

# A 4-guideline Downlink Scheduling Strategy to Support Fairness and QoS for LTE Networks

You-Chiun Wang and Dai-Rong Zhong

Department of Computer Science and Engineering,

National Sun Yat-Sen University, Kaohsiung, 80424, Taiwan

Email: ycwang@cse.nsysu.edu.tw; m023040024@student.nsysu.edu.tw

**Abstract**—LTE has been the main protocol for 4G networks, which supports large-demand multimedia service. How to schedule its downlink spectral resource, namely *physical resource blocks (PRBs)*, has great impact on performance but is not specified in standards. This paper thus proposes a *4-guideline LTE downlink scheduling (4G-LDS)* strategy to support fairness and QoS in LTE networks. For fairness, 4G-LDS uses a credit-based guideline to adjust the amount of resource given to each user, and a cell-division guideline to reserve bandwidth for cell-edge users to avoid starving them. For QoS, 4G-LDS adopts flow-weight and packet-fitness guidelines to allocate each PRB to a packet based on its flow type and size. Simulation results show that 4G-LDS outperforms previous methods in terms of network throughput, system fairness, and QoS support.

**Index Terms**—downlink communication, fairness, long term evolution (LTE), quality of service (QoS), resource scheduling.

## I. INTRODUCTION

Today, people can enjoy broadband communication via their smart phones. Cisco also points out that multimedia streaming and video downloads have been dominating Internet traffics [1]. Accordingly, 3GPP continues to develop *long term evolution (LTE)* to provide high-speed wireless access, especially for large-demand downlink service, for 4G networks.

LTE uses OFDMA (orthogonal frequency division multiple access) for downlink communication. A *physical resource block (PRB)* is the unit to allocate resource to each *user equipment (UE)*. With a different modulation and coding scheme, each PRB can carry different amount of data. How to allocate PRBs to UEs based on user demands is called the *LTE downlink scheduling problem*. This problem greatly affects 4G performance, but 3GPP leaves it to LTE implementers.

There are some popular solutions to the problem [2]. The max-CQI method uses a greedy policy by giving each PRB to the UE with the best *channel quality indication (CQI)*. It improves network throughput but may starve the UEs with bad channel condition. Thus, the PF (proportional fair) method compares the data rate  $r_i$  of each UE with its past rate  $r_i^{avg}$ , and allocates a PRB to the UE by  $u_i = \arg \max_i \{r_i/r_i^{avg}\}$ . However, PF does not address the delay issue. So, the M-LWDF (modified largest weighted delay first) method introduces a weight  $w_i$  and the *head-of-line (HOL)* packet delay  $d_i$  into PF by  $u_i = \arg \max_i \{w_i d_i \times r_i/r_i^{avg}\}$ . To favor real-time flows, the EXP/PF (exponential proportional fair) method

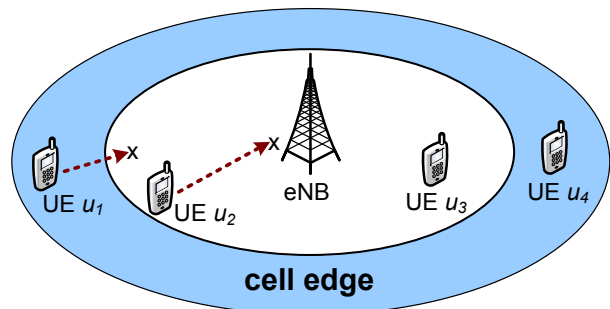


Fig. 1: Examples to show unfair transmission by the PF method.

modifies M-LWDF by adding the average HOL packet delay  $d_i^{avg}$  of all real-time flows as follows:

$$u_i = \arg \max_i \left\{ \exp \left[ \frac{w_i d_i - d_i^{avg}}{1 + \sqrt{d_i^{avg}}} \right] \times \frac{r_i}{r_i^{avg}} \right\}.$$

However, some issues are arisen in these methods. First, most of them are modified from PF, which prefers giving PRBs to the UEs whose channel condition improves (i.e.,  $r_i/r_i^{avg}$  increases). But, PF lacks a *global view* to control the amount of resource given to UEs to provide fair transmission. Let us consider an example in Fig. 1, where two UEs  $u_1$  and  $u_2$  stay in their original positions for a while and move toward the eNB such that  $r_1/r_1^{avg} < r_2/r_2^{avg}$ . Thus, PF will favor  $u_2$ , which enlarges the difference between the transmission amount of  $u_1$  and  $u_2$  (i.e., more unfair). Actually, when the channel condition of  $u_1$  and  $u_2$  improves, we can give more resource to  $u_1$  to keep fairness, as  $u_1$  has received less data than  $u_2$ . Second, these methods do not give special treatment of *cell-edge UEs*, and such UEs may not get enough resource due to bad channel quality (i.e., starvation). Fig. 1 shows an example, where two UEs  $u_3$  and  $u_4$  keep static. In this case, we have  $r_3/r_3^{avg} \approx r_4/r_4^{avg} \approx 1$ . Thus, PF (and its variation) views  $u_3$  and  $u_4$  as no difference from the metric  $r_i/r_i^{avg}$ . In fact,  $u_4$  (i.e., a cell-edge UE) receives much less data than  $u_3$ . Third, some methods like max-CQI do not *differentiate* flows by their types. Consequently, they would not well support quality of service (QoS) for real-time applications. Finally, no method considers the *size-relationship* between PRBs and

TABLE I: LTE CQI table.

index	modulation	code rate ( $\times 1024$ )	efficiency
1	QPSK	78	0.1523
2	QPSK	120	0.2344
3	QPSK	193	0.3770
4	QPSK	308	0.6016
5	QPSK	449	0.8770
6	QPSK	602	1.1758
7	16QAM	378	1.4766
8	16QAM	490	1.9141
9	16QAM	616	2.4063
10	64QAM	466	2.7305
11	64QAM	567	3.3223
12	64QAM	666	3.9023
13	64QAM	772	4.5234
14	64QAM	873	5.1152
15	64QAM	948	5.5547

packets. When a PRB with a more complex modulation is used to send a small packet, the PRB is wasted.

From the above observations, we develop a *4-guideline LTE downlink scheduling (4G-LDS) strategy* to support fairness and QoS for LTE networks. In the *credit-based guideline*, each UE records a credit value to adjust the amount of resource that it can obtain in a scheduling period, so as to keep system fairness. In the *cell-division guideline*, we reserve a dynamic portion of network resource for the UEs close to the cell's edge, in order to prevent them from starvation. In the *flow-weight guideline*, each flow is given a weight based on its type, data rate, and HOL packet delay. Real-time flows should have larger weights for QoS concern [3]. In the *packet-fitness guideline*, we match each PRB with a packet by referring to the PRB's capacity (i.e., the number of bits carried) and the packet's size. The objective is to make good use of the PRB without wasting its capacity. Then, 4G-LDS quantifies each guideline to a variable, and combines them in an equation to help the LTE base station (called *eNB*) allocate PRBs. Through simulation, we verify that the 4G-LDS strategy can efficiently increase network throughput, improve system fairness, and reduce packet dropping of real-time flows.

## II. SYSTEM MODEL

We consider an LTE macro-cell controlled by one eNB, which manages downlink resource for the cell's UEs. The resource is divided into a 2D array of PRBs, where each PRB has 0.5ms duration and 180KHz width. Depending on the downlink channel bandwidth, the eNB provides different number of PRBs in a scheduling period, called transmission time interval (TTI = 1ms). If the channel has bandwidth of 1.4, 3, 5, 10, 15, and 20MHz, the eNB respectively provides 6, 15, 25, 50, 75, and 100 PRBs. When SU-MIMO (single-user multiple-input and multiple-output) is used, PRBs will be non-sharable, so a PRB can be allocated to at most one UE.

Each UE reports CQI in every TTI to reveal its channel condition, where Table I lists CQIs supported by LTE [4]. A higher CQI indicates that the UE has better channel quality. The eNB then refers to CQI to determine the modulation and coding scheme used to transmit the UE's data. From Table I,

LTE supports three modulations, each with different coding rates: QPSK (CQI = 1~6), 16QAM (CQI = 7~9), and 64QAM (CQI = 10~15). An PRB can carry more data bits when a more complex modulation and a higher code rate are used.

Given the CQI and demand of each UE, the LTE downlink scheduling problem asks how to allocate PRBs to meet all UEs' demands, under the SU-MIMO assumption. There exist feasible solutions only when we have enough PRBs. In case of insufficient PRBs, our goals are to increase network throughput, improve system fairness, and support QoS for real-time flows. We use Jain's fairness index [5] to evaluate fairness:  $FI = \frac{(\sum_{i=1}^n x_i)^2}{n \sum_{i=1}^n x_i^2}$ , where  $x_i$  is the normalized throughput of a flow and  $n$  is the number of flows. We have  $0 < FI \leq 1$ , and larger FI implies that the network is more fair. Besides, we use the packet dropping ratio of a real-time flow to evaluate its degree of QoS support. Apparently, a lower ratio implies that the flow has higher QoS support.

## III. RELATED WORK

The LTE downlink scheduling problem has attracted lots of attention, and many solutions are proposed to support QoS for real-time service. The work of [6] aims at delivering video over LTE. It considers the rate, delay, and distortion of each video flow, and determines resource allocation and coding for that flow. Piro et al. [7] propose a two-layer scheduling method for LTE multimedia service. The first layer finds the amount of data that a multimedia flow has to send within a period to meet its delay constraint. The next layer then uses PF to allocate PRBs accordingly. The work of [8] uses a virtual queue to predict future incoming packets, and discards the packets that fail to meet their delay limit. Then, it adopts max-CQI to allocate PRBs to UEs. Wang and Hsieh [9] also use max-CQI to decide PRB allocation, and ask non-urgent flows to give back a portion of allocated PRBs. Such PRBs are redistributed among the flows being threatened by packet dropping.

Some studies try to achieve fair transmission in LTE. The work of [10] converts the LTE downlink scheduling problem to a bankruptcy game problem, and solves it by the Shapley value to provide fair resource allocation. Schwarz et al. [11] use the  $\alpha$ -fair utility function in [12] to deal out resource, which depends on the user's average throughput and the parameter  $\alpha$ . The work of [13] uses a utility function to find the satisfaction degree of each flow, and allows flows to bid for resource by their utility values. Liu et al. [14] apply the earliest-deadline-first (EDF) concept to the PF method, which first picks the packet with the most urgent deadline to transmit. So, it could combine both the fairness feature of PF and the bounded-delay feature of EDF. Distinguishing from the previous work, this paper seeks to improve fairness by not only using a credit idea but also taking care of cell-edge UEs. Moreover, through flow-weight and packet-fitness guidelines, our 4G-LDS strategy can increase network throughput while support QoS for real-time service. Experimental results in Section V will also demonstrate its effectiveness.

#### IV. THE 4G-LDS STRATEGY

Our strategy follows four guidelines to schedule LTE resource. Below, we first detail each guideline, and then discuss how 4G-LDS combines these guidelines and its rationale.

##### A. Credit-based Guideline

Inheriting from *weighted fair queuing* [15], [16], we expect that each UE  $u_i$  can acquire a fixed amount  $\tau_i$  of resource in every TTI, where  $\tau_i$  depends on  $u_i$ 's demand. Ideally, if all UEs each exactly receives  $k\tau_i$  amount of data in  $k$  TTIs, for  $k \geq 1$ , the network is said to achieve fairness. However, due to the change of channel condition and different packet sizes, it is difficult to ask the eNB to transmit exact  $\tau_i$  amount of data to a UE  $u_i$  in each TTI. Thus, we keep a variable  $D_i$  for each UE  $u_i$  to record the accumulative difference between the amount of downlink data actually received by  $u_i$  and the amount of resource that  $u_i$  is expected to obtain. Let  $r_{i,k}$  be the amount of downlink data transmitted to  $u_i$  in the  $k$ th TTI. Then, we can compute the accumulative difference by  $D_i = \sum_k r_{i,k} - \tau_i$ . Here,  $D_i > 0$  means that  $u_i$  uses more resource than expectation, so the eNB should give its resource to others for fairness.  $D_i < 0$  implies that  $u_i$  does not receive enough data, so it is better to allocate more resource to  $u_i$  later.

However, the variation of  $D_i$  may be large, especially when some UEs have much better channel quality but others do not. Thus, we translate  $D_i$  into a *normalized credit* as follows:

$$\hat{C}_i = 2 - \frac{D_i - D_{\min}}{D_{\max} - D_{\min}}, \quad (1)$$

where  $D_{\min} = \min_{\forall i} \{D_i\}$  and  $D_{\max} = \max_{\forall i} \{D_i\}$ . So, we can limit the credit  $\hat{C}_i$  to  $[1, 2]$ . Here, a larger credit implies that UE  $u_i$  has a higher priority to get resource, as it receives less data than expectation, and vice versa. We give an example with three UEs, where  $\tau_i = 30\text{Kb/TTI}$  for  $i = 1, 2, 3$ . In the first TTI, UEs  $u_1$ ,  $u_2$ , and  $u_3$  receive 30, 50, and 10Kb data, respectively. Then, we can derive that  $D_1 = 0$ ,  $D_2 = 20$ , and  $D_3 = -20$ . Thus, the normalized credits will be  $\hat{C}_1 = 1.5$ ,  $\hat{C}_2 = 1$ , and  $\hat{C}_3 = 2$ . In this case, these UEs have priorities of  $u_3 > u_1 > u_2$ . So,  $u_3$  and  $u_2$  will be given more and less resource later to keep fairness, respectively.

##### B. Cell-division Guideline

When a UE moves close to cell edge, its channel condition may degrade due to path loss or interference. We call such UEs *cell-edge UEs*. However, many methods like max-CQI disfavor cell-edge UEs, or view them as no difference than other UEs (e.g., the PF method, where UEs  $u_3$  and  $u_4$  in Fig. 1 give an example). These UEs may be the underdog in competing network resource, thereby causing starvation to them. Thus, the cell-division guideline solves this problem by reserving a small, dynamic portion of resource for cell-edge UEs. Let  $\alpha_i$  be the demand of a UE  $u_i$ . Also, we denote by  $\mathcal{U}$  and  $\mathcal{U}_e$  the sets of all UEs and cell-edge UEs, respectively. Then, the eNB reserves a number of  $\gamma$  PRBs for  $\mathcal{U}_e$  by:

$$\gamma = \left\lceil \min \left\{ \frac{\sum_{u_i \in \mathcal{U}_e} \alpha_i}{\sum_{u_i \in \mathcal{U}} \alpha_i}, \beta \right\} \times m \right\rceil, \quad (2)$$

where  $m$  is the number of available PRBs in a TTI. Here,  $\gamma$  is proportional to the demands of all cell-edge UEs. However, since cell-edge UEs can still compete with others for the  $(m - \gamma)$  unreserved PRBs, it is unwise to reserve much resource for them, or network throughput will greatly degrade. Thus, we add a small threshold  $0 < \beta \leq 0.1$  in Eq. (2).

Another problem is how to find the UEs of  $\mathcal{U}_e$ . An intuitive solution is to use the distance between a UE and the eNB. However, this solution has two drawbacks. First, it requires the knowledge of UEs' locations. Second, it is not easy to choose the distance threshold. Thus, we propose a solution by using CQI. When a UE reports  $\text{CQI} \leq 6$ , it is considered as a cell-edge UE. Our solution adds no overhead, since every UE has to report its CQI to the eNB in each TTI. Besides, when  $\text{CQI} \leq 6$ , the eNB has to choose the simplest modulation, QPSK, to send data (referring to Table I). Thus, there is a high possibility that the UE stays near the cell's edge.

##### C. Flow-weight Guideline

In LTE, a UE can have multiple flows, which share the resource gotten by the UE. However, real-flows have the feature of stringent delay requirement. Therefore, they should be given with a higher priority for transmission depending on their HOL packet delays. To do so, we differentiate flows by giving them with different *flow weights* based on their types. Specifically, let  $r_i$  be the data rate of a UE  $u_i$ , and  $d_{i,j}$  be the HOL packet delay of  $u_i$ 's flow  $f_{i,j}$ . Then, if  $f_{i,j}$  is a real-time flow, we set  $\hat{W}_{i,j} = w_{i,j} \times d_{i,j} \times r_i$ , where  $w_{i,j}$  is a weight defined in the M-LWDF method (referring to Section I). On the other hand, if  $f_{i,j}$  is a non-real-time flow, we simply set  $\hat{W}_{i,j} = r_i$ .

In fact, our flow-weight guideline is a mixture of both 'revised' M-LWDF and max-CQI methods. We apply M-LWDF to real-time flows to give them a larger weight by considering their HOL packet delays  $d_{i,j}$ . However, instead of using the original term  $r_i/r_i^{avg}$  in M-LWDF, we replace it by  $r_i$  due to two reasons. First, we have adopted the normalized credit  $\hat{C}_i$  to keep fair transmission, so there is no need to use the PF's metric  $r_i/r_i^{avg}$ . Second, the formats of flow weights can be consistent. In other words, the weight of a real-time flow can be viewed as adding a scaling factor ( $w_{i,j} \times d_{i,j}$ ) to the weight of a non-real-time flow.

##### D. Packet-fitness Guideline

Existing methods allocate each PRB to a flow by its channel quality, data rate, or HOL packet delay. However, they do not care whether the PRB is able to 'fit' the sending packet by its size. When a large-capacity PRB is used to send a small-size packet, the PRB is wasted. Thus, we calculate a *fitness degree*  $\hat{F}_{i,j}$  for each flow  $f_{i,j}$  in this guideline. A larger  $\hat{F}_{i,j}$  degree means that the PRB is more suitable to send the flow's packet.

Let  $l_{i,j}$  be the size of  $f_{i,j}$ 's HOL packet, and  $\theta_i$  be the PRB's capacity. We denote by  $\Theta_{64QAM}$  and  $\Theta_{16QAM}$  the minimum capacity of an PRB when 64QAM and 16QAM are adopted, respectively. Then, we compute  $\hat{F}_{i,j}$  by three cases below.

- 1)  $\theta_i \geq \Theta_{64QAM}$ : The PRB has large capacity, so we prefer matching it with a large-size packet to reduce wastage. Thus, the fitness degree of  $f_{i,j}$  is set to  $\hat{F}_{i,j} = \min\{l_{i,j}/\theta_i, 1\}$ . Here, when the HOL packets of multiple flows have sizes larger than  $\theta_i$ , it does not matter to choose which flow, as we need multiple PRBs to finish sending out each of these packets. That is why we select the minimum value between  $l_{i,j}/\theta_i$  and 1.
- 2)  $\Theta_{16QAM} \leq \theta_i < \Theta_{64QAM}$ : Two subcases are considered. When  $l_{i,j} \geq \Theta_{64QAM}$ , the packet size exceeds the PRB's capacity. So, the fitness degree is set to  $\theta_i/\Theta_{64QAM}$  to let the PRB send out more data bits of the packet. Otherwise, we set  $\hat{F}_{i,j} = \min\{l_{i,j}/\theta_i, 1\}$  to reduce capacity's wastage.
- 3)  $\theta_i < \Theta_{16QAM}$ : Similarly, we set  $\hat{F}_{i,j}$  to  $\theta_i/\Theta_{16QAM}$  if  $l_{i,j} \geq \Theta_{16QAM}$ , or  $\min\{l_{i,j}/\theta_i, 1\}$  otherwise.

### E. Discussion

In a TTI, our 4G-LDS strategy works as follows. It first checks for each flow whether it has *overdue* packets. Here, we say that the HOL packet of a flow  $f_{i,j}$  is overdue if  $d_{i,j} + \rho(p_{i,j}^{HOL}) > d_{i,j}^{max}$ , where  $\rho(p_{i,j}^{HOL})$  is the propagation delay to send out the HOL packet, and  $d_{i,j}^{max}$  is the corresponding delay tolerant time. If so, this HOL packet is discarded, and we iteratively check the next packet, until the HOL packet is not overdue. Then, 4G-LDS deals out  $\gamma$  reserved PRBs to the cell-edge UEs in  $\mathcal{U}_e$ , and allows all UEs in the cell to compete for  $(m - \gamma)$  unreserved PRBs. In the above allocation, we give each PRB to a flow  $f_{i,j}$  (belonging to UE  $u_i$ ) according to the equation:

$$f_{i,j} = \arg \max_{i,j} \begin{cases} \hat{C}_i \times \hat{F}_{i,j} \times r_i & \text{if } u_i \in \mathcal{U}_e, \\ \hat{C}_i \times \hat{F}_{i,j} \times \hat{W}_{i,j} & \text{otherwise.} \end{cases} \quad (3)$$

4G-LDS possesses some special designs. First, we employ the idea from [9] to drop the packets that will pass their deadlines, so as to save bandwidth. Second, the eNB spends  $\gamma$  PRBs to let cell-edge UEs have an opportunity to send their data. From Eq. (2), we have  $\gamma \leq 0.1m$ , so the reservation will not greatly degrade throughput. Moreover, cell-edge UEs can compete for unreserved PRBs to handle the cases when most UEs are close to the cell's edge or they have large amount of demands. Third, since cell-edge UEs can use only the simplest modulation (i.e., QPSK), it would have slight impact to differentiate flows by their types. Thus, we do not apply the flow-weight guideline (i.e.,  $\hat{W}_{i,j}$ ) to cell-edge UEs. Instead, we enhance max-CQI by adding  $\hat{C}_i$  (credit) and  $\hat{F}_{i,j}$  (fitness index) to increase the cell-edge throughput. Finally,  $\hat{C}_i$  provides a global view for the eNB to find the UEs without obtaining enough resource. In Section V, we will show that the credit mechanism provides more fair transmission than the PF policy (i.e. using  $r_i/r_i^{avg}$ ).

## V. PERFORMANCE EVALUATION

We adopt LTE-Sim [19] to evaluate system performance. Table II summarizes our parameters in the experiments. We consider a macro-cell, where 50~100 UEs move in speed of

TABLE II: Parameters used in our simulation.

Parameters for the eNB:	
cell range	3km
channel bandwidth	20MHz
number of PRBs	100
frame structure	frequency division duplexing (FDD)
Parameters for UEs:	
number of UEs	50, 60, 70, 80, 90, 100
mobility model	random walking [17]
moving speed	3km/h
flows	video streaming (242Kbps) and CBR (12Kbps)
Parameters for transmission:	
path loss	$128.1 + 37.6 \log L$ , $L$ is the distance in km
penetration loss	10dB
slow fading	log-normal distribution (zero mean)
fast fading	Jakes fading model [18]

3km/h (i.e., walking). Three eNBs are placed near the cell to generate interference. Each UE has two flows: 242Kbps real-time flow (video streaming) and 12Kbps non-real-time flow (constant-bit-rate, CBR). When the delay of a real-time packet exceeds 100ms, it will be dropped. We compare 4G-LDS with max-CQI, M-LWDF, and EXP/PF. The total simulation time is 100 seconds.

We first measure the average throughput of all UEs in Fig. 2(a). Because the amount of resource is constant, the average throughput decreases when the number of UEs grows. For the max-CQI methods, it greedily assigns PRBs to the UEs with the best CQI, resulting in higher throughput than both M-LWDF and EXP/PF methods. Our 4G-LDS strategy not only discards overdue packets but also saves PRBs' capacity by the packet-fitness guideline. Thus, it achieves the highest throughput. On the average, 4G-LDS improves 5.14%, 6.81%, and 8.39% of overall throughput than max-CQI, M-LWDF, and EXP/PF, respectively.

We then measure the average throughput of cell-edge UEs in Fig. 2(b). Since there is a high possibility that these UEs stay in the cell-edge region, their channel quality may keep bad for a long time. Due to the PF policy (i.e.,  $r_i/r_i^{avg}$ ), M-LWDF and EXP/PF almost do not allocate PRBs to cell-edge UEs. To avoid starving cell-edge UEs, 4G-LDS reserves a portion of PRBs for them. From Eq. (2), when there are more cell-edge UEs, we use a threshold  $\beta$  to limit the number of reserved PRBs. That is why the cell-edge throughput decreases when the number of UEs grows. Averagely, 4G-LDS has 8.46, 15.46, and 16.97 times the cell-edge throughput comparing with max-CQI, M-LWDF, and EXP/PF, respectively.

Fig. 2(c) evaluates fairness based on Jain's index. The max-CQI method always has the lowest index, because it tries to increase throughput at the expense of UEs with bad channel quality. M-LWDF and EXP/PF extend the PF idea by comparing  $r_i/r_i^{avg}$ , and thus they can improve fairness than max-CQI. However, PF lacks the global view to control the amount of resource given to UEs for fair transmission. Therefore, 4G-LDS adopts credits to find the difference between expected and actual amount of resource given to each UE, and reduces such difference. On the average, 4G-LDS increases 13.60%, 2.93%, and 3.97% of fairness as compared with max-CQI,

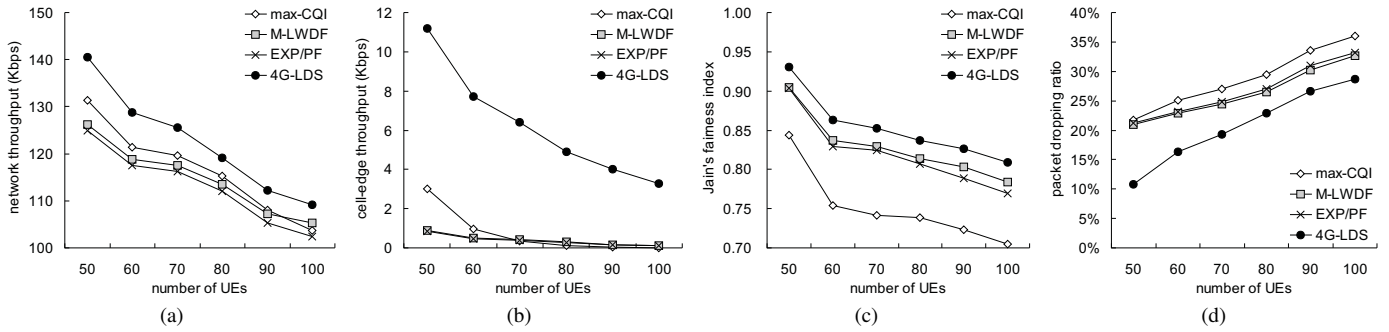


Fig. 2: Experimental results: (a) average throughput of all UEs, (b) average throughput of cell-edge UEs, (c) Jain's fairness index, and (d) dropping ratio of real-time packets.

M-LWDF, and EXP/PF, respectively.

Finally, we compare the packet dropping ratio of real-time flows, where the experimental results are presented in Fig. 2(d). The max-CQI method does not differentiate real-time flows from others, so it results in the highest ratio. Both M-LWDF and EXP/PF consider the HOL packet delay, so they alleviate real-time packet dropping. By the flow-weight guideline, our 4G-LDS strategy exploits M-LWDF's property to favor real-time flows. Besides, thanks to the packet-fitness guideline, 4G-LDS can make good use of PRBs' capacity, thereby increasing the throughput of real-time flows. That is why 4G-LDS has a lower ratio than others. In particular, 4G-LDS reduces 27.96%, 20.99%, and 22.39% of real-time packet dropping than max-CQI, M-LWDF, and EXP/PF, respectively.

## VI. CONCLUSION AND FUTURE WORK

The LTE downlink scheduling problem plays an important role in performance but is not addressed in standards. This paper points out some drawbacks of existing solutions, and develops the 4G-LDS strategy that obeys four guidelines to efficiently allocate PRBs to each flow. Through LTE-Sim experiments, we show that 4G-LDS increases throughput, especially for cell-edge UEs, improves fairness, and alleviates dropping of real-time packets, as comparing with max-CQI, M-LWDF, and EXP/PF. For the future work, we will aim at resource scheduling in LTE heterogeneous networks [20], where macro-cells and pico-cells can cooperate to provide network service.

## ACKNOWLEDGEMENTS

You-Chiun Wang's research is co-sponsored by the Ministry of Science and Technology under Grant No. MOST 104-2221-E-110-036-MY2 and 104-2628-E-110-001-MY2, Taiwan.

## REFERENCES

- [1] Cisco, "Visual networking index: forecast and methodology, 2014–2019 white paper," 2015. [Online]. Available: <http://www.cisco.com/>
- [2] F. Capozzi, G. Piro, L. A. Grieco, G. Boggia, and P. Camarda, "Downlink packet scheduling in LTE cellular networks: key design issues and a survey," *IEEE Communications Surveys & Tutorials*, vol. 15, no. 2, pp. 678–700, 2013.
- [3] Y. C. Wang and S. Y. Hsieh, "QoS-provisioning downlink resource management in 4G cellular systems," in *IEEE International Wireless Communications & Mobile Computing Conference*, 2015, pp. 67–72.
- [4] ETSI, "LTE; Evolved Universal Terrestrial Radio Access (E-UTRA); Physical layer procedures," TS 136 213 V12.4.0, 2015.
- [5] R. Jain, D. M. Chiu, and W. Hawe, "A quantitative measure of fairness and discrimination for resource allocation in shared computer system," Digital Equipment Corporation, Tech. Rep. TR-301, 1984.
- [6] H. Luo, S. Ci, D. Wu, J. Wu, and H. Tang, "Quality-driven cross-layer optimized video delivery over LTE," *IEEE Communications Magazine*, vol. 48, no. 2, pp. 102–109, 2010.
- [7] 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.
- [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] 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.
- [10] M. Iturralde, T. A. Yahiya, A. Wei, and A. L. Beylot, "Resource allocation using Shapley value in LTE networks," in *IEEE International Symposium on Personal Indoor and Mobile Radio Communications*, 2011, pp. 31–35.
- [11] S. Schwarz, C. Mehlhauer, and M. Rupp, "Throughput maximizing multiuser scheduling with adjustable fairness," in *IEEE International Conference on Communications*, 2011, pp. 1–5.
- [12] T. Lan, D. Kao, M. Chiang, and A. Sabharwal, "An axiomatic theory of fairness in network resource allocation," in *IEEE INFOCOM*, 2010, pp. 1–9.
- [13] S. Ali and M. Zeeshan, "A utility based resource allocation scheme with delay scheduler for LTE service-class support," in *IEEE Wireless Communications and Networking Conference*, 2012, pp. 1450–1455.
- [14] B. Liu, H. Tian, and L. Xu, "An efficient downlink packet scheduling algorithm for real time traffics in LTE systems," in *IEEE Consumer Communications and Networking Conference*, 2013, pp. 364–369.
- [15] Y. C. Wang, S. R. Ye, and Y. C. Tseng, "A fair scheduling algorithm with traffic classification in wireless networks," *Computer Communications*, vol. 28, no. 10, pp. 1225–1239, 2005.
- [16] Y. C. Wang, Y. C. Tseng, and W. T. Chen, "MR-FQ: a fair scheduling algorithm for wireless networks with variable transmission rates," *Simulation*, vol. 81, no. 8, pp. 587–608, 2005.
- [17] 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," *International Journal on Communication Systems*, vol. 23, no. 2, pp. 213–230, 2010.
- [18] P. Dent, G. E. Bottomley, and T. Croft, "Jakes fading model revisited," *Electronics Letters*, vol. 29, no. 13, pp. 1162–1163, 1993.
- [19] LTE-Sim. [Online]. Available: <http://telematics.poliba.it/index.php/en/lte-sim>
- [20] Y. C. Wang and C. A. Chuang, "Efficient eNB deployment strategy for heterogeneous cells in 4G LTE systems," *Computer Networks*, vol. 79, no. 14, pp. 297–312, 2015.