows in every short TTI. Furthermore, through LTE-Sim experiments, we demonstrate that our RB allocation algorithm increases network throughput, especially for those UEs in the cell-edge region, improves system fairness, and reduces packet dropping and delays of real-time flows, as comparing with popular solutions including max-CQI, PF, M-LWDF, EXP/PF, and log-rule.

This paper aims at resource scheduling for a single downlink channel in LTE. To support much larger bandwidth, LTE-A, the advanced version of LTE, employs carrier aggregation by integrating multiple channels (probably in different bands) for communication. With carrier aggregation, the scheduling problem has to address how to select and combine different channels and allocate RBs accordingly to transmit data. This issue deserves further investigation in the future.

References

[1] Cisco Systems, "Cisco visual networking index: forecast and methodology," May 2015. [Online]. Available: http://www.cisco.com/

[2] P. Kela, J. Puttonen, N. Kolehmainen, T. Ristaniemi, T. Henttonen, and M. Moisio, "Dynamic packet scheduling performance in UTRA long term evolution downlink," *Proc. IEEE Int'l Symp. Wireless Pervasive Computing*, 2008, pp. 308–313.

[3] H.J. Kushner and P.A. Whiting, "Convergence of proportional-fair sharing algorithms under general conditions," *IEEE Trans. Wireless Comm.*, vol. 3, no. 4, pp. 1250–1259, 2004.

[4] M. Andrews, K. Kumaran, K. Ramanan, A. Stolyar, P. Whiting, and R. Vijayakumar, "Providing quality of service over a shared wireless link," *IEEE Comm. Magazine*, vol. 39, no. 2, pp. 150–154, 2001.

[5] J.H. Rhee, J.M. Holtzman, and D.K. Kim, "Scheduling of real/non-real time services: adaptive EXP/PF algorithm," *Proc. IEEE Vehicular Technology Conf.*, 2003, pp. 462–466.

[6] European Telecommunications Standards Institute, "LTE; Evolved Universal Terrestrial Radio Access (E-UTRA); Physical layer procedures," Technical Report, ETSI TS 136 213 V12.4.0, 2015.

[7] R. Jain, D.M. Chiu, and W. Hawe, "A quantitative measure of fairness and discrimination for resource allocation in shared computer system," DEC Research Report, TR-301, 1984.

[8] W.K. Lai and C.L. Tang, "QoS-aware downlink packet scheduling for LTE networks," *Computer Networks*, vol. 57, no. 7, pp. 1689–1698, 2013.

[9] C. Wang and Y.C. Huang, "Delay-scheduler coupled throughput-fairness resource allocation algorithm in the long-term evolution wireless networks," *IET Comm.*, vol. 8, no. 17, pp. 3105–3112, 2014.

[10] D. Samia, B. Ridha, and A. Wei, "Resource allocation using Nucleolus value in downlink LTE networks," *Proc. IEEE Symp. Computers and Comm.*, 2016, pp. 250–254.

[11] M. Leng and M. Parlar, "Analytic solution for the nucleolus of a three-player cooperative game," *Naval Research Logistics*, vol. 57, no. 7, pp. 667–672, 2010.

[12] Y.C. Wang and S.Y. Hsieh, "Service-differentiated downlink flow scheduling to support QoS in long term evolution," *Computer Networks*, vol. 94, pp. 344–359, 2016.

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

[14] 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 Trans. Multimedia*, vol. 13, no. 5, pp. 1052–1065, 2011.

[15] Q. Liu and C. W. Chen, "Smart downlink scheduling for multimedia streaming over LTE networks with hard handoff," *IEEE Trans. Circuits and Systems for Video Technology*, vol. 25, no. 11, pp. 1815–1829, 2015.

[16] X. Liu, "Solving the nonlinear LTE resource allocation problem with a linear approach," *Proc. IEEE Vehicular Technology Conf.*, 2016, pp. 1–5.

[17] M. Iturralde, T.A. Yahiya, A. Wei, and A.L. Beylot, "Resource allocation using Shapley value in LTE networks," *Proc. IEEE Int'l Symp. Personal Indoor and Mobile Radio Comm.*, 2011, pp. 31–35.

[18] A.E. Roth, *The Shapley Value*, Cambridge University Press, 2005.

[19] F. Huang, V. Veque, and J. Tomasik, "A Pareto-optimal approach for resource allocation on the LTE downlink," *Proc. IEEE Int'l Conf. Comm.*, 2016, pp. 1–7.

[20] Y.C. Wang, "A two-phase dispatch heuristic to schedule the movement of multi-attribute mobile sensors in a hybrid wireless sensor network," *IEEE Trans. Mobile Computing*, vol. 13, no. 4, pp. 709–722, 2014.
[21] S. Schwarz, C. Mehlfuhrer, and M. Rupp, "Throughput maximizing multiuser scheduling with adjustable fairness," *Proc. IEEE Int'l Conf. Comm.*, 2011, pp. 1–5.
[22] T. Lan, D. Kao, M. Chiang, and A. Sabharwal, "An axiomatic theory of fairness in network resource allocation," *Proc. IEEE INFOCOM*, 2010, pp. 1–9.
[23] S. Ali and M. Zeeshan, "A utility based resource allocation scheme with delay scheduler for LTE service-class support," *Proc. IEEE Wireless Comm. and Networking Conf.*, 2012, pp. 1450–1455.
[24] B. Liu, H. Tian, and L. Xu, "An efficient downlink packet scheduling algorithm for real time traffics in LTE systems," *Proc. IEEE Consumer Comm. and Networking Conf.*, 2013, pp. 364–369.
[25] D. Liu and Y.H. Lee, "An efficient scheduling discipline for packet switching networks using earliest deadline first round robin," *Proc. IEEE Int'l Conf. Computer Comm. and Networks*, 2003, pp. 5–10.
[26] M.B. Shahab, M.A. Wahla, and M.T. Mushtaq, "Downlink resource scheduling technique for maximized throughput with improved fairness and reduced BLER in LTE," *Proc. IEEE Int'l Conf. Telecomm. and Signal Processing*, 2015, pp. 163–167.
[27] Y.C. Wang and Y.C. Tseng, "Packet fair queuing algorithms for wireless networks," in *Design and Analysis of Wireless Networks*, Nova Science Publishers, 2005, pp. 113–128.
[28] R. Giuliano and F. Mazzenga, "Exponential effective SINR approximations for OFDM/OFDMA-based cellular system planning," *IEEE Trans. Wireless Comm.*, vol. 8, no. 9, pp. 4434–4439, 2009.
[29] W.H. Yang, Y.C. Wang, Y.C. Tseng, and B.S.P. Lin, "Energy-efficient network selection with mobility pattern awareness in an integrated WiMAX and WiFi network," *Int'l J. Comm. Systems*, vol. 23, no. 2, pp. 213–230, 2010.
[30] European Telecommunications Standards Institute, "LTE; Evolved Universal Terrestrial Radio Access (E-UTRA); Radio Frequency (RF) requirements for LTE Pico Node B," Technical Report, ETSI TR 136 931 V9.0.0, 2011.
[31] Y.C. Wang and C.A. Chuang, "Efficient eNB deployment strategy for heterogeneous cells in 4G LTE systems," *Computer Networks*, vol. 79, pp. 297–312, 2015.
[32] G. Piro, L.A. Grieco, G. Boggia, F. Capozzi, and P. Camarda, "Simulating LTE cellular systems: an open-source framework," *IEEE Trans. Vehicular Technology*, vol. 60, no. 2, pp. 498–513, 2011.
[33] B. Sadiq, S.J. Baek, and G. de Veciana, "Delay-optimal opportunistic scheduling and approximations: the log rule," *IEEE/ACM Trans. Networking*, vol. 19, no. 2, pp. 405–418, 2011.