IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received August 13, 2019, accepted September 11, 2019, date of publication September 16, 2019,
date of current version September 27, 2019.

*Digital Object Identifier 10.1109/ACCESS.2019.2941451*

# A Novel Scheduling Algorithm to Improve Average User Perceived Throughput for LTE Systems

**YAKE LI, XINPENG FANG, AND SIYU LIANG**
School of Aerospace Science and Technology, Xidian University, Xi'an 710071, China

Corresponding author: Xinpeng Fang (xinpengfang@163.com)

This work was supported in part by the National Natural Science Foundation of China under Grant 61703326, in part by the Fundamental Research Funds for the Central Universities under Grant JB181306, and in part by the Innovation Fund of Xidian University.

**ABSTRACT** In recent years, the network operators have paid more attention to user experience. The design of scheduling algorithm is a great challenge in long term evolution (LTE) system to improve the overall system Quality of Experience (QoE). Many papers have been developed to solve the scheduling problem for the full buffer traffic. In this paper, a new scheduling algorithm named percent proportional fair (PPF) is proposed for the reality burst traffic. To evaluate the user experience, we introduce two measure indicators, average user perceived throughput (UPT) and user perceived throughput-cut (UPT-cut). In order to improve the average UPT with fairness guarantee, we design the scheduling priority by considering the ratio of the amount of transmitted data to the total amount of burst data. At the same time, an allocation matrix is introduced so as to avoid bandwidth waste. Finally, comparing the PPF scheduler with the classical proportional fair (PF) scheduler and fair allocation high throughput (FAHT) scheduler, the simulation results show the improvement of average UPT while ensuring fairness among the users. Besides, the performance of PPF scheduling is analyzed in various system scenarios.

**INDEX TERMS** Quality of Experience, LTE systems, percent proportional fair, user perceived throughput, scheduling priority.

## I. INTRODUCTION

The wireless communication technology that supports high-speed download will be commercially available in the near future. However, the network-centric operation can not adapt to the increasing demands of customers. A poor user experience leads to the loss of customers. Hence, mobile network operators are more concerned about the improvement of Quality of Experience (QoE). For the first time, the network management index, scheduled internet protocol throughput, is introduced for the non-full (burst) traffic to measure user experience in [1]. The downlink scheduling plays a vital role in radio resource management. Thus, it is necessary for the industry and academia to design schedulers so as to improve the QoE.

The common scheduling schemes are round robin (RR), maximum carrier to interference (Max C/I), and proportional fair (PF) [2]. PF scheduling algorithm for orthogonal frequency division multiple (OFDM) system is proposed firstly in [3]. PF is a commonly used scheduling algorithm for the industry because it achieves a good trade-off between system throughput and user fairness.

Later on, several PF variations have been proposed to guarantee QoE for different requirements. Largest weighted delay first (LWDF), exponential-proportional fair (EXP-PF), and log-proportional fair (LOG-PF) rule all mainly focus on the delay and channel awareness for real-time operating system to guarantee QoE [4]. The authors in [5] introduce a fair allocation high throughput (FAHT) algorithm by using geometric mean with a faster convergence for user throughput to improve the throughput while ensure the fairness. In [6], the authors design the scheduling policy to meet QoE requirements and maintain fairness for the heterogeneous traffic in overload states. In [7], the waiting time of the real-time traffic is specially considered while designing the scheduling criterion. In [8], the authors design the scheduling priority

The associate editor coordinating the review of this manuscript and approving it for publication was Kanapathippillai Cumanan.

  

according to the estimation of the amount of video data stored in the player buffer, which avoids the playback interruptions.

According to the literature review [4], the authors summarize that most schedulers are designed from the following aspects: (1) information of transmission queues; (2) users'channel quality; (3) resources allocation history; (4) buffer state; (5) quality of service requirements. Many other factors are considered in the design of schedulers for wireless communication, for example, the operator's revenue, the information content of traffic, and so on. The scheduling module in [9] combines traffic amount, channel condition, with user level that achieves a good trade-off between performance and operator's revenue. In [10], the authors take video contents and its features into account in the process of packet scheduling. In [11], the authors design the scheduling priority according to the data importance for the Machine to Machine (M2M) communication in long term evolution (LTE) networks.

There are several performance metrics to evaluate QoE. System throughput, fairness index, packet loss rate, and packet dalay are the metrics to evaluate the QoE of the end users in [12]. In [13], the average user throughput is seen as the performance metric to evaluate the proposed coordinated multi-point (CoMP) scheduling approach. The authors in [14], [15] choose the peak signal-to-noise ratio (PSNR) for video application to evaluate the QoE. Another widely used metric is the mean opinion score (MoS) [16]. The authors adopt MoS to evaluate the quality perceived by users in [17]. An accumulative QoE model is presented based on Video quality, stability, and playback continuity in [18].

The following performance metrics are usually considered for burst traffic models [19]: (1) mean, 5, 50, 95% user throughput ( user throughput = amount of data / time needed to download data); (2) served cell throughput ( served cell throughput = total amount data for all users / total amount of observation time / number of cells); (3) harmonic mean normalized cell throughput; (4) Normalized cell throughput. The first metric aims at the single user. The second metric is meaningless if the total amount data for the user is fixed, since the served cell throughput is constant regardless the scheduler when the system load is not high. The third and the fourth ones are the simple extensions of the second one. Thus, it is significant to define a effective metric for burst traffic to measure QoE for the cell users.

In this work, we introduce two new indices average user perceived throughput (UPT) and average user perceived throughput-cut (UPT-cut) to measure QoE for the cell users. The two indices consider the cell user throughput, waiting time and transmission time for the burst traffic. Then we propose a novel scheduling algorithm to improve the average UPT/UPT-cut with fairness guarantee. As it has shown in [20] that internet traffic exhibits high variability: most flows are short, while more than 50% of the bytes are carried by less than 5% of the largest flows. Therefore, we create a burst traffic model with long bursts and short bursts. The contributions are summarized as follows:

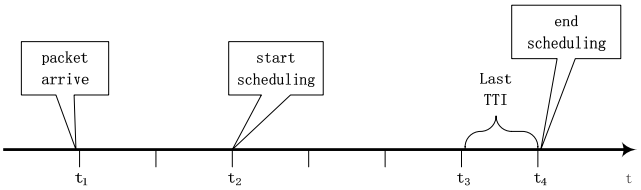


**FIGURE 1.** The time axis for the burst scheduling process.

- Two performance metrics, average UPT/UPT-cut at cell level, are introduced to evaluate user experience. For simplicity, the two novel metrics are named UPT/UPT-cut in this paper.
- To improve average UPT/UPT-cut at cell level, we propose a PPF scheduling algorithm, which takes into account the ratio of the amount of transmitted data to the total amount of burst data.
- The resources allocation matrix is added to the scheduling priority to utilize the limited resources more effectively.
- We compare the PPF with PF and FAHT in terms of UPT, UPT-cut, and fairness index. At the same time, we study the impact of tuning parameters on PPF performance and analyze the effectiveness of PPF.

The rest of the work is organized as follows: Section II gives necessary definitions and explanations in this paper. The PPF scheduling process and the design of scheduling priority are described in Section III. In Section IV, we present the simulation results, followed by conclusions drawn in Section V.

## II. PRELIMINARY WORK

In this section, we present two definitions named average UPT and average UPT-cut. Meanwhile, the physical resource block (PRB) usage and the measuring standard of fairness are introduced. To explain the definitions clearly, a time axis for the burst scheduling process is given in Fig. 1. A data burst scheduling begins at the point in time when the first transmission begins after a Packet Data Convergence Protocol Service Data Units becomes available for transmission. The data burst ends at the point in time when transmissions are successfully completed. In Fig. 1, the waiting time is from $t_1$ to $t_2$, the transmission time is from $t_2$ to $t_4$ and the scheduling time is from $t_1$ to $t_4$.

*Definition 1 (Average UPT-Cut):* The average UPT-cut at cell level is defined as follows

$$v^c = \sum_{i=1}^{N} b_i^c \Big/ \sum_{i=1}^{N} t_i^c, \tag{1}$$

where $b_i^c$ is the downlink traffic volume of transmitted burst data besides last transmission time interval (TTI) of user $i$, $t_i^c$ is the data transmit duration except the last TTI before the downlink buffer of user $i$ is empty, which represents the time period from $t_2$ to $t_3$ in Fig 1, and $N$ is the total number of users in the time period during which the measurement is performed.

*Remark 1:* Data bursts that are large enough require transmissions to be split across several TTIs. Throughput in the last TTI are usually negligible due to a small amount of data by 3GPP [1]. Therefore, we also eliminate the bits of last TTI at Definition 1.

*Definition 2 (Average UPT):* The average UPT at cell level is defined as follows

$$v = \sum_{i=1}^{N} b_i \Big/ \sum_{i=1}^{N} t_i, \tag{2}$$

where $b_i$ is the downlink traffic volume of transmitted burst data of user $i$, and $t_i$ is the sum of waiting time and data transmit duration of user $i$, which is related to the time period from $t_1$ to $t_4$ in Fig. 1.

*Remark 2:* In general, the operators are willing to use the average UPT-cut at cell level as a key performance index (KPI) because of the small proportion of the head time to total time, i.e., $\left(\sum_{i=1}^{N} t_i - \sum_{i=1}^{N} t_i^c\right) \Big/ \sum_{i=1}^{N} t_i$ is small, but we also consider the average UPT at cell level in this paper because it could reflect user experience more effectively.

*Remark 3:* The UPT/UPT-cut in this paper can be expressed as follows:

1) UPT = total amount of data for all users / total scheduling time for all users;
2) UPT-cut = total amount of data (except the data for the last TTI) for all users / total transmission time (except the last TTI) for all users.

*Definition 3 (RB Usage):* The system load with the burst traffic can be measured by PRB usage. The PRB usage for the downlink can be defined as following [1]

$$M(T) = \frac{M1(T)}{P(T)} * 100\%, \tag{3}$$

where $M1(T)$ denotes a count of PRBs number used for the downlink transmission, $P(T)$ represents all available PRBs number during time period $T$, and $T$ is the time period during which the measurements are performed.

*Definition 4 (Fairness Index):* The fairness index among users is calculated according to Rai Jain index [21]

$$FI = \left[\sum_{i=1}^{N} v_i\right]^2 \Big/ \left[N \sum_{i=1}^{N} (v_i)^2\right], \tag{4}$$

where $v_i$ is the throughput during the observation time for full buffer traffic. In this paper, for the burst traffic, let $v_i$ represent the throughput of user $i$ (i.e.$v_i = \frac{b_i}{t_i}$). It is apparently known that the scheduling approach is absolutely fair if the value of fairness index is 1.

## III. THE PPF SCHEDULING ALGORITHM

### A. SCHEDULING PROCESS IN LTE SYSTEM

LTE standard uses Orthogonal Frequency Division Multiplexing technology at the physical layer to modulate the symbols for each TTI. The Scheduling mechanism is two dimensional for the LTE system, which means that resources

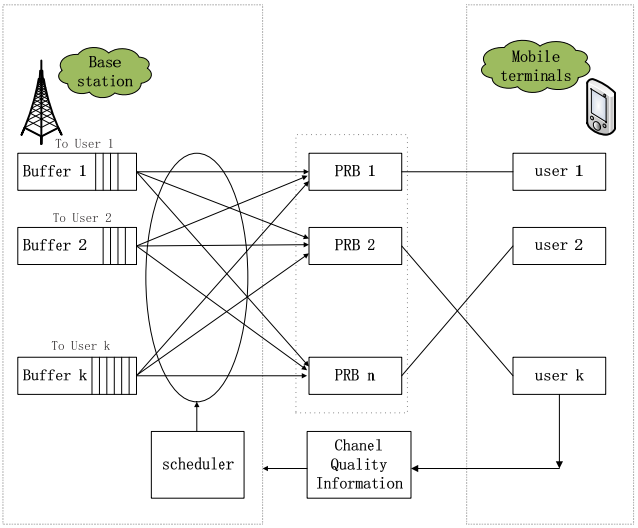


**FIGURE 2.** Scheduling process of an OFDM system.

are shared across the users in time domain and frequency domain. In this paper, we pay attention to the frequency scheduling. The scheduling process for the downlink OFDM system is exhibited in Fig. 2. The scheduler is placed at the base station (BS) to allocate radio resources (i.e. PRBs) among users. If a user's terminal sends a data transfer request to the BS, the user will be active. Meanwhile, the BS will build a buffer zone for the user, which is a relay station to store data temporarily. When the BS receives the channel quality information (CQI) reported by the users, the scheduler at the BS will allocate the PRBs to users according to the user priority (see Section III-B) at each scheduling epoch. A PRB can be used by no more than one user and a user can use multiple PRBs to transmit data during one scheduling epoch. Let $\mathcal{N}$ denote the users set, $\mathcal{K}$ denote the active users set, and $\mathcal{M} = \{1, 2, \ldots, M\}$ denote the PRBs set. The key mathematical symbols are listed in Table1. Bold symbols represent matrices or vectors. Besides, "/" denotes "or", and "/" denotes a division symbol.

### B. SCHEDULING PRIORITY

Suppose transaction 1 and transaction 2 arrive simultaneously of size $F_1$ and $F_2$ bits ($F_1 > F_2$), respectively. We also assume that the network has a fixed network capacity of $M$ bits per TTI. Next, we consider two scheduling policies: policy 1- transaction 1 has a higher priority than transaction 2, which means that the longest burst is served first; policy 2- transaction 2 has a higher priority than transaction 1, which means that the shortest burst is served first. Then the UPT of the two policies can be calculated as:

$$v^{P1} = \frac{F_1 + F_2}{\frac{F1}{M} \times 2 + \frac{F_2}{M}} = \frac{M(F1 + F_2)}{2F_1 + F_2} \tag{5}$$

$$v^{P2} = \frac{F_1 + F_2}{\frac{F2}{M} \times 2 + \frac{F_1}{M}} = \frac{M(F1 + F_2)}{2F_2 + F_1} \tag{6}$$

**TABLE 1. Allocation data(bits).**

| Comment | PRB 1 | PRB 2 | PRB 3 | Remaining data |
|---|---|---|---|---|
| UE 1 | 5 | 8 | 2 | 10 |
| UE 2 | 4 | 2 | 6 | 15 |
| UE 3 | 3 | 6 | 7 | 5 |
| UE 4 | 2 | 3 | 5 | 8 |

By combining (5) and (6) with the assumption $F_1 > F_2$, we can obtain $v^{P1} < v^{P2}$. According the above analysis, policy 2 apparently performs better than policy 1 in term of UPT. First, the observation motivates a scheduling algorithm that gives a higher priority to the user with the less remaining data. In fact, the actual network capacity is not fixed due to users' joining and leaving and the users' time-variable channel. Second, the transmission time required for a long burst should be longer than a short burst. In order to maintain fairness between the long bursts and short bursts, we consider the ratio of transmitted data to the total burst data volume rather than the remaining data. Third, history transmission rate should be taken into account for fairness guarantee among long bursts and short bursts.

The PPF scheduling algorithm is put forward on the basis of the above three motives. The PPF scheduling is implemented by selecting the user from the active user set who has the following highest priority for the $m$th PRB at each scheduling epoch,

$$K = \frac{r_{k,m}}{R_k}(1 + \beta x_k^{\alpha})A_m, \quad \forall m \in \mathcal{M} \tag{7}$$

where $r_{k,m}$ is the instantaneous supported data rate of user $k$ for PRB $m$, $R_k$ denotes the low-pass filtered average data rate of user $k$, and $x_k$ denotes the ratio of transmitted data volume to total burst data volume for user $i$. $x_k$ can be expressed as

$$x_k(t) = \frac{Q_k(t)}{d_k} \tag{8}$$

where $Q_k$ denotes the transmitted data volume for the user $k$ and $d_k$ denotes the total burst data volume for user $k$. $\alpha$ and $\beta$ represent two tuning parameters investigated in detail later. $A = [A_1, A_2, \ldots, A_m, \ldots, A_M]$ denotes the allocation matrix with the form as

$$A_{K\times M} = (a_{ij}), \quad a_{ij} \in \{0, 1\}, \tag{9}$$

where $n$ denotes the number of the current active users and $m$ denotes the number of the PRBs in the system. Next, we present the construction method of allocation matrix $A$. Suppose that supported data rate in each PRB and the remaining data for user $i$ are $\boldsymbol{b^K} = (b_1^i, b_2^i, \ldots, b_M^i)$ and $R^i$, respectively. Without loss of generality, sort the vector $\boldsymbol{b^k}$ in descending order as following

$$b_{q_1}^i \geq b_{q_2}^i \geq \ldots \geq b_{q_M}^i \tag{10}$$

where $b_{q_i}^i \in \{b_1^i, b_2^i, \ldots, b_M^i\}$. If the following inequalities are satisfied

$$\begin{cases} b_{q_1}^i + b_{q_2}^i + \ldots + b_{q_s}^i & \geq R^i, \\ b_{q_1}^i + b_{q_2}^i + \ldots + b_{q_{s-1}}^i & < R^i \end{cases} \tag{11}$$

then we have

$$\forall i \in \mathcal{K}, a_{i,j} = \begin{cases} 1, & j \in \{q_1, q_2, \ldots, q_s\} \\ 0, & otherwise \end{cases} \tag{12}$$

$a_{ij}$ equals to one, which means that the user $i$ will compete the $j$th PRB with other active users, whereas $a_{ij}$ equals to zero, which means the $j$th PRB will be not allocated for the user $i$ regardless of the scheduling rules. As an example, suppose a system with 3 PRBs and 4 active users. The supported data rate in each PRB and the remaining data for each user are listed in the Table 1, then the allocation matrix can be written as

$$A = \begin{pmatrix} 1 & 1 & 0 \\ 1 & 1 & 1 \\ 0 & 0 & 1 \\ 0 & 1 & 1 \end{pmatrix}. \tag{13}$$

$R_i$ gets updates at every TTI in accordance with the following equation

$$R_i(t) = (1 - \frac{1}{t_c}) \times R_i(t-1) + \frac{1}{t_c} \times r_i(t-1), \tag{14}$$

where $r_i$ denotes the actual received data rate of user $i$ at time $t-1$, and $t_c$ is the time window.

**Algorithm 1** PPF Scheduling Policy

**Input:** *Initial:* $Q(0) = 0$; $\mathcal{N} = \{1, 2, 3, \ldots, N\}$. For each time slot $t$, $r(t) = (r_{i,j})_{N\times M}$; $Q(t) = [Q_1, Q_2, \ldots, Q_N]$; $R(t) = [R_1, R_2, \ldots, R_N]$ ; $\mathcal{K} = \varnothing$;
**Output:** $I(t) = (I_{i,j})_{K\times M}$;
1: for $n = 1$ to $N$ do
2: if $Q_n > 0$ then
3: user $n$ join in the active user set, $n \in \mathcal{K}$
4: end for
5: for $m = 1$ to $M$ do
6: chose the user $k^*$ according to (7);
7: if $k = k^*$ then
8: $I_{k,m} = 1$;
9: else
10: $I_{k,m} = 0$
11: end if
12: end for
13: for $k = 1$ to $K$ do
14: update $R_k(t)$ according to (14);
15: update $Q_k(t)$ according to $Q_k(t+1) = Q_k(t) + \sum_{m=1}^{M} I_{k,m} r_{k,m}$;
16: end for
17: **return** $I(t) = (I_{i,j})_{K\times M}$, $R(t+1)$, and $Q(t+1)$.

### C. ANALYSIS OF THE PROPOSED SCHEME

The scheduling algorithm has little impact on the total traffic volume (i.e., the numerator part of average UPT/UPT-cut expression) because of the little transmission requirement, especially when the system loads are light and medium.

**TABLE 2.** Key mathematical symbols for the network.

| Notation | Definition | Notation | Definition |
|---|---|---|---|
| $\mathcal{N}$ | number of user | $n$ | index of user n |
| $\mathcal{K}$ | number of active user | $k$ | index of active user $k$ |
| $\mathcal{M}$ | number of PRB | $m$ | index of PRB m |
| $\boldsymbol{Q(t)}$ | transmitted data volume vector at time slot $t$ | $Q_k(t)$ | transmitted data volume for user $k$ |
| $\boldsymbol{R}(t)$ | low-pass filtered average data rate vector | $R_k(t)$ | low-pass filtered average data rate for user $k$ |
| $\boldsymbol{r}$ | instantaneous supported data rate matrix | $r_{k,m}$ | instantaneous supported data rate of user $k$ for PRB $m$ |
| $\boldsymbol{I}$ | PRB allocation matrix | $\boldsymbol{A}$ | allocation matrix |
| $d_k$ | total burst data volume for user $k$ | | |

The average UPT/UPT-cut can be improved by decreasing the sum of effective transmission time (i.e., the denominator part of average UPT/UPT-cut expression). The factor $x_i$ plays an important role in PPF scheduling. The user with the higher transmit complete ratio will have a higher priority, then the transmission will be completed as soon as possible, which means the sum time period from $t_2$ to $t_3$ for all users is shortened in Fig. 1. Thus the average UPT/UPT-cut is improved on the whole.

The allocation matrix $A$ shows that the users only compete the PRBs that can meet their requirements via as few resources as possible. With the table 1 for example, the user 1 only compete the PRB 1 and PRB 2, because the transmission on PRB 1 and PRB 2 can satisfy the requirement of user 1. On the other hand, PRB3 must be allocated to one from other active users (i.e.,user 2, 3, and 4), which can achieve a higher data transmission. Hence, the allocation matrix $A$ makes the best use of resources by allowing each PRB transmit as much data as possible.

We have selected an exponential function ($y = 1 + \beta x^{\alpha}$) because the dependent variable $y$ approaches to the maximum value smoothly with the independent variable $x$ increasing, where $x$ is from 0 to 1. The exponential functions with $\beta = 4$ are shown in Fig. 3. As it is shown in Fig. 3, $\alpha$ controls the slope of the curve and $\beta$ determines the range of $1 + \beta x^{\alpha}$. From the Fig. 3, we can observe: for $\forall x \in (0, 1]$, (a) if $\alpha \to 0$, then $y \to 1 + \beta$; (b) if $\alpha \to +\infty$, then $y \to 1$; (c) if $\beta \to 0$, then $y \to 1$; (b) if $\beta \to +\infty$, then $y \to +\infty$. To improve the UPT while ensure the fairness, the users' priority should be set higher if the ratio of the amount of transmitted data to total burst data volume is closer to 1. The ideal exponential function satisfy: (a)the curve for the ideal exponential function is smooth, which means the slope of the curve changes uniformly; (b) $y$ is larger if $x$ is closer to 1, which means the slope of the curve changes rapidly especially when x is almost 1. Compared the curves in the Fig. 3, we can observe that the optimal value for $\alpha$ is neither too small or too large. Similarly, the conclusions are the same for $\beta$. Hence, the PPF scheduling algorithm is convergent when $\beta$ approaches to 0 and $\alpha$ approaches 0 or infinity, which means the PPF scheduler is equal to the PF scheduler. It can be described in the mathematical term as following: for $\forall j$, $x \in [0, 1)$, if $\alpha \to +\infty$õr $0, or\beta \to 0$, then

$$\arg\max_i \frac{r_{i,j}}{R_i}(1 + \beta {x_i}^{\alpha})A_j = \arg\max_i \frac{r_{i,j}}{R_i}A_j \qquad (15)$$

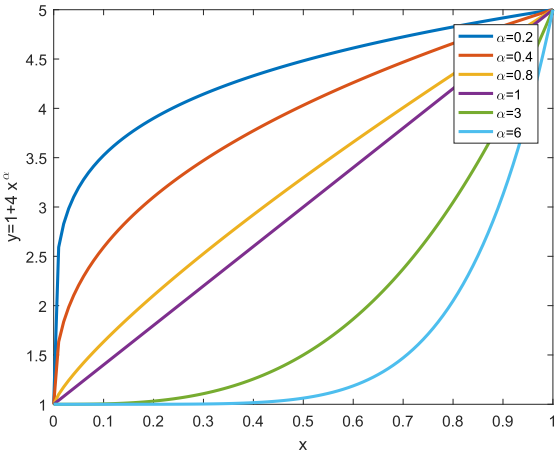


**FIGURE 3.** $y = 1 + 4x^{\alpha}$ with different $\alpha$ ($\beta = 4$).

*Remark 4:* The users' priority should be set higher as the process of transmission goes on. For a burst case, $x$ changes gradually from 0 to 1 as the process of transmission goes on. However, for a multi bursts case, $x$ may not satisfy the property mentioned above. For example, half of the data has been transmitted and another burst arrives, then $x$ changes from 0.5 to a value less than 0.5.

## IV. SIMULATION RESULTS

In this section, in order to investigate the effectiveness of the proposed algorithm, we consider an OFDM-based LTE downlink system. We analyze the performance of the PPF, PF and FAHT algorithms in terms of UPT, UPT-cut, as well as the fairness index. We consider the allocation matrix for the three algorithms in the simulations, because it is not emphatically analyzed in this work. We also assess the PPF performance with different tuning parameters and ratios of the number of long bursts to the number of short bursts.

In this paper, gain shows the superiority of the PPF/ FAHT scheme compared with the traditional PF algorithm. Gain is calculated as follows:

$$Gain = \frac{V^{PPF/FAHT} - V^{PF}}{V^{PF}} * 100\%, \qquad (16)$$

where $V^{PPF}/V^{FAHT}/V^{PF}$ denotes the UPT/ UPT-cut by PPF/ FAHT/ PF algorithm.

**TABLE 3. Simulation parameters.**

| Parameter | Value/Comment |
|---|---|
| Cell topology | Single cell |
| Number of sector | 3 |
| PRB number per sector | 10 |
| Sector Radius | 250m |
| Height of BS | 35m |
| Number of users per sector | Change from 10 to 52 |
| UE distribution | Random |
| Path Loss Model | $34.5 + 35 * log10(d)$(dB) |
| Shadow fading standard deviation | 8 dB |
| Thermal noise density | -174dBm/Hz |
| System and link level mapping interface | EESM |
| Target Bit-Error-Rate | 10% |
| PRB bandwidth | 180kHz |
| Horizontal antenna pattern | 70° |
| Base station antenna gain | 15 dBi |
| Mobile station antenna gain | 0 dBi |
| Power allocation | equally distributed |
| $t_c$ | 5 |
| Simulation time | 100 (TTI) |
| Number of users/bursts in cell | 10, 13, 16 , ... , 52 |

**TABLE 4. Traffic model& performance metric.**

| Parameter | Comment |
|---|---|
| long bursty traffic | log-normal distribution of bursts at MAC layer ($\mu$=3000 bytes, max=6000 bytes, $\sigma$=1000 bytes) |
| short bursty traffic | log-normal distribution of bursts at MAC layer ($\mu$=200 bytes, max=400 bytes, $\sigma$=100 bytes) |
| performance metric | user perceived throughput at cell level (UPT); user perceived throughput-cut at cell level (UPT-cut); Fairness index |

### A. SIMULATION MODEL

We assume that a cell consists of three sectors and one directional antenna is deployed in each sector. The channel model considers path loss and shadow-fading. Shadow fading is normal distribution with mean 0 and variance 8. To reduce the computation load, we set 10 PRBs in each sector. Users are uniformly distributed at the cell and users arrive uniformly and randomly during the simulation time. The number of users in each sector varies from 10 to 52, corresponding to the RB usage that ranges from approximately 15% to 75%. The system load (i.e., RB usage) is increased with the increase of the number of users. Besides, we assume that every user has one burst, so the number of users is restricted to be equal to the total number of bursts. More detailed simulation parameters are listed in Table 3.

The burst traffic model and performance metrics are provided in Table 4. There are two types of bursts: long bursts and short bursts. For each type of bursts, the size of bursts is log-norm distribution with expectation $\mu$ and variance $\sigma$. For the long bursts, the expectation and variance of the size are 3000 and 1000, and the max size is 6000. For the short bursts, the expectation and variance of the size are 200 and 100, and the max size is 400. A short burst can be transmitted completely during a little TTIs ( about 2-5TTIs) under medium system load in this simulation model. The default ratio of the number of long bursts to the number of short bursts is 3:7. The performance of the proposed algorithm is analyzed by the statistic method because the arrivals, sizes of bursts, and channel states of users are random. The simulation time is 100 TTIs in each simulation process. We repeat the simulation process 50 times for a fixed number of users, and get an average of the 50 experimental results (i.e.,UPT, UPT-cut, fairness index and RB usage).

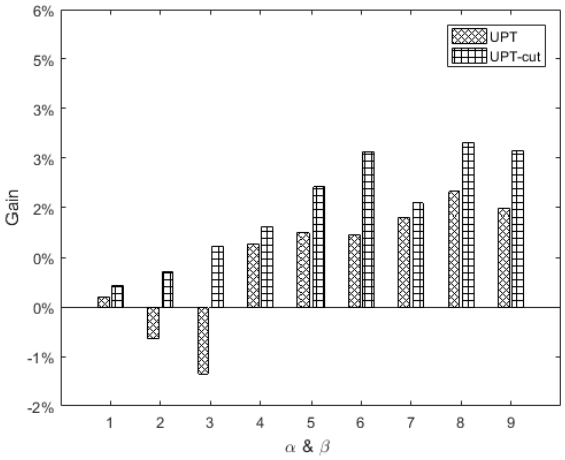


**FIGURE 4. Gain with various $\alpha$, $\beta$ (RB usage = 28%).**

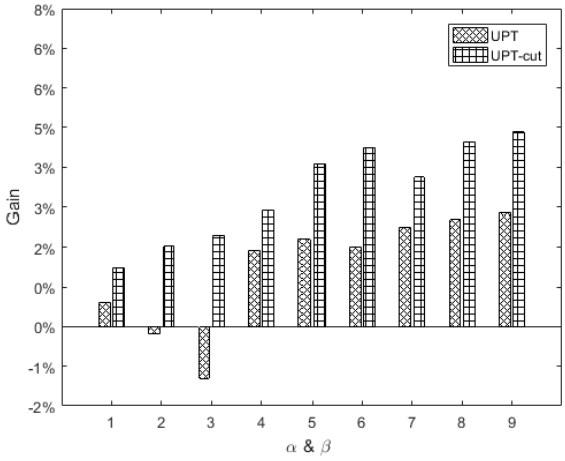


**FIGURE 5. Gain with various $\alpha$, $\beta$ (RB usage = 36%).**

### B. SIMULATION RESULTS AND ANALYSIS

We set $(\alpha, \beta) \in \{(1, 2), (1, 4), (1, 6), (2, 2), (2, 4), (2, 6), (3, 2), (3, 4), (3, 6)\}$ (total nine combinations) to study the performance of PPF algorithm. Fig. 4-Fig. 7 illustrate the gains of the investigated PPF in terms of UPT and UPT-cut with RB usages 28%, 36%, 57%, and 72%, respectively. The four RB usages correspond to inferiority, low, medium, and high system loads. In Fig. 4-Fig. 7, the labels of X-axis denote the nine different tuning parameters, which correspond to $(1, 2), (1, 4), (1, 6), (2, 2), (2, 4), (2, 6), (3, 2), (3, 4), (3, 6)$ From the Fig. 4-Fig. 7, we can see that the gain in term of UPT/UPT-cut is almost the same with $\beta = 4$ and $\beta = 6$ regardless of the system loads. These results are also similar with $\alpha = 2$ and $\alpha = 3$. Compared with

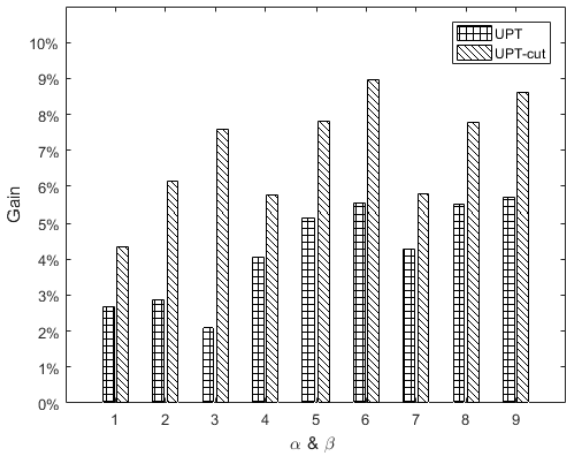


**FIGURE 6.** Gain with various $\alpha$, $\beta$ ( RB usage = 57%).

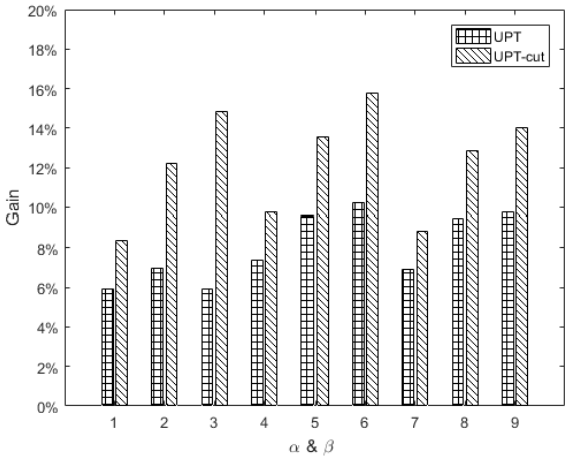


**FIGURE 7.** Gain with various $\alpha$, $\beta$ (RB usage = 72%).

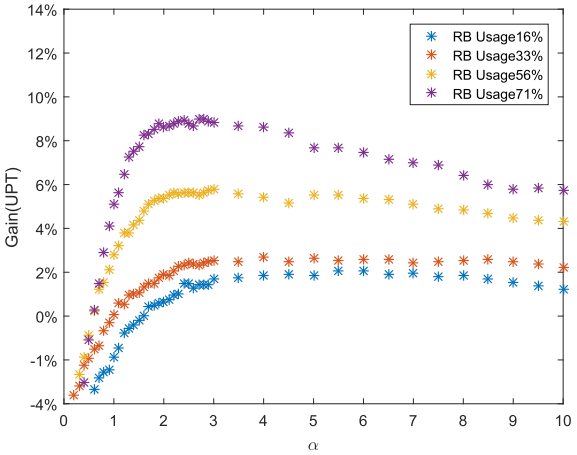


**FIGURE 8.** UPT gain with various $\alpha$ for PPF ($\beta = 4$).

PF, little gain occurred for PPF under low system loads; however, considerable gain can be observed under middle and high system loads regardless of the tuning parameters. In addition, the higher the system load, the more the gain.

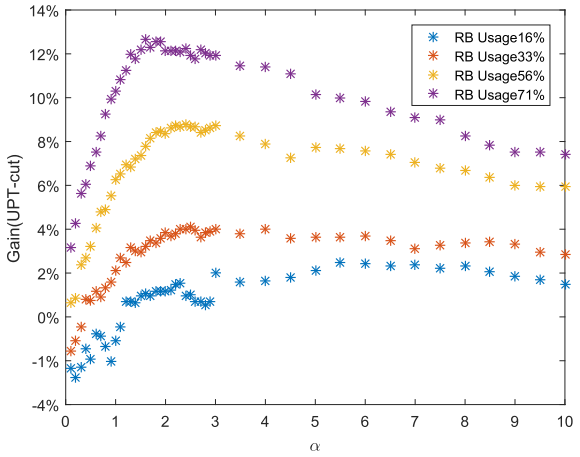


**FIGURE 9.** UPT-cut gain with various $\alpha$ for PPF ($\beta = 4$).

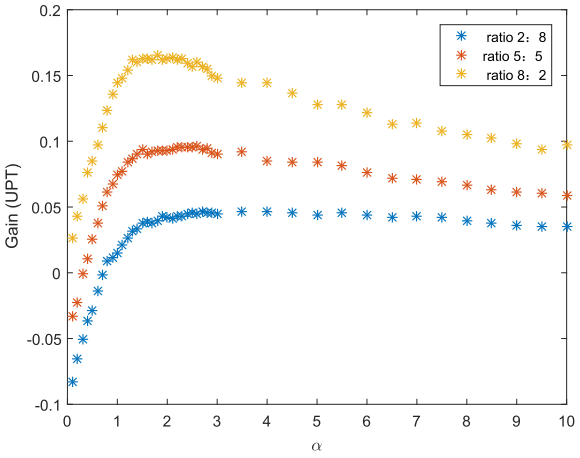


**FIGURE 10.** UPT gain with various burst size ratios for PPF ($\beta = 4$).

These results indicate that the room for improvement is larger if the system load is higher (i.e., there are more active users). Taking above results into account, we set $\beta = 4$ and $\alpha = \{0.5, 1, 1.5, 2, 2.5, 3, 4, 5, 6, 7, 8, 9, 10\}$ to investigate the impact of $\alpha$ on PPF performance. Fig. 8 and Fig. 9 demonstrate the UPT gain and UPT-cut gain with various $\alpha$ for different system loads. From the Fig. 8 and Fig. 9, we can see that PPF outperforms both on UPT and UPT-cut with $\alpha = 2$ and $\beta = 4$. Fig. 10 and Fig. 11 display the UPT gain and UPT-cut gain with various $\alpha$ for different ratios of the number of long bursts to the number of short bursts. As we can see from the Fig. 10 and Fig. 11, the UPT gain and UPT-cut gain are higher when the number of long bursts is greater. Besides, the UPT/UPT-cut gain reaches the approximate optimal point at $\alpha = 2$ for the three different ratios.

Next, we set $\alpha = 2$ and $\beta = 4$ to evaluate the performance under PPF. Table. 5 demonstrates the confidence intervals of the UPT gain and UPT-cut gain by PPF with various system loads (RB usage). From the Table. 5, we can observe that the results widely spread around the average. Compared PPF

**TABLE 5.** Confidence intervals of the UPT gain / UPT-cut gain (%) by PPF with various RB usages.

| Scene & Gain \ Confidence level | | 68% | 90% | 95% | 99% | average level |
|---|---|---|---|---|---|---|
| RB usage 28% | UPT gain | (1.36,1.87) | (-2.59,5.83) | (-3.4,6.63) | (-4.97,8.2) | 1.62 |
| | UPT-cut gain | (2.36,3.1) | (-3.4,8.86) | (-4.57,10.03) | (-6.86,12.31) | 2.73 |
| RB usage 36% | UPT gain | (2.09,2.66) | (-2.31,7.05) | (-3.21,7.95) | (-4.95,9.7) | 2.37 |
| | UPT-cut gain | (4.09,4.97) | (-2.7,11.75) | (-4.08,13.14) | (-6.77,15.83) | 4.53 |
| RB usage 57% | UPT gain | (4.93,5.67) | (-0.75,11.35) | (-1.91,12.51) | (-4.16,14.77) | 5.3 |
| | UPT-cut gain | (7.72,8.68) | (0.27,16.12) | (-1.24,17.64) | (-4.2,20.59) | 8.2 |
| RB usage 72% | UPT gain | (9.55,10.31) | (3.71,16.16) | (2.51,17.35) | (0.19,19.67) | 9.93 |
| | UPT-cut gain | (13.78,14.77) | (6.08,22.47) | (4.51,24.04) | (1.46,27.09) | 14.28 |

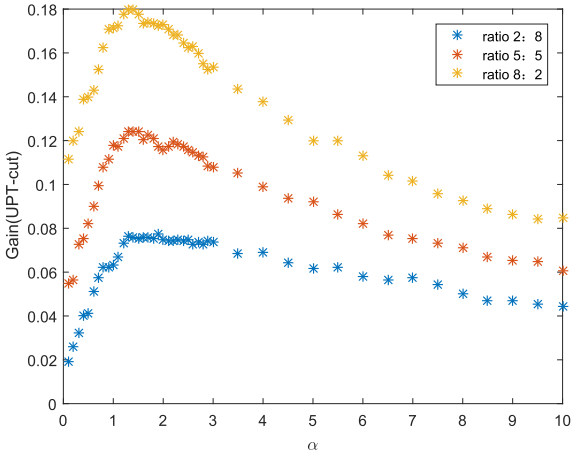


**FIGURE 11.** UPT-cut gain with various burst size ratios for PPF($\beta = 4$).

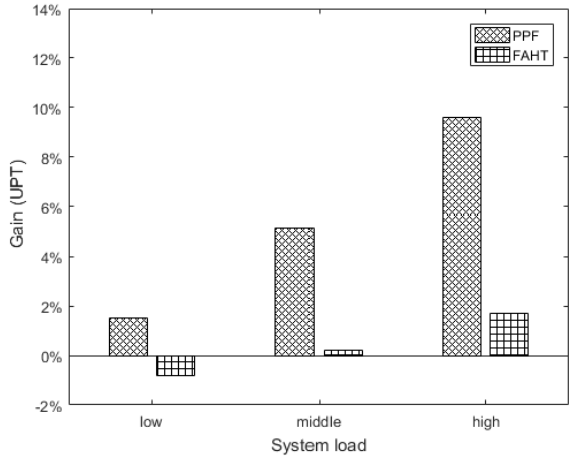


**FIGURE 12.** UPT gain with various system loads under PPF ($\alpha = 2$, $\beta = 4$) and FAHT.

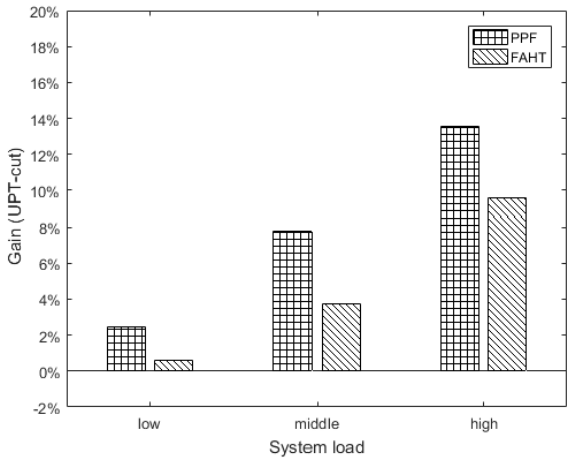


**FIGURE 13.** UPT-cut gain with various system loads for PPF ($\alpha = 2$, $\beta = 4$) and FAHT.

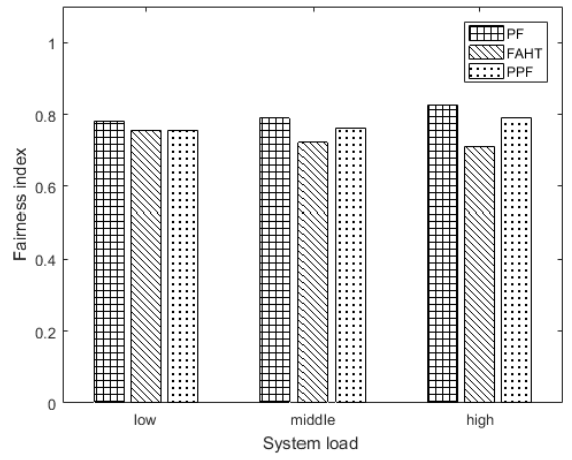


**FIGURE 14.** Fairness index with various system loads for PPF ($\alpha = 2$, $\beta = 4$), PF, and FAHT.

and FAHT with PF, Figs.12, 13, and 14 demonstrate the UPT gain, UPT-cut gain, and fairness with various system loads. The RB usages are 28%, 57%, and 74% under the low, middle, and high system loads, respectively. We observe that the gain for PPF is higher than that of FAHT in term of UPT/UPT-cut regardless of the system loads. It is obvious that the fairness indexes for PF and PPF are very close under low and middle system loads. The fairness of PF is optimal, while worst for FAHT under the high system load. Fig. 15 displays the fairness indexes with various RB usages by PPF, PF and FAHT. The fairness indexes for the three scheduler are very close while PF achieves the maximum fairness index regardless of the RB usage. Fairness indexes with the long bursts and short bursts for the three scheduler are compared in Fig. 16. We observe that fairness indexes of FAHT for long bursts and short bursts are nearly identical, while the difference between the fairness index for long bursts and the

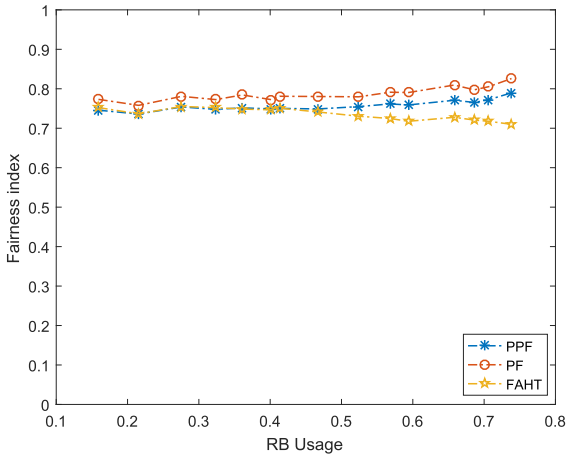


**FIGURE 15.** Fairness index with various RB usage for PPF ($\alpha = 2$, $\beta = 4$), PF, and FAHT.

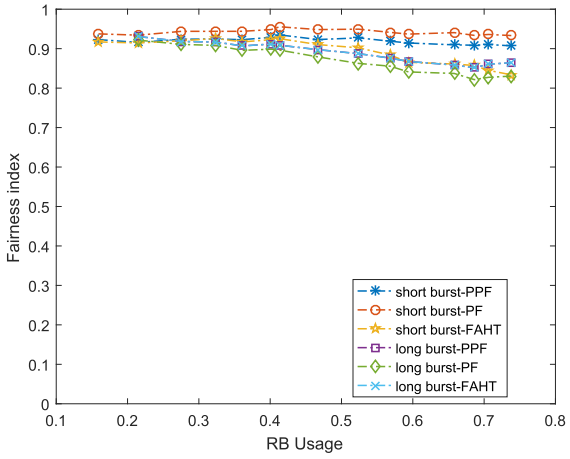


**FIGURE 16.** Fairness index with various RB usage, $(\alpha, \beta) = (2, 4)$.

fairness index for short bursts for PF is biggest. In general, compared with PF, PPF and FAHT achieve a better balance between the long bursts and short bursts.

In summary, PF achieves more fairness but not perform well in term of UPT/UPT-cut. FAHT creates better performance in UPT/UPT-cut, and the optimal fairness between long bursts and short bursts. The PPF scheduling algorithm can yield optimal UPT/UPT-cut while ensuring fairness among the users.

## V. CONCLUSION

We have defined the average UPT/UPT-cut at cell level for the burst traffic to measure the QoE. A PPF scheduling algorithm was developed to improve the average UPT while ensure the fairness among users. The proposed PPF scheduling specially considered the ratio of the amount of transmitted data to the total burst data volume. The resources allocation matrix was considered to take full advantage of the limited resources. The impact of tuning parameters on PPF performance was one of our research focuses in this paper. Further more, we researched the PPF performance with long bursts and short bursts, respectively. It has been verified that the average UPT/UPT-cut was enhanced with fairness guarantee. In this paper, we only study the case that every user has one data burst. However the PPF scheduling is not suitable for the multi data bursts with arrivals that are a certain distribution, such as Poisson distribution. In the future, the multi data bursts will be deal with seriously.

## REFERENCES

[1] *Evolved Universal Terrestrial Radio Access (E-UTRA) Layer 2, Measurements (Release 14), Version 14.4.0*, document 36.300, Technical Specification Group Radio Access Network, Apr. 2017.
[2] C. So-In, R. Jain, and A.-K. Tamimi, "Scheduling in IEEE 802.16e mobile WiMAX networks: Key issues and a survey," *IEEE J. Sel. Areas Commun.*, vol. 27, no. 2, pp. 156–171, Feb. 2009.
[3] H. Kim and Y. Han, "A proportional fair scheduling for multicarrier transmission systems," *IEEE Commun. Lett.*, vol. 9, no. 3, pp. 210–212, Mar. 2005.
[4] 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 Commun. Surveys Tuts.*, vol. 15, no. 2, pp. 678–700, 2nd Quart., 2013.
[5] M. Ayhan, Y. Zhao, and H.-A. Choi, "Utilizing geometric mean in proportional fair scheduling: Enhanced throughput and fairness in LTE DL," in *Proc. IEEE Global Commun. Conf.*, Dec. 2017, pp. 1–6.
[6] N. Ferdosian, M. Othman, B. Ali, and K. Y. Lun, "Multi-targeted downlink scheduling for overload-states in LTE networks: Proportional fractional knapsack algorithm with Gaussian weights," *IEEE Access*, vol. 5, pp. 3016–3027, 2017.
[7] R. Kausar, Y. Chen, and K. K. Chai, "QoS aware packet scheduling with adaptive resource allocation for OFDMA based LTE-advanced networks," in *Proc. IET Int. Conf. Commun. Technol. Appl. (ICCTA)*, 2011, pp. 207–212.
[8] J. Navarro-Ortiz, P. Ameigeiras, J. M. Lopez-Soler, J. Lorca-Hernando, Q. Perez-Tarrero, and R. Garcia-Perez, "A QoE-aware scheduler for HTTP progressive video in OFDMA systems," *IEEE Commun. Lett.*, vol. 17, no. 4, pp. 677–680, Apr. 2013.
[9] Y.-C. Wang and T.-Y. Tsai, "A pricing-aware resource scheduling framework for LTE networks," *IEEE/ACM Trans. Netw.*, vol. 25, no. 3, pp. 1445–1458, Jun. 2017.
[10] M. M. Nasralla, M. Razaak, I. U. Rehman, and M. G. Martini, "Content-aware packet scheduling strategy for medical ultrasound videos over LTE wireless networks," *Comput. Netw.*, vol. 140, pp. 126–137, Jul. 2018.
[11] A. E. Mostafa and Y. Gadallah, "A statistical priority-based scheduling metric for m2m communications in LTE networks," *IEEE Access*, vol. 5, pp. 8106–8117, 2017.
[12] M. S. Mushtaq, S. Fowler, A. Mellouk, and B. Augustin, "QoE/QoS-aware LTE downlink scheduler for VoIP with power saving," *J. Netw. Comput. Appl.*, vol. 51, pp. 29–46, May 2015.
[13] S. Brueck, L. Zhao, J. Giese, and M. A. Amin, "Centralized scheduling for joint transmission coordinated multi-point in LTE-advanced," in *Proc. Int. ITG Workshop Smart Antennas (WSA)*, Feb. 2010, pp. 177–184.
[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, Oct. 2011.
[15] P. Barsocchi and G. Oligeri, "Quality of experience in multicast hybrid networks: Avoiding bandwidth wasting with a double-stage FEC scheme," *IET Commun.*, vol. 4, no. 13, pp. 1573–1579, Sep. 2010.
[16] P. Itu-T, *Methods for Subjective Determination of Transmission Quality*, document ITU-T Recommendation 800, 1996.
[17] P. Oliver-Balsalobre, M. Toril, S. Luna-Ramírez, and R. G. Garaluz, "Self-tuning of service priority parameters for optimizing quality of experience in LTE," *IEEE Trans. Veh. Technol.*, vol. 67, no. 4, pp. 3534–3544, Apr. 2018.
[18] L. He and G. Liu, "QoE-driven joint optimization of segment adaptation and resource allocation for DASH clients over LTE networks," *IEEE Trans. Veh. Technol.*, vol. 67, no. 11, pp. 11035–11048, Nov. 2018.
[19] *Evolved Universal Terrestrial Radio Access (E-UTRA); Further Advancements for E-UTRA Physical Layer Aspects (Release 9), Version 9.0.0*, document 36.814, Technical Specification Group Radio Access Network, Oct. 2010.

[20] I. A. Rai, E. W. Biersack, and G. Urvoy-Keller, "Size-based scheduling to improve the performance of short TCP flows," *IEEE Netw.*, vol. 19, no. 1, pp. 12–17, Jan./Feb. 2005.
[21] R. Jain, D.-M. Chiu, and W. R. Hawe, "A quantitative measure of fairness and discrimination for resource allocation in shared computer system," Eastern Res. Lab., Digit. Equip. Corp., Hudson, MA, USA, Res. Rep. TR-301, 1984, vol. 38.

**YAKE LI** received the B.S. degree in mathematics from the Henan University of Economics and Law, Zhengzhou, China, in 2015. She is currently pursuing the Ph.D. degree with the School of Aerospace Science and Technology, Xidian University, Xi'an, China. Her research interests include wireless communication, queue theory, and scheduling algorithms.

**XINPENG FANG** received the B.S. degree in automation from Central South University, Changsha, in 2009, and the Ph.D. degree in control theory and control engineering from Northwestern Polytechnical University, Xi'an, China, in 2016. He is currently a Lecturer with the School of Aerospace Science and Technology, Xidian University, Xi'an. His current research interests include target localization, motion control, and wireless communication.

**SIYU LIANG** received the B.S. degree from the School of Opto-Electronic Information Science and Technology, Yantai University, Yantai, China, in 2016. She is currently pursuing the M.S. degree with the School of Aerospace Science and Technology, Xidian University, Xi'an, China. Her research interests include wireless communication, machine learning, and scheduling algorithms.