A Downlink Scheduling Algorithm Based on Network Slicing for 5G

Shanwei Wang(✉), Bing Xi, Zhizhong Zhang, and Bingguang Deng

School of Communication and Information Engineering, Chongqing University
of Posts and Telecommunications, Chongqing, China
S180101165@stu.cqupt.edu.cn

Abstract. Current cellular mobile network should satisfy the service requirements of the User Equipment (UE) applications through Radio Resource Management (RRM) mechanisms as much as possible. In order to improve the resource utilization rate and Quality of Experience (QoE) for downlink Real-Time (RT) services in 5G system. In this paper, based on the Modified Largest Weighted Delay First (M-LWDF) algorithm, a slicing-oriented resource scheduling algorithm-S-MLWDF is proposed with using 5G network slicing technology. S-MLWDF takes RB groups as the basic units of RA (resource allocation) and takes slices as the allocation object. During the process of in-slice scheduling, on account of the Channel Quality Indication (CQI) obtained from Base Station (BS) feedback and the allocation of RBs over time, the generated weighting factor can guarantee the edge users to get equal scheduling opportunities. Meanwhile, the modified queue delay and HARQ retransmission packets delay can solve the problem of surge in Packet Loss Rate (PLR) near the delay threshold. The simulated results show that the performance of the proposed algorithm is better than the traditional scheduling algorithms. Especially compared with M-LWDF, the fairness and PLR of S-MLWDF are optimized by about 10% and 16.3%, which can better meet the needs of users.

Keywords: 5G · Network-slicing · S-MLWDF algorithm · Resource scheduling · Delay · Resource blocks allocation

1 Introduction

With the global commercial deployment of 5G, the Ministry of Industry and Information Technology (MIIT) issued the occupation permit of radio frequency to China Telecom, China Unicom and China Broadcast Network respectively in February 2020. This means that 5G network has entered the stage of large-scale deployment. In the process of mobile communication system evolving to the fifth generation, it is necessary to meet the great challenges of differentiated service requirements in multiple scenarios. The different requirements such as safety, mobility, transmission delay, instantaneous speed and so on in Enhanced Mobile Broadband (eMBB), Massive Machine Type Communication (mMTC) and Ultra-reliable and Low Latency Communication (uRLLC) should be met as possible. As it happens, network slicing technology can realize the customization capability through effective management of current wireless resources [1, 2].

H. Gao et al. (Eds.): ChinaCom 2020, LNICST 352, pp. 212–225, 2021.
https://doi.org/10.1007/978-3-030-67720-6_15

5G RRM focus on latency-sensitive applications and massive data. As a mechanism to provide Quality of Service (QoS) requirements and improve system throughput in multi-user networks, packet scheduling will meet strict requirements of delay and PLR. In order to meet the QoS demands for RT communication, various packet scheduling algorithms have been used to allocate limited frequency and time sources for all data transfer devices including mobile and wireless networks [3–5]. On the premise of 5G high-speed rate, high-capacity, low latency, high reliability demand, the limited external environment and public resources, the importance of scheduling algorithms to allocate resources reasonably for users is self-evident. Proportional Fair (PF), the classical scheduling algorithm, achieves the trade-off between throughput and user fairness in Non-Real-Time (NRT) services [6]. In Modified Largest Weight Delay First (M-LWDF) algorithm, the delay of packets is the main parameter [7]. However, when the number of users increased, the PLR of queues near delay threshold in buffer increased greatly. The two-stage downlink scheduling algorithm ensured fairness without reducing system throughput [8]. The authors of [9] proposed a downlink channel queue-aware scheduling algorithm based on service curve and overflow probability of buffer, which maximized throughput, provided lower latency and buffer overflow rate. The authors of [10] analyzed the buffer state of user queues which took delay as the weight, and proposed a scheduling scheme based on delay and QoS-aware to minimize the delay of RT traffic. An enhanced delay sensitive algorithm which is used to increase the energy efficiency, network lifetime and throughput was proposed [11]. However, when number of nodes increased, delay also increased. In addition, a new channel-aware scheduling algorithm for improving the cell edge throughput and fairness has been proposed [12], but the problem of PLR growth was neglected. At the same time, scholars at home and abroad have also done research on slice scheduling. A two-layer MAC scheduling framework was proposed in [13] to handle uplink and downlink transmission of network slices with different characteristics in RAN, which can meet the strict latency and reliability requirements in uRLLC. In [14], the authors proposed a strategy which prioritized slices for different service providers. Slices of RT service are scheduled first, and slices of Non-Real-Time (NRT) service are scheduled later. Their studies mainly focused on one aspect of performance Their studies mainly focused on one aspect of performance optimization, however, they did not comprehensively consider the impact of various parameters in 5G different services on RA.

Hence, in this paper, 5G network slicing technology is adopted to divide the wireless resource scheduling process into slice-level scheduling and in-slice resource scheduling. During in-slice scheduling, on the basis of M-LWDF algorithm, we improve the scheduling priority of queues which will approach the delay threshold, consider the current channel quality and RBs allocated to users in the previous period. Simulation results show that the proposed algorithm can keep satisfactory balance between fairness and throughput at the same time of maximizing resource utilization and reduce the PLR effectively.

The rest of the paper is organized as follows. In Sect. 2 the system model of packet scheduling for 5G system is provided, and the resource grid in the time-frequency domain is introduced. The proposed algorithm is discussed in Sect. 3. In Sect. 4 simulation results are shown to compare the proposed algorithm with the existing algorithms in terms of resource utilization, user fairness, system throughput and PLR. Finally, in Sect. 5 we draw our conclusion.

2 System Model

As the smallest unit of resources that can be allocated to a user [15], RBs can form a time-frequency resource grid to represent downlink physical resources. In 5G system, Resource Element (RE) is the basic unit of resource mapping in physical layer. A RE consists of an OFDM symbol in the time domain and a subcarrier in the frequency domain, an RB is composed of OFDM symbols and 12 subcarriers. The length of the RB in time is called the Transmission Time Interval (TTI), all frequency blocks at a given TTI are called a subframe [16]. When the resource blocks take different μ values, it can meet the data transmission rate and throughput rate required by the system for different scenes. The time-frequency wireless resource grid is shown in Fig. 1.

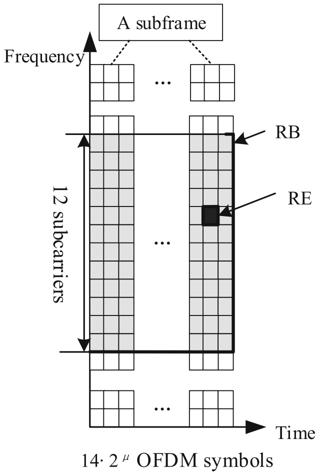

Fig. 1. Wireless resource grid in the time-frequency domain

The MAC scheduler is the controlling entity for multi-users radio RA, which assigns shared physical resources to different users in the cell by prioritizing them. The 5G PHY layer offers a large set of new options for the MAC scheduler, which enable significant improvements for efficiently multiplexing users with highly diverse service requirements [17]. When a service needs to send downlink data, for different service and users, the queue pass through the classifier from the upper layer to different buffer on the base station side. By now the MAC scheduler implements resource scheduling process by accepting CQI reported by UEs and selecting scheduling algorithms (Fig. 2).

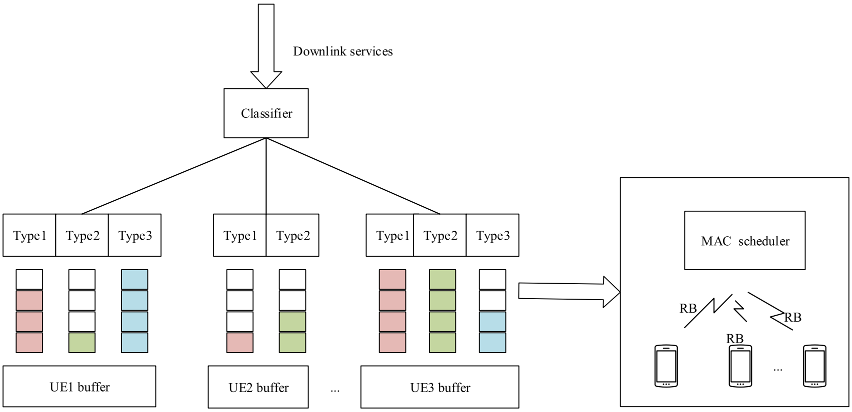

Fig. 2. Downlink scheduling model

During the period of downlink resource scheduling, BS timely adjusts and updates the channel state information of users according to the feedback CQI. Then the Adaptive Modulation and Coding (AMC) module selects the right Modulation and Coding Scheme (MCS) within a given Block Error Rate (BLER), according to the CQI reported, in order to maximize the throughput [18].

3 S-MLWDF Algorithm

Users in different slices in 5G network slicing architecture have different requirements for various services, which makes each user has different priorities in resource scheduling. Therefore, it is necessary to allocate resources adequately according to the differentiated demands of users in slices to ensure the QoE of users in wireless resource scheduling. Based on this, the proposed algorithm includes slice-level scheduling process and in-slice user resource scheduling process.

3.1 Slice-Level Resource Block Group Scheduling

Within a TTI, resources in current time-slot are divided into several RB groups according to the number of candidates, and slices are selected according to the priority order of each slice on each RB group for scheduling until the whole of RB groups are allocated.

We assume that the number of scheduled candidates in current TTI is $I(t)$, the total number of RBs available is R, then the number of an RB groups at time t is $c = \max\left(1, \frac{R}{I(t)}\right)$. Considering that the three scenarios in 5G have different service requirements, we propose the priority metric of slice a in the b-th RB group.

$$m_k = \frac{\sum_{n=1}^{c} \sum_{k=1}^{I_a(t)} \left(\alpha \cdot r_{m,k}(t) + \beta \cdot W_k(t)\right)}{c} \quad (1)$$

Where m_k is the metric of k-th user, $I_a(t)$ is the number of candidates in slice a, $r_{m,k}(t)$ is the instantaneous transmission rate of the k-th user on the m-th RB. $W_k(t)$ is the delay of the Head of Line (HOL) packet, α, β are transmission rate and weight factor of queue delay respectively, and $\alpha + \beta = 1$. The in-slice scheduling process is as follows:

Algorithm 1 Slice-level Resource Block Group Scheduling

```
1: Input: $T, R, I(t), W_k(t), r_{m,k}(t), \tau_k$;
2: Output: $m_k$
3: while $t < T$ do
4:     for $k = 20, 30, 40, \ldots, 100$ do
5:         $c = R / I(t)$;
6:         Calculate the priority of slices $m_k$ on each RB group;
7:         while $c$
8:             Assign RB group to the $a$-th slice;
9:         end while
10:    end for
11: end while
```

3.2 In-Slice User Scheduling

After the slice-level process, the slices with the highest priority are allocated a set of RBs. At this time, each slice can be regarded as a cell, and the RB group obtained through the slice-level process is used to conduct resource scheduling for users in the corresponding slice. In scheduling process of traditional cells, the M-LWDF is a channel-aware extension of PF and provides bounded packet delivering delay [19]. This algorithm is dedicated for RT services. It can ensure balance between spectrum efficiency, system fairness and QoS. M-LWDF metric can be easily expressed as in Eq. (2).

$$m_k = \arg\max_k \left[\alpha_k W_k(t) \frac{r_k(t)}{\overline{R_k(t)}} \right] \quad (2)$$

Where $r_k(t)$ is the instantaneous transmission rate of user k at time t, the factor $\overline{R_k(t)}$, that represents the past average throughput experienced by the k-th user at time t, and α_k weights metric so that k-th RT user with the most pressing needs in terms of delay threshold and acceptable loss rate, it can be given by

$$\alpha_k = \frac{-\log \sigma_k}{\tau_k} \quad (3)$$

where σ_k is QoS parameter of k-th user, and τ_k is the delay threshold for k-th user.

M-LWDF algorithm is an important guarantee to meet QoS requirements in RT service. The longer a packet waits in the current queue, the higher its priority will be. Nevertheless, packets delay in M-LWDF does not have a significant role. If a user cannot get a scheduling opportunity, a packet will be dropped from the queue of buffer due to deadline expiration. Therefore, on the basis of M-LWDF algorithm, an algorithm proposed in this paper solves the problems of packet loss caused by delay and unfair scheduling opportunity of edge users. An outline of in-slice scheduling as shown in Algorithm 2.

Algorithm 2 In-slice Scheduling

```
/*Simulation duration $T$, Number of UEs $N_{UE}$, Amount of RBs in $a$-th slice $S_{RB}$ */;
Input: $T$, $N_{UE}$, $S_{RB}$, $W_k(t)$, $\tau_k$, $a$;
Output: $m_k$
while $t < T$ do
    for $k = 20, 30, \ldots, 100$ do
        Update the priority of $k$-th UE;
        Get the CQI value of UEs feedback;
        Calculate $\overline{R_k(t)}$, $\tau_k$, $W_k(t)$ and $r_k(t)$ within $t_c$;
        Calculate $\mathrm{mod\_}W_k(t)$;
        Determine HARQ priority factor $\theta$;
        Calculate the weighted factor $\gamma$ and the predicted RBs $N_{i+1}$;
        Calculate the priority $m_k$ and allocate RBs;
        Delete the allocated RBs from RB group and determine whether the RB group is empty. If so, go to the next TTI; if not, go back to the previous step.
    end for
end while
```

Congestion and queuing are the main causes of packet loss in 5G video service. So, reducing PLR is one of the problems that the proposed algorithm must solve. The packet loss caused by too much queuing delay of users can be replaced by sigmoid function in the original algorithm. The modified queuing delay $\mathrm{mod\_}W_k(t)$ is:

$$\mathrm{mod\_}W_k(t) = \frac{1}{1 + \exp[-\psi(W_k - \tau_k)]} \tag{4}$$

Where ψ is an adjustable threshold factor with the range of (0, 1), which determines the degree of inclination of sigmoid function. The bigger ψ is, the steeper of the function is and the more sensitive the delay is. As can be seen from Fig. 3, with the increase of queue delay, in Eq. (4), users have higher satisfaction with delay. In other words, when the queue delay is close to the maximum delay the user can tolerate (τ_k), it can greatly improve the urgency of the packet, the priority of k-th user, and reduce PLR.

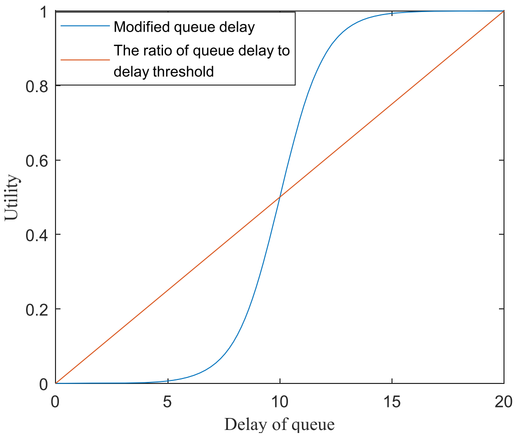

Fig. 3. The utility of W_k/τ_k and $\mathrm{mod}\_W_k(t)$

In addition to queuing delay, transmission delay in HARQ process also affect the priority of users. After the terminal receives downlink data, the MAC layer judges whether the data is retransmitted according to the New Data Indicator (NDI) sent by BS. Practically, retransmission packets priority are greater than the new transmission packets in order not to undermine the resources already used in the initial transmission of the packets [20]. Also, the delay of the retransmission packet is larger, so for meeting the demand of 5G low latency, such retransmitted data must be prioritized and scheduled. At this point, a HARQ priority factor θ is introduced to improve the priority of retransmitted data.

$$\theta = \begin{cases} 1, new \\ 2, retransmission \end{cases} \tag{5}$$

The factor θ is used to distinguish whether new or retransmitted packets received due to errors in the previous transmitted packets. In this paper, parameter θ is set as 1 for the new transmission packets and 2 for the retransmitted packets.

Since M-LWDF algorithm continuously schedules a user when current channel quality is great, other users cannot get fair opportunities. Therefore, a weighted factor γ is introduced to influence the fairness of users by combining the current channel quality and the state of RBs allocation in previous time.

$$\gamma_k(t) = \begin{cases} 1 - \frac{S_i}{N'_{i+1}}, CQI_k(t) \geq 10 \\ \exp\left(1 - \frac{S_i}{N'_{i+1}}\right), else \end{cases} \tag{6}$$

where $CQI_k(t)$ represents CQI value feedback from k-th user at the moment, S_i is the number of RBs allocated to k-th user in i-th time-slot, and N'_{i+1} is the number of available RBs predicted in the next slot. If equation $CQI_k(t) \geq 10$ is true, k-th user will be scheduled within several consecutive TTIs. At this point, the RBs allocation in the previous slot should be taken into account to influence the priority of such user. When the channel quality is poor, users at the edge of cell will not be scheduled for a long time. In order to avoid "starvation", an exponential function should be used to appropriately improve the priority of the users on the premise of considering the RA over time. With the increase of RBs allocation, the growth rate of weighting factor becomes slower. In other words, if the current channel quality is poor, the more RBs allocated in previous slot, the less priority of users will be improved. Conversely, the fewer resource blocks allocated, the greater the chance that users will be scheduled. Consequently, we use Bayes' theorem to calculate the probability of users in various states, and then the maximum likelihood estimation method is used to find the possible values of available resource blocks in the current slot. In this paper, the service status of the k-th user is represented by $s_k \cdot s_k = 0$ indicates that the user hasn't been allocated RBs, and $s_k = 1$ indicates that the user has been scheduled successfully. The conditional probability formula for the state of a user is:

$$P(s_k|N_i) = \begin{cases} \left(1 - \frac{1}{N_u}\right)^{N_i}, s_k = 0 \\ 1 - \left(1 - \frac{1}{N_u}\right)^{N_i}, s_k = 1 \end{cases} \tag{7}$$

Where N_u is the number of candidates which will be scheduled, and N_i is the number of available RBs in time t. Imagine there are p_1 users with $s_k = 0$, and p_2 users with $s_k = 1$, the user state probability is given by Eq. (8).

$$P(s|N_i) = P(s_k = 0|N_i)^{p_1} \cdot P(s_k = 1|N_i)^{p_2} \tag{8}$$

When the maximum value N_i of Eq. (8) is obtained, the estimated value N_i' of the number of available RBs is also clear.

$$N_i' = \arg\max \prod_{k=0}^{1} P(s|N_i) \tag{9}$$

The logarithm of both the left and right sides of Eq. (9) can be obtained as follows:

$$\ln N_i' = \arg\max \sum_{k=0}^{1} P(s|N_i) \tag{10}$$

Equation (11) can be obtained by deriving Eq. (10). This function is convex, so when the derivative is 0, N_i is the maximum value obtained by Eq. (10), namely N_i'.

$$\frac{\partial}{\partial N_i}\left(\sum_{k=0}^{1} \ln P(s|N_i)\right) = 0 \tag{11}$$

The number of RBs successfully allocated to users in the i-th slot is denoted as S_i. After the number of available resource blocks are obtained according to the scheduled state of users in i-th slot, the prediction of available RBs in next slot can be calculated by Eq. (12).

$$N'_{i+1} = N'_i - S_i \tag{12}$$

To sum up, the priority of proposed scheduling algorithm is:

$$m_k = \arg\max_k \left\{ \alpha_k \bmod \_W_k(t) \frac{r_k^{\theta}(t)}{R_k(t)} \gamma_k(t) \right\} \tag{13}$$

4 Numerical Results and Analysis

4.1 Simulation Parameters

We consider a wireless system with pedestrian and high moving users in a cell with radius of 500 m [21]. It is also assumed that the number of eMBB, mMTC and uRLLC users in the target cell is the same. Due to the random distribution of BS in a cell and users keep moving, the channel quality of each user is different, so the environment set in this paper is closer to the actual system (Table 1).

Table 1. Simulation parameters

Parameter	Value
Simulation duration	100TTIs
Number of UE	20, 30, 40…
TTI	1 ms
Radius of cell	500 m
Delay threshold	50 ms
Bandwidth	20 MHz
Subcarrier spacing	15 kHz
Number of RBs	100
Delay threshold factor ψ	0.9
Transmission power of UE	0.1 W

4.2 Simulation Results and Analysis

The algorithm proposed in this paper is to calculate the scheduling priority of users in N TTIs. The time complexity of a user's priority within a TTI is $O(n)$, so the time complexity of S-MLWDF algorithm is $O(n^2)$. The same is true for traditional PF and M-LWDF algorithms. Figure 4, Fig. 5, Fig. 6 and Fig. 7 depict the performance of the algorithm for resource utilization, fairness, throughput and PLR with the number of users increase from 20 to 100.

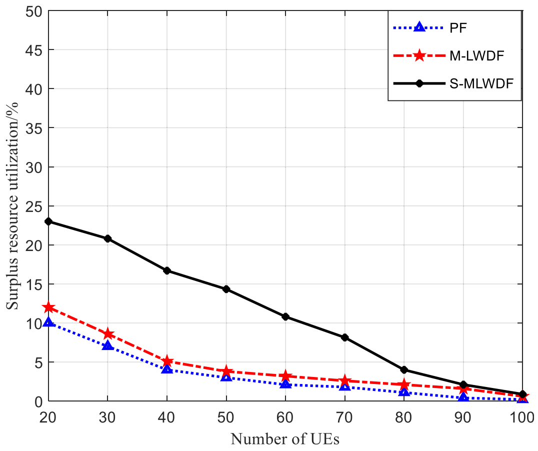

Fig. 4. Surplus resource utilization

Figure 4 presents the utilization of residual resource in the system for these three algorithm strategies. As shown in the simulation, the remaining resource utilization of three algorithms all decreases with the increase of the access users, but there is not much difference between PF and M-LWDF. This is because 5G slicing technology allocates RB groups for slices with the highest priority according to the actual needs in different services, so that the available RBs in each slice can meet the users' needs to the maximum extent. With the increase of users, the resources in the system are severely depleted, so the utilization of surplus resources in these three algorithms tend to be similar. In general, compared with PF and M-LWDF algorithms, S-MLWDF algorithm improves the residual resource utilization by about 8% and 7% on average.

Figure 5 shows the fairness index between users. With the increase of the users, the fairness of PF algorithm and M-LWDF algorithm decreases sharply, while the proposed improved algorithm S-MLWDF performs better than PF and M-LWDF with aggregate percentage is increased by 13% and 10%. Because users are distributed in different locations in cell, the distance from BS is random, and the channel quality is different. When the number of users increases, such users who are at the edge of cell are given priority in the process of in-slice scheduling for take care of them. It can ensure the fairness of scheduling users with poor channel quality or serious queuing delay.

The throughput is presented in Fig. 6. It can be seen that the throughput of three algorithms is proportional to the number of users. Compared with the other two algorithms, S-MLWDF algorithm improves the system throughput by about 33% and 27%. However, as the number of users of the network continue to increase, the RBs available in the system are not sufficient to support all service requirements. By now the network slicing technology can maximize the resource utilization by grouping RBs according to the different demands of access services, so as to make effective use of the empty space resources in the cell. Therefore, slicing-oriented scheduling algorithm is more suitable for the actual scene.

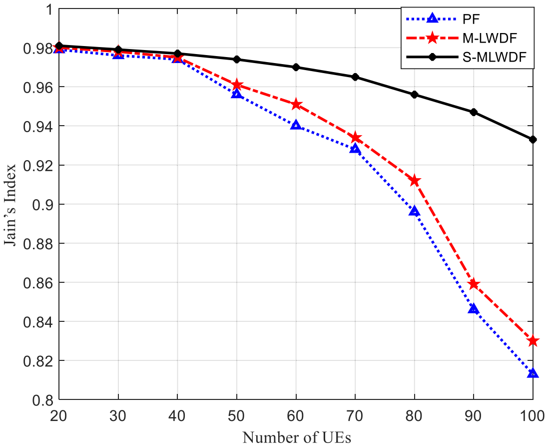

Fig. 5. Fairness

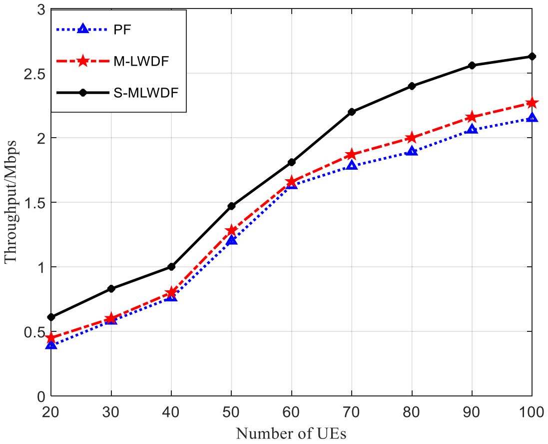

Fig. 6. Throughput

The PLR for all users is shown in Fig. 7. The lower the PLR for each algorithm, the better the performance of the algorithm in terms of PLR. The results show that the PLR of S-MLWDF algorithm is about 22% and 16.3% lower than PF and M-LWDF algorithm, respectively. As more and more users applying for accessing the system, the queuing delay of users increases. When delay approaches the threshold, the corresponding packets will be discarded. Meanwhile, the transmission delay of HARQ retransmission packets will also increase dramatically in PLR. The process of in-slice scheduling in S-MLWDF greatly improves the urgency of the packets and the priority of HARQ retransmission packets when approaching the deadline expiration, so the proposed algorithm can support more users for scheduling, and send more packets in the case of limitation for the number of RBs.

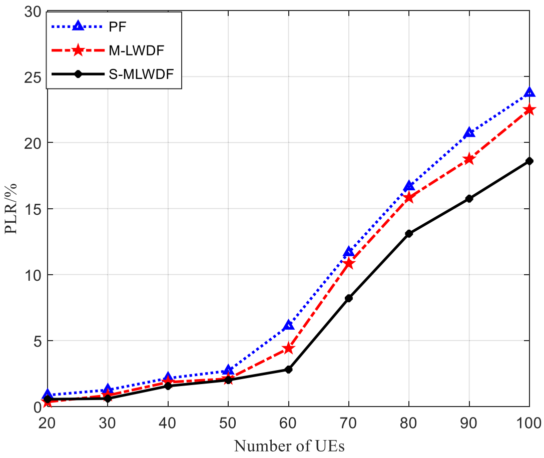

Fig. 7. PLR

5 Conclusion

In this paper, the performance of 5G downlink resource scheduling algorithm is studied. Combing 5G network slicing which is widely used with M-LWDF algorithm, we proposed a novel scheduling strategy named as S-MLWDF algorithm which fully considers the delay, current channel quality status, RBs allocation and the available RBs in next time-slot. At the same time, PF and M-LWDF algorithms are used as comparisons to simulate and analyze for resource utilization, fairness, throughput and PLR. Simulation results show that the proposed algorithm is feasible. Under the environment of multi-users, S-MLWDF algorithm performs better than PF and M-LWDF algorithm in improving fairness between users who are at the edge of cell, reducing PLR caused by delay of queues. Simultaneously, it can satisfy QoE for different users and save a lot of network resources. Besides, the computational complexity is not high, and there are unique advantages for the development of varied services in 5G scenarios. In future work, we need to consider the use of buffer state while scheduling is obligatory to maximize the resource utilization.

Acknowledgement. This work comes from a major project in Chongqing, “R&D and application of 5G road test instruments (No. cstc2019jscx-zdztzxX0002)”.

References

1. Gupta, A., Jha, R.K.: A survey of 5G network: architecture and emerging technologies. IEEE Access **3**, 1206–1232 (2015)
2. Foukas, X., Patounas, G., Elmokashfi, A., et al.: Network slicing in 5G: survey and challenges. IEEE Commun. Mag. **55**(5), 94–100 (2017)

3. Comsa, I., Zhang, S.J., Aydin, M.E.: Towards 5G: a reinforcement learning-based scheduling solution for data traffic management. IEEE Trans. Netw. Serv. Manag. **15**(4), 1661–1675 (2018)
4. Wang, C., Shih, K., Li, C.: User location recommendation combined with MLWDF packet scheduling in LTE downlink communication. In: 2018 Asia-Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA ASC), pp. 450–453. IEEE (2018)
5. Samia, D., Ridha, B.: A new scheduling algorithm for real-time communication in LTE networks. In: 2015 IEEE 29th International Conference on Advanced Information Networking and Applications Workshops, pp. 267–271. IEEE (2015)
6. Shen, L., Wang, T., Wang, S.: Proactive proportional fair: a novel scheduling algorithm based on future channel information in OFDMA systems. In: 2019 IEEE/CIC International Conference on Communications in China (ICCC), pp. 925–930. IEEE (2019)
7. Nasralla, M.M., Rehman, I.U.: QCI and QoS aware downlink packet scheduling algorithms for multi-traffic classes over 4G and beyond wireless networks. In: 2018 International Conference on Innovation and Intelligence for Informatics, Computing, and Technologies (3ICT), pp. 1–7. IEEE (2018)
8. Ferdosian, N., Othman, M., Ali, B.M., et al.: Fair-QoS broker algorithm for overload-state downlink resource scheduling in LTE networks. IEEE Syst. J. **15**(4), 3238–3249 (2018)
9. Rocha, F.G.C., Vieira, F.H.T.: A channel and queue-aware scheduling for the LTE downlink based on service curve and buffer overflow probability. IEEE Wirel. Commun. Lett. **8**(3), 729–732 (2019)
10. Madi, N.K.M., Hanapi, Z.M., Othman, M., Subramaniam, S.K.: Delay-based and QoS-aware packet scheduling for RT and NRT multimedia services in LTE downlink systems. EURASIP J. Wirel. Commun. Netw. **2018**(1), 1–21 (2018). https://doi.org/10.1186/s13638-018-1185-3
11. Padmavathy, C., Jayashree, L.S.: An enhanced delay sensitive data packet scheduling algorithm to maximizing the network lifetime. Wireless Pers. Commun. **94**(4), 2213–2227 (2016). https://doi.org/10.1007/s11277-016-3376-8
12. Ferreira, F.A., Guardieiro, P.R.: A new channel-aware downlink scheduling algorithm for LTE-A and 5G HetNets. In: Chaubey, N., Parikh, S., Amin, K. (eds.) COMS2 2020. CCIS, vol. 1235, pp. 173–183. Springer, Singapore (2020). https://doi.org/10.1007/978-981-15-6648-6_14
13. Ksentini, A., Frangoudis, P.A., Nikaein, N., et al.: Providing low latency guarantees for slicing-ready 5G systems via two-level MAC scheduling. IEEE Network **32**(6), 116–123 (2018)
14. Abdelhamid, A., Krishnamurthy, P., Tilpper, D.: Resource allocation for heterogeneous traffic in LTE virtual networks. In: 2015 16th IEEE International Conference on Mobile Data Management (MDM), vol. 1, pp. 173–178. IEEE (2015)
15. Kayali, M.O., Shmeiss, Z., Safa, H., et al.: Downlink scheduling in LTE: challenges, improvement, and analysis. In: 2017 13th International Wireless Communications and Mobile Computing Conference (IWCMC), pp. 323–328. IEEE (2017)
16. Hadar, I., Raviv, L., Leshem, A.: Scheduling for 5G cellular networks with priority and deadline constraints. In: 2018 IEEE International Conference on the Science of Electrical Engineering in Israel (ICSEE), pp. 1–5. IEEE (2018)
17. Pedersen, K., Pocovi, G., Steiner, J., et al.: Agile 5G scheduler for improved E2E performance and flexibility for different network implementations. IEEE Commun. Mag. **56** (3), 210–217 (2018)

18. Mamane, A., Ghazi, M.E., Barb, G., et al.: 5G heterogeneous networks: an overview on radio resource management scheduling schemes. In: 2019 7th Mediterranean Congress of Telecommunications (CMT), pp. 1–5. IEEE (2019)
19. 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. **15**(2), 678–700 (2013)
20. Nnamani, C.O., Anioke, C.L., Ani, C.I.: Improved MLWDF scheduler for LTE downlink transmission. Int. J. Electron. **103**(10–12), 1857–1867 (2016)
21. Nguyen, T.T., Ha, V.N., Le, L.B.: Wireless scheduling for heterogeneous services with mixed numerology in 5G wireless networks. IEEE Commun. Lett. **24**(2), 410–413 (2020)