SCIENCE CHINA
Information Sciences

• RESEARCH PAPER •

April 2019, Vol. 62 042306:1–042306:10
https://doi.org/10.1007/s11432-018-9746-9

Exponentially weighted proportional fair scheduling algorithm for the OFDMA system

Siyu LIANG, Weisheng CHEN*, Yake LI & Xinpeng FANG*

School of Aerospace Science and Technology, Xidian University, Xi'an 710071*, China*

Received 19 August 2018/Revised 1 November 2018/Accepted 6 December 2018/Published online 7 March 2019

Abstract This study aims to propose an exponentially weighted proportional fair (EWPF) scheduling algorithm for orthogonal frequency division multiple access (OFDMA) system for long-term evolution downlink transmission. The proposed algorithm improves the system performance in the user-perceived throughput (UPT) by adding exponential weights to different types of services. The UPT employed to measure the system capability is novel and customer oriented; it reflects the experiences of users in an efficient manner and determines whether the user's scheduling is reasonable. The EWPF algorithm is compared with two other schedulers, and our simulation results showed that the EWPF can increase the overall UPT and can maintain the fairness for prioritizing the transmission of some types of traffic.

Keywords exponentially weighted proportional fair, user perceived throughput, downlink scheduling algorithm, scheduled IP throughput

Citation Liang S Y, Chen W S, Li Y K, et al. Exponentially weighted proportional fair scheduling algorithm for the OFDMA system. Sci China Inf Sci, 2019, 62(4): 042306, https://doi.org/10.1007/s11432-018-9746-9

1 Introduction

With the emergence of high-speed multimedia data services, the rate requirements for the air interface in wireless communication systems have increased significantly. According to Shannon's information theory [1], if the wireless channel capacity is to be increased, either the system bandwidth or the system spectral efficiency must be increased. However, although the system bandwidth resources have been expanded, the spectrum resources are still limited and cannot be widened without any restrictions. Therefore, many recent technologies, such as the orthogonal frequency division multiple access (OFDMA) scheme are being used to improve the bandwidth utilization of the wireless network. For example, to reduce the inter symbol interference caused by the multipath transmission effect, OFDMA and single-carrier frequency division multiple access (SC-FDMA) were used for diverse multiusers in the long-term evolution (LTE) cellular system for the downlink and the uplink, respectively [2]. Some algorithms, such as the classical water-injection algorithm and power-allocation algorithm, which are proposed in [3], were used to assign power to users to minimize the system power consumption to the extent possible while satisfying the users' reception quality. Under these prerequisites, the user equipment (UE) is allocated to transfer data determined whether the system could give the users a better experience for the service provided by the operator and whether the spectrum utilization could be maximized. In other words, using different scheduling algorithms will greatly affect the performance of the LTE system. Therefore, many researchers are conducting studies that are based on allocating resources among users in an efficient manner to maximize the service quality under a fixed capacity.

* Corresponding author (email: wshchen@126.com, xinpengfang@163.com)

Three classical packet schedulers in communication networks are the round-robin (RR), the maximum carrier to interference (Max C/I), and the proportional fair (PF) algorithms. Max C/I and RR schedulers are based on throughput maximization and optimal fairness, respectively. Moreover, the PF algorithm can satisy the high-quality data service demands to the extent possible while considering the user experience of a poor channel quality [4]. General proportional fair approach, as presented in [5], was extended from the PF scheduler, which not only realizes the tradeoff between the system throughput and the fairness through parameter adjustment but also provides a significant multiuser diversity gain. Instead of using the arithmetic mean, the authors introduced two kinds of fair allocation high throughput (FAHT) algorithms using a geometric mean with a faster convergence for user throughput in [6]. By solving the optimization problem of maximizing a strictly concave smoothing utility function under constraints, the authors proposed a gradient algorithm with minimum/maximum rate (GMR) constraints that limit the user rates to a certain range [7]. Ref. [8] considered the statistical channel state information (CSI) and proposed the user scheduling algorithms that guaranteed the achievable sum rate and the fairness among the users. In addition, some algorithms focused on meeting the different quality of service (QoS) requirements [9, 10]. Ref. [11] proposed a QoS-aware packet scheduler that paid attention to the quality of experience (QoE), but for the network performance, this scheduler considered the users' subjective feelings more frequently. Ref. [12] also considered QoE to drive the radio resource allocation for different mobile applications. A calculation method for the LTE advanced spectrum requirements was put forward in [13]; a probability distribution was also given for the number of scheduled resource blocks (RBs) for a device at a time satisfying the QoS requirement. To a certain extent, this series of algorithms has improved the cell throughput, fairness, and system delay, among others.

However, the throughput indicators used by the algorithms above were mainly from the operator's point of view; it ignored the actual user experience of the current network. Herein, we will use a definition of the user-perceived throughput UPT[1)] to measure the system performance. The UPT can, to a certain extent, reveal the users' effective time in transmission, which can be used as a measure of the user experience. It should be declared that the UPT/UPT-cut is a cell-level index, i.e., to enhance the UPT, the sum of the transmission and waiting times of all users should be reduced for the users to have a better experience. In practical applications, many aspects involved in the transmission process, such as the fairness in resource allocation, the user's channel knowledge, and the delay, among others, should be taken into account. To solve this problem, we propose the EWPF scheduling algorithm obtained by theoretical derivation based on the definition of PF. Our proposed algorithm can change the priority to increase the transmission efficiency and the user experience by giving different weights to different traffic types. Additionally, we designed simulation experiments to evaluate the proposed algorithm aimed at raising the transmission rate of the real-time (RT) traffic. Finally, the UPT/UPT-cut and the fairness were employed to measure the effectiveness of the proposed algorithm. To the best of our knowledge, few researchers have studied the scheduling algorithm from the perspective of the UPT. The simulation results illustrate that the EWPF algorithm increased the UPT of all users and maintained the fairness between the users on the basis of enhancing the RT traffic at a high level. We analyzed the system performance theoretically, and we provided the specific transmission process is to verify the correctness of the analysis.

The rest of this paper is organized as follows. Section 2 presents a brief description of the downlink scheduling architecture and gives some definitions of throughput adopted in this paper. Section 3 gives the improved scheduling algorithm. In Section 4, we elaborate our LTE simulation platform and demonstrate some comparison results. Finally, Section 5 concludes the paper.

2 Preliminary

2.1 Downlink scheduling architecture

In the LTE system, the scheduler is deployed in the base station (BS). Media access control is responsible

1) Herein, the performance metrics UPT and UPT-cut are publicly released by Huawei Technologies Co. Ltd., and all the average UPT/UPT-cuts are at the cell level, unless mentioned otherwise.

for the management and scheduling of wireless resources. Each UE calculates the channel quality indicator (CQI) at the current moment and sends it to the BS through uplink [14]. The scheduler uses the reported CQI information to make an allocation decision subsequently. Time/domain resources are allocated to users according to the users' channel quality and buffer queue, among other information [15]. More precisely, at each transmission time interval (TTI), RBs, which are the smallest radio resource units are assigned to users based on the scheduler. An RB contains one TTI, i.e., two time slots having a duration of 0.5 ms each, in the time domain and 12 subcarriers in which the carrier spacing is 15 kHz in the frequency domain, i.e., bandwidth is 180 kHz. When a large number of users have independent Rayleigh fading in a network waiting for transmission, there is always one user with a better channel condition. If the user mentioned above is allowed to transmit, channel resources will be efficiently utilized. When the BS allocates resources, the scheduler should not only improve the throughput but should also consider the fairness and QoE requirements of users. Therefore, the scheduler should meet the demands of the users and reduce the invalid transmission under bad channel conditions. If a scheduler is determined, all the users will have transmission priority in each RB, and the RBs will arrange the users who have the highest transmission priority.

2.2 UPT/UPT-cut in downlink

The proposed scheduling algorithms and the performance indexes are mainly established from the operator's perspective. These algorithms seldom consider the users' actual opinion about the current wireless network, i.e., a good network performance indicator is not necessarily identical to a good user experience. Therefore, in this subsection, we will give two performance indexes to measure the system. These metrics are determined on the basis of the scheduled Internet protocol (IP) throughput, which is defined by the third-generation partnership project (3GPP) protocol [16].

Definition 1 (Scheduled IP throughput in downlink)**.** For any user $i \in U$, the scheduled IP throughput V_i (user-level) within the time period $\overline{T_i} + T$ is defined as follows:

$$V_i = \begin{cases} \dfrac{\overline{b}_i + b_i}{\overline{T}_i + T}, & i \in U_S, \\ \dfrac{\overline{b}_i}{\overline{T}_i + T}, & \text{otherwise}, \end{cases} \tag{1}$$

where U is the user set, and U_S is the set of users selected by an arbitrary scheduler S; $\overline{T_i}$ is the effective transmission time that the user i has spent for data transmission; T is the interval between the two scheduling times; $\overline{b}_i$ is the amount of data transmitted; $b_i = \sum_{k \in C_i} b_{i,k}$ is the amount of data transmitted by the user in the time interval T, and C_i is the set of carriers allocated to the user i. The BS can obtain $b_{i,k}$ from the modulation and coding scheme based on the CQI, which is fed back by users.

Definition 1 defines the scheduled IP throughput that reflects the data transmission of each user. Based on Definition 1, we propose Definition 2 to display the overall transmission of all users in a cell over a certain period. This definition can measure the user's experience better because it considers the users' waiting time during the transmission, which is a main factor for the users to judge the network status.

Definition 2 (User perceived throughput)**.** For any user $i \in U$, the UPT V (cell-level) within the time period T_{total} is defined as follows:

$$V = \frac{\sum_{T=1}^{T_{\text{total}}} \sum_{i=1}^{U} b_i^T}{\sum_{T=1}^{T_{\text{total}}} \sum_{i=1}^{U} t_i^T}, \tag{2}$$

where b_i^T is the data transferred by the user i, and t_i^T is the effective transmission time of the user i during the time interval T.

Remark 1. As shown in Figure 1, this expression can be classified into two criteria, UPT and UPT-cut. The UPT-cut criterion ignores the users' waiting time and the transmission of the last TTI, that is, the time from $T2$ to $T3 - 1$. The UPT criterion is used to calculate the throughput from the user from the initiation of the request ($T1$) to the end of the transmission ($T3$). That is, if there is a user with

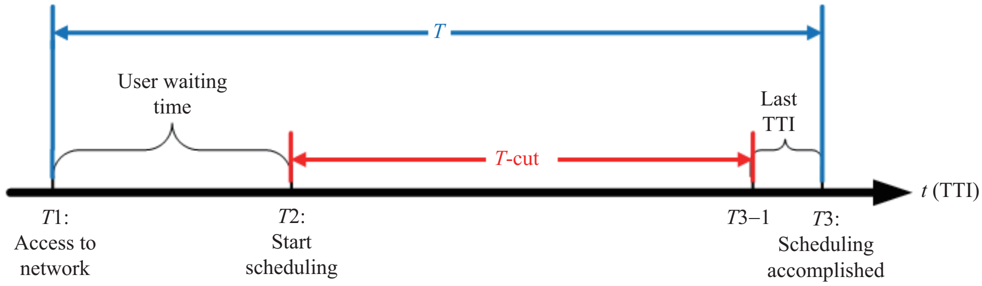

Figure 1 (Color online) The difference between the UPT and UPT-cut.

a transmission request, the system will start timing until the end of the transmission. We will mainly use the UPT-cut to measure the performance of the cellular network system in Section 4. Definition 2 is the ratio of the transferred data amount to the effective transmission time. The intuitive experience of the network users mainly depends on the response time of the network after the users apply for network access. We can see from the structure of (2) that the UPT can reflect both the total amount of data transmission and the cumulative waiting time and the transmission time of all users in the system. After generating all the user information on the platform, the scheduler starts to schedule users in the cell. We used different scheduling algorithms in the following two cases and used the UPT to measure the system. If all these users could complete the transmission within a certain period, the total amount of data transferred would be a fixed value regardless of the scheduling algorithm used. In this case, when calculating the UPT, $\sum_{T=1}^{T_{\rm total}}\sum_{i=1}^{U} b_i^T$ of each scheduling algorithm is the same. Then, the scheduler with less accumulated waiting time for all users will inevitably receive a higher UPT. If the total amount of data transmission of the user differs under these scheduling algorithms within a certain period, we can see that as the UPT increases, the system transmission amount increases and the user cumulative transmission time decreases. In this case, if the system measures a high UPT, most users will feel that the service quality of the operator is better under the current scheduling policy. Therefore, to a certain extent, the UPT and the UPT-cut can both reflect the users' QoE and the system throughput better. Moreover, the UPT is a cell-level indicator only when the system does reasonable scheduling on every TTI and each PRB. Hence, to obtain a high UPT, as long as there is a transmission requirement, the system must minimize all the users' waiting time before starting the transmission and during the transmission regardless of whether it is a cell center user or a cell edge user.

2.3 System metrics

The purpose of radio resource management for LTE systems is to ensure the efficient use of limited radio resources. However, resource scheduling is one of the key problems that have not properly solved, because some of the traditional scheduling algorithms cannot perform well in both of system throughput and user fairness. Hence, we need to focus on the fairness while increasing the system throughput, especially for a high load.

The RB usage, which is defined in [16] to reflect the system load, is as follows.

Definition 3 (RB usage). The RB usage $M(T)$ over the time interval T is defined as

$$M(T)=\frac{\sum_{t=1}^{T}m_t}{\sum_{t=1}^{T}p_t}\times 100\%, \tag{3}$$

where m_t and p_t represent the number of RBs used by all users and the RBs available at any time t, respectively.

For fairness under different system loads, we employed the Raj Jain index to measure whether the users in the same cell could share the network resources fairly when congestion occurred in the network [17].

Definition 4 (Fairness index). The fairness among N users is defined according to Raj Jain index as follows:

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

where v_i is the data-rate achieved by the user i within a given time interval. The fairness value of 1 corresponds to the optimal fairness, that is, all the users in a cell are allocated equally.

3 EWPF scheduling algorithm

The system considered in this paper comprises a cellular network with N users associated with an eNodeB, which is located at the center of the cell with three sectors. To meet the needs of different user services, we define the weighted proportional fairness, which is similar to the proportional fairness [18], as follows.

Definition 5 (Exponentially weighted proportional fair)**.** The scheduling P is exponentially weighted proportional fair (EWPF) if and only if for any feasible scheduling S, it satisfies

$$\sum_{i\in U}\omega_i\frac{R_i^{(S)} - R_i^{(P)}}{R_i^{(P)}} \leqslant 0, \tag{5}$$

where $R_i^{(S)}$ is the average rate of the user i under the scheduler S, and ω_i is the exponential weight of the user i.

It is easy to derive that an EWPF scheduler P can maximize the sum of the weighted logarithmic average user rates, that is, the solution of (5) can be obtained by solving the following optimization problem:

$$P = \arg\max_{S}\sum_{i\in U}\omega_i\log R_i^{(S)}. \tag{6}$$

Theorem 1 (EWPF scheduling for multi carrier transmission systems)**.** In a multi-carrier transmission system, the scheduler P is EWPF if and only if it satisfies

$$P = \arg\max_{S}\prod_{i\in U_S}\left(1 + \frac{\sum_{k\in C_i} r_{i,k}}{(T_c - 1)\bar{R}_i}\right)^{\omega_i}, \tag{7}$$

where $r_{i,k}$ represents the highest rate that the user i could support on the subcarrier $k \in C_i$ at the current TTI; $\bar{R}_i$ is the average throughput of the user i recorded by the BS within the time window T_c.

Proof. This part is similar to the proof process of the traditional proportional fairness algorithm in [19]. As described above, to achieve weighted proportional fairness, Eq. (6) must be satisfied. If this is equivalent to the expression

$$\sum_{i\in U}\omega_i\log R_i^{(P)} \geqslant \sum_{i\in U}\omega_i\log R_i^{(S)}, \tag{8}$$

then it can be transformed into

$$\prod_{i\in U} R_i^{(P)\omega_i} \geqslant \prod_{i\in U} R_i^{(S)\omega_i}. \tag{9}$$

We assume that the users scheduled by the schedulers P and S belong to U_P and U_S, respectively. As shown in Figure 2, the different schedulers may lead to different selections of users by the BS but all the users would have a historical average rate R_i, which is not affected by the changing of the scheduler.

$$\prod_{i\in U_P\cup U_S} R_i^{(P)\omega_i} \geqslant \prod_{i\in U_P\cup U_S} R_i^{(S)\omega_i}. \tag{10}$$

Note that $U_S\bigcup U_P$ can be rewritten as $U_P\bigcup(U_S - U_P)$ or $U_S\bigcup(U_P - U_S)$. Then, we have

$$\prod_{i\in U_P} R_i^{(P)\omega_i}\prod_{i\in U_S - U_P} R_i^{(P)\omega_i} \geqslant \prod_{i\in U_S} R_i^{(S)\omega_i}\prod_{i\in U_P - U_S} R_i^{(S)\omega_i}. \tag{11}$$

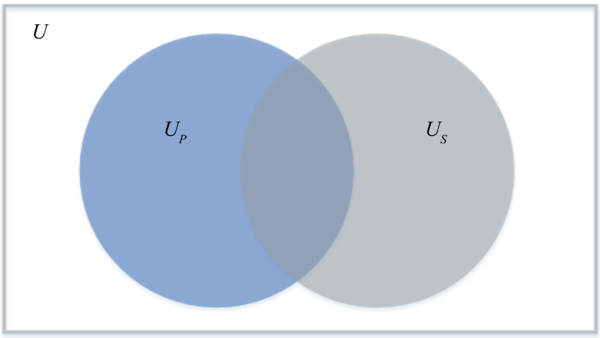

Figure 2 (Color online) Users attached to different schedulers.

The average rate of each user is updated according to the following equation:

$$R_i^{(P)} = \begin{cases} \dfrac{(T_c-1)\bar{R}_i + \sum_{k\in C_i} r_{i,k}}{T_c}, & \text{if } i \in U_P, \\ \dfrac{(T_c-1)\bar{R}_i}{T_c}, & \text{otherwise.} \end{cases} \tag{12}$$

By substituting R_i in (12) into (11), we obtain the following inequality:

$$\begin{aligned} &\prod_{i\in U_P}\left(\frac{(T_c-1)\bar{R}_i + \sum_{k\in C_i} r_{i,k}}{T_c}\right)^{\omega_i} \prod_{i\in U_S-U_P}\left(\frac{(T_c-1)\bar{R}_i}{T_c}\right)^{\omega_i} \\ &\geqslant \prod_{i\in U_S}\left(\frac{(T_c-1)\bar{R}_i + \sum_{k\in C_i} r_{i,k}}{T_c}\right)^{\omega_i} \prod_{i\in U_P-U_S}\left(\frac{(T_c-1)\bar{R}_i}{T_c}\right)^{\omega_i}. \end{aligned} \tag{13}$$

Multiplying both sides of (13) by $T_c \prod_{i\in U_P\cup U_S}(T_c-1)\bar{R}_i$, we have

$$\begin{aligned} &\prod_{i\in U_P}\left((T_c-1)\bar{R}_i + \sum_{k\in C_i} r_{i,k}\right)^{\omega_i} \prod_{i\in U_S}\left((T_c-1)\bar{R}_i\right)^{\omega_i} \\ &\geqslant \prod_{i\in U_S}\left((T_c-1)\bar{R}_i + \sum_{k\in C_i} r_{i,k}\right)^{\omega_i} \prod_{i\in U_P}\left((T_c-1)\bar{R}_i\right)^{\omega_i}. \end{aligned} \tag{14}$$

Next, we divide both sides of the inequality (14) by $\prod_{i\in U_P}(T_c-1)\bar{R}_i \prod_{i\in U_S}(T_c-1)\bar{R}_i$. Then, we obtain

$$\prod_{i\in U_P}\left(1+\frac{\sum_{k\in C_i} r_{i,k}}{(T_c-1)\bar{R}_i}\right)^{\omega_i} \geqslant \prod_{i\in U_S}\left(1+\frac{\sum_{k\in C_i} r_{i,k}}{(T_c-1)\bar{R}_i}\right)^{\omega_i}. \tag{15}$$

Consequently, in the multi-carrier system, the EWPF scheduler can be described as

$$P = \arg\max_{S} \prod_{i\in U_S}\left(1+\frac{\sum_{k\in C_i} r_{i,k}}{(T_c-1)\bar{R}_i}\right)^{\omega_i}, \tag{16}$$

which is the same as that in (7). It is easy to obtain the following scheduler for the single-carrier system:

$$P = \arg\max_{S}\left(1+\frac{r_{i,k}}{(T_c-1)\bar{R}_i}\right)^{\omega_i}. \tag{17}$$

This algorithm is based on the PF fairness, and we can adjust user priorities for different kinds of users or traffic by adding the exponential weights. EWPF can improve the system performance and ensure fairness while increasing the priority of some kinds of traffic that require high transmission efficiency.

Hence, this algorithm not only ensures the fairness of all users but also considers the transmission requirements of users, which the PF algorithm cannot do. The values of the exponential weights should not differ greatly; otherwise, it would be easy for the BS to schedule one kind of service and ignore the other users so they would not be scheduled even if their channel conditions were better.

The BS would schedule users based on the EWPF algorithm stated above. For example, each RB will provide service to the users who have the highest priority. First, each user obtains an exponential weight ω_i according to the type of traffic. Then, the users will be sorted on every RB by the priority values calculated by (7). For the exponential weights, we can set the weight of users with large packets as 1 and the weight of users with small packets as 3. Thus, the users with small packets will receive prior scheduling opportunities. In addition, ω_i can be taken arbitrarily as long as the priority value of the service that needs to be preferentially scheduled can be numerically greater than the others. If there is a combination of two kinds of traffic, the weights can be set as $\omega_i = u_i v_i$ accordingly.

4 Simulation results

In this part, to verify the validity of the proposed EWPF algorithm, we took mainly the RT traffic and the non-real-time (NRT) traffic as examples to prove the effectiveness of the proposed method. We assigned the users with the RT traffic a high priority by setting the exponential weights to different services to reduce the transmission time and to give the users a better network experience. For this purpose, we used MATLAB to build an LTE simulation platform in this section. We conducted performance comparisons from three aspects: the UPT-cut, fairness, and RB usage.

In our platform, the BS contained three directional antennas; we also took the co-channel interference into account. Besides, the simulations also considered the path loss, shadow fading, and fast fading. For convenience, each user initiated a transfer request only once in our scenario. The simulation duration was chosen as 100 TTIs. The time when the user sent a transmission request followed the Poisson distribution, and the time interval between two users who arrived at adjacent times was subject to an exponential distribution. The user locations were uniformly distributed throughout each sector. To make the simulation closer to the practical application, we considered both large and small data packets in the 3 : 7 ratio; the sizes of the data packets were generated by a normal distribution. Moreover, the user traffic was divided into two categories, namely RT and NRT traffics, and each type of traffic accounted for 50% of the total traffic. The other standard parameters and general algorithms used for the simulation are summarized in Table 1 [20–22]. Some parameters, such as the number of RB and number of users per sector, among others, are the corresponding configurations based on the RB usage. The role of the simulation platform was to generate users randomly; the settings of the parameters and the variable distribution described above are aimed to bring the simulation scenario as close as possible to the actual circumstances. After completing the parameter configuration, all the users will be scheduled in the same network environment, and the results will depend entirely on the scheduling algorithm used in the experiment.

To weaken the influence of the system randomness, such as the time when the users arrived and the variation of the mobile channel caused by different user distributions, we changed the scheduling environment 50 times and obtained an average of the results. According to some experiences, the exponential weights of the RT and NRT traffics were taken as 3.2 and 1.5, respectively, which is a reasonable setup in the following simulation results. Additionally, we also tested the FAHT and PF algorithms in the proposed platform for comparing with the EWPF algorithm, which aimed to assess the performance of the presented algorithm. The number of users in each sector varied from 3 to 30, and the average RB usage calculated by (3) were 11.8%, 21.1%, 32.3%, 40.6%, 48.7%, 62.7%, 66.8%, 72.1%, 76.3%, and 84.4%.

During the simulation, the UPT-cut of the three algorithms are shown in Figure 3. We can clearly see that the EWPF algorithm has the highest average UPT-cut in the proposed system simulation platform. Figure 4 is the average fairness calculated during the simulation time, and the fairness of our algorithm does not reduce much as compared with the other two algorithms under a high system load. Even under

Table 1 Simulation parameters for LTE downlink system

Item	Parameter settings
Number of RB	5
Maximum power of directional antenna	43 dBm (20 W)
Total bandwidth	180 kHz
Subcarrier bandwidth	15 kHz
Sector radius	200 m
Height of BS	35 m
BS antenna gain	15 dBi
Mobile station antenna gain	0
Path-loss model	34.5+35×log10(d) (dB)
Shadow-fading standard deciation	8 dB
Fast fading channel model	Rayleigh distribution
Thermal noise density	−174 dBm/Hz
System and link level mapping interface	EESM
Target bit-error-rate (BER)	10%
Frequency reuse factor	1
Power allocation policy	Equally distributed
Number of users per sector	Change from 3 to 30
User distribution	Uniformly distributed per sector
Average window size	5 TTI

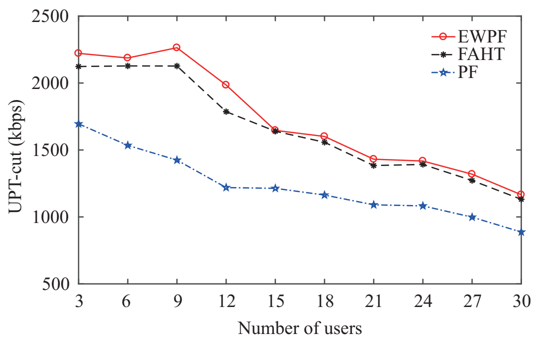

Figure 3 (Color online) The UPT-cut of the three algorithms.

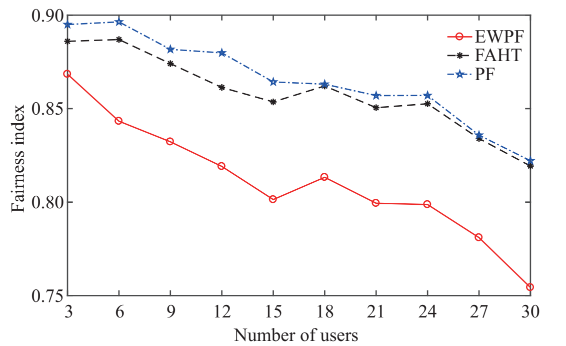

Figure 4 (Color online) The fairness index of the three algorithms.

a high load of approximately 70%, the fairness of the EWPF algorithm reduced only by approximately 5%. Furthermore, as illustrated in Figure 5(a), we separately counted the UPT of the users who sent the RT service transmission request. We found that the EWPF algorithm achieved better performance for the RT traffic. This was because under the condition of limited bandwidth resources, some NRT traffics would inevitably be delayed while some RT traffic would be prioritized. However, as shown in Figure 5(b), the NRT traffic does not delay much, and the decline in the NRT traffic under low and medium loads is less obvious than the decline under high loads. Combined with the fairness shown in Figure 4, we can reasonably conclude that the deduction part was based on proportional fairness, and our algorithm did not sacrifice much fairness in exchange for system throughput. To clarify the validity of the EWPF algorithm and to make the scheduler process more obvious, we selected several representative data transmission processes to indicate the difference between the three algorithms in user scheduling. Each algorithm has a different emphasis; therefore, the process of user scheduling and data transmission will be distinct. Comparisons of the data transmission performances of the three algorithms are shown in Figure 6. We plotted the time and the amount of the residual data on the horizontal axis and the vertical axis, respectively. Each user is represented by a line of a different color. The real and dotted lines represent the RT and NRT traffic, respectively. Figure 6 shows how users transmit data under three

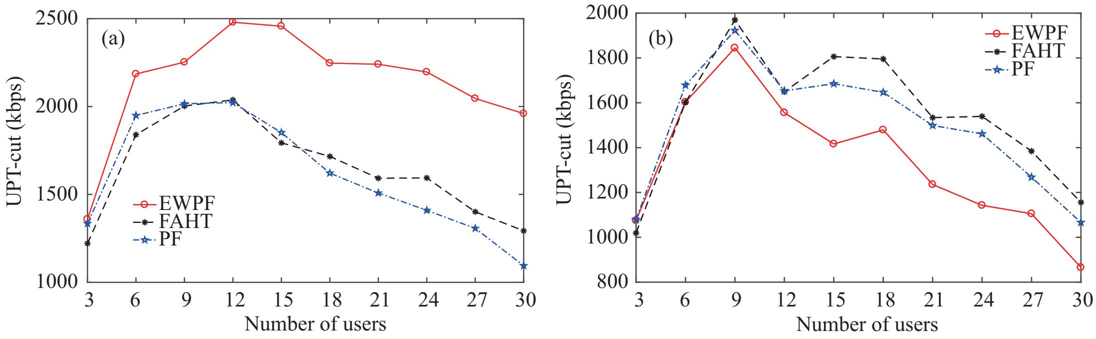

Figure 5 (Color online) Comparisons between PF, FAHT and EWPF algorithms. (a) UPT-cut of RT Traffic; (b) UPT-cut of NRT Traffic.

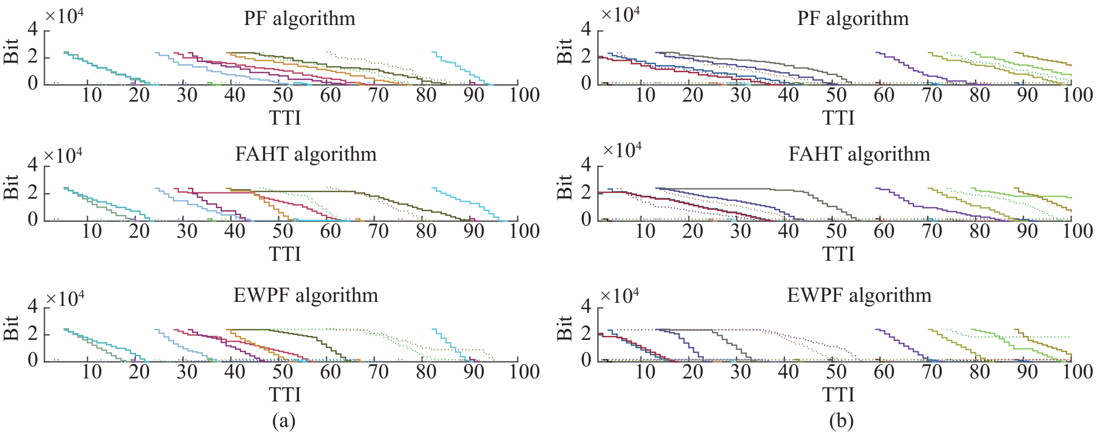

Figure 6 (Color online) The amount of data left. (a) Number of users = 27; (b) number of users = 30. Each user is represented by a line of a different color. The real and dotted lines represent the RT and NRT traffic, respectively.

algorithms and the difference in the scheduling order. It is clear that the RT traffic has priority over the NRT traffic, and the effective transmission time of the RT traffic is shorter than that of the other two algorithms. Moreover, NRT users are not ignored by schedulers because of the reasonable numerical settings. When combined with fairness, we know that the derivation based on the PF algorithm and the exponential weights of the different traffic do play a role in the scheduling process. From the above figures, we concluded that the EWPF algorithm performs well for the RT traffic, and it also receives a substantial increase in UPT-cut for the overall view.

5 Conclusion

Herein, we proposed an improved EWPF scheduling algorithm for the OFDMA downlink system. By employing the exponential weight, the EWPF algorithm made differences in the user priority among the various types of traffic. The EWPF algorithm is derived from the definition of proportional fairness, and the fairness among users is guaranteed accordingly. Meanwhile, we considered the QoE requirements of the users with different traffic types. Moreover, by considering users rather than the mobile communication operators, we used the UPT/UPT-cut to measure the network capability. Obviously, the simulation results demonstrated that the proposed algorithm maintained the system performance at a high level in terms of the UPT and the fairness.

Acknowledgements This work was partially supported by National Natural Science Foundation of China (Grant Nos. 61703326, 61673308, 61673014), Natural Science Foundation of Shaanxi Province (Grant No. 2017JQ5037), and Fundamental Research Funds for the Central Universities (Grant No. 20101186377).

References

1 Dobrushin R L. General formulation of Shannon's main theorem in information theory. Amer Math Soc Trans, 1963, 33: 323–438
2 Hara S, Prasad R. Overview of multicarrier CDMA. IEEE Commun Mag, 1997, 35: 126–133
3 Wei H X, Li Y Z, Xiao L M, et al. Queue-aware energy-efficient scheduling and power allocation with feedback reduction in small-cell networks. Sci China Inf Sci, 2018, 61: 048301
4 Jalali A, Padovani R, Pankaj R. Data throughput of CDMA-HDR a high efficiency-high data rate personal communication wireless system. In: Proceedings of Vehicular Technology Conference, Tokyo, 2000. 1854–1858
5 Wengerter C, Ohlhorst J, Elbwart A G E. Fairness and throughput analysis for generalized proportional fair frequency scheduling in OFDMA. In: Proceedings of Vehicular Technology Conference, Stockholm, 2005. 1903–1907
6 Ayhan M, Zhao Y, Choi H A. Utilizing geometric mean in proportional fair scheduling: enhanced throughput and fairness in LTE DL. In: Proceedings of Global Communications Conference, Washington, 2016
7 Andrews M, Qian L, Stolyar A. Optimal utility based multi-user throughput allocation subject to throughput constraints. In: Proceedings of the 24th Annual Joint Conference of the IEEE Computer and Communications Societies, Miami, 2005. 2415–2424
8 Li X, Sun T T, Qin N N, et al. User scheduling for downlink FD-MIMO systems under Rician fading exploiting statistical CSI. Sci China Inf Sci, 2018, 61: 082302
9 Lorenz D H, Orda A. Optimal partition of QoS requirements on unicast paths and multicast trees. IEEE/ACM Trans Netw, 2002, 10: 102–114
10 Yu W J, Musavian L, Ni Q. Statistical delay QoS driven energy efficiency and effective capacity tradeoff for uplink multi-user multi-carrier systems. IEEE Trans Commun, 2017, 65: 3494–3508
11 Piro G, Grieco L A, Boggia G, et al. Two-level downlink scheduling for real-time multimedia services in LTE networks. IEEE Trans Multimedia, 2011, 13: 1052–1065
12 Fei Z S, Xing C W, Li N. QoE-driven resource allocation for mobile IP services in wireless network. Sci China Inf Sci, 2015, 58: 012301
13 Kim Y, Park S. Analytical calculation of spectrum requirements for LTE-A using the probability distribution on the scheduled resource blocks. IEEE Commun Lett, 2018, 22: 602–605
14 Capozzi F, Piro G, Grieco L A, et al. Downlink packet scheduling in LTE cellular networks: key design issues and a survey. IEEE Commun Surv Tutor, 2013, 15: 678–700
15 Ekstrom H. QoS control in the 3GPP evolved packet system. IEEE Commun Mag, 2009, 47: 76–83
16 3GPP. Evolved universal terrestrial radio access (E-UTRA); user equipment (UE) radio transmission and reception. 3GPP TS V10, 2011. http://arib.or.jp/english/html/overview/doc/STD-T104v1_00/2_T104/ARIB-STD-T104/Rel10/36/A36101-a30.pdf
17 Jain R, Chiu D M, Hawe W. A Quantitative Measure of Fairness and Discrimination for Resource Allocation in Shared Systems. Technical Report DEC-TR-301, 1984
18 Kelly F P, Maulloo A K, Tan D K H. Rate control for communication networks: shadow prices, proportional fairness and stability. J Oper Res Soc, 1998, 49: 237–252
19 Kim H, Han Y. A proportional fair scheduling for multicarrier transmission systems. IEEE Commun Lett, 2005, 9: 210–212
20 Zheng Y R, Xiao C S. Improved models for the generation of multiple uncorrelated Rayleigh fading waveforms. IEEE Commun Lett, 2002, 6: 256–258
21 Li Y, Yu F, Zheng S L, et al. LTE system level simulation with MATLAB. In: Proceedings of International Conference on Internet Technology and Applications, Wuhan, 2011
22 Prabhu G S, Shankar P M. Simulation of flat fading using MATLAB for classroom instruction. IEEE Trans Educ, 2002, 45: 19–25