The Korean Institute of Communications and Information Sciences

Available online at www.sciencedirect.com

ScienceDirect

ICT Express 4 (2018) 175–180

www.elsevier.com/locate/icte

QoS-guaranteed scheduling for small cell networks

Een-Kee Hong[a], Jung-Yeon Baek[a,*], Ye-Ok Jang[a], Jee-Hyeon Na[b], Kyung-Sook Kim[b]

[a] *Department of Electronic Engineering, KyungHee University, Yongin, Republic of Korea*
[b] *ETRI, Daejeon, Republic of Korea*

Received 28 February 2017; accepted 30 November 2017
Available online 1 January 2018

Abstract

This paper proposes a QoS-guaranteed scheduling algorithm that considers the available resources and traffic load of small cell. Even though the available resources are scarce or traffics to be scheduled are overloaded, Guaranteed Bit Rate (GBR) traffic such as VoIP and video should be supported. The proposed algorithm provides a flexible weight ratio between GBR and Non-GBR traffic according to resources and traffic load. Furthermore, since QoS is considered by simple QoS Class Identifier (QCI), low complexity is achieved. Simulation results show that the proposed algorithm improves QoS including delay and Packet Loss Rate (PLR) performance effectively.

Keywords: QoS; Scheduling; Traffic load; Guaranteed Bit Rate; Small cell

1. Introduction

In recent years, with the increasing number of smart devices and contents, mobile traffic is rapidly increasing. There are some solutions to accommodate the continuous growth of traffic demand, such as securing wider bandwidth, segmenting use of frequency resources, advanced modulation schemes and cell densification. Among them, small cell technology, referred to as cell densification, is considered one of the most promising options due to its various methods of deployment and cost-efficiency. This technology, which locates small cells on hotspots or in coverage holes of macro cells, can increase the network capacity and improve spectrum efficiency by increasing the number of overlaid small cell Base Stations (BSs). In a small cell network, however, macro cells might seriously interfere with small cells since these cells share radio resources. In order to solve this problem, an Inter Cell Interference Coordination (ICIC) technique known as resource partitioning is used to separate radio resources between macro and small cells [1,2].

There are three approaches in resource partitioning, Co-channel Deployment (CCD), Orthogonal Deployment (OD) and Partially Shared Deployment (PSD). When the total number of resource block groups (RBGs) is M, macro and small cell BSs simultaneously use all available radio resources (i.e., M RBGs) in the CCD approach. In OD, small cell BSs use K RBGs, and macro BSs use M–K RBGs. Since they each use an exclusive set of resources, there is no interference between macro and small cells. In contrast, in PSD, macro and small cells share K RBGs when M–K RBGs are assigned to the macro cell only or vice versa. In this approach, capacity gain can be achieved by sharing resources [3].

As the various applications have appeared, satisfying user QoS requirements becomes more and more important. QoS indicates the service qualities that are expressed in terms of delay, data rate, and packet loss rate. Nine classes of QoS have been identified in the QoS Class Identifier (QCI) table defined by 3GPP according to traffic type, priority, delay budget, and acceptable packet loss rate [4]. Also, traffic type is divided into Guaranteed Bit Rate (GBR) and Non-GBR. GBR traffic such as VoIP (Voice over Internet Protocol), video, and games requires the minimum data rate. Non-GBR traffic corresponds to the best

---

[*] Corresponding author.
E-mail address: mmmmjy119@gmail.com (J.-Y. Baek).
Peer review under responsibility of The Korean Institute of Communications and Information Sciences (KICS).

https://doi.org/10.1016/j.icte.2017.11.017

effort services (e.g., FTP, HTTP, e-mail), which may not need any guaranteed rate [4].

Scheduling algorithms can be classified in three groups depending on channel-awareness and QoS-awareness. First, a general scheduler of channel-unaware groups is Round Robin (RR) [5]. This scheduling algorithm provides the best performance in terms of fairness, but it does not consider channel condition, amount of available resources, and requirements characterized by traffic type. The primary representatives of scheduling algorithms accounting for channel quality are Maximum Throughput (MT) and Proportional Fairness (PF) [6]. MT scheduling, referred to as greedy scheduling, maximizes system throughput by assigning a resource to a user who has the best channel quality in a Transmission Time Interval (TTI). In this case, however, some users with bad channel quality may not be assigned. PF scheduling includes the average throughput in the MT metric in order to provide fairness [7]. However, like MT scheduling, it cannot guarantee QoS and does not consider network conditions such as traffic load and the amount of resources for small cells [8]. Some schedulers consider either channel quality or QoS. Modified-Largest Weighted Delay First (M-LWDF) and Exponential/PF (EXP/PF), introduced in [9,10], were designed to provide packet latency guarantees, lower packet loss rate, and good fairness by using the acceptable packet loss probability and Head of Line (HOL) delay. In [11], Frame Level Scheduler (FLS) is proposed, in which each user calculates the required data rate to meet their own delay budget on the upper layer, and resources are assigned to selected users at every TTI on the lower layer. Until all real-time traffic is transmitted, the MT metric is used to schedule users, and then the remaining resources are assigned by PF metric. Those schedulers account for QoS requirements or queue status, but they do not consider network conditions such as the amount of radio resources allocated to small cells or traffic load to be scheduled. Since the scheduler is located on the base station, network conditions related to the base station directly affect the scheduling process. For example, assuming constant number of users and traffic load generated by those users, not only non-GBR traffic, but also GBR traffic will lose the chance to be scheduled as the available resources decrease. That is, GBR traffic becomes difficult to guarantee because the available resources decrease as all available network resources decrease. In a resource-partitioning scheme between macro and small cells, the scheduler in small cell BS should consider the amount of radio resources allocated to a small cell.

In this paper, we propose the QoS scheduling algorithm for small cells in OD resource partitioning network configuration, where the amount of radio resources and traffic load are taken into account. As the radio resources allocated to small cells are reduced, the priority of GBR traffic increases. In addition, when the radio resources are expected to be sufficient compared to the traffic load for scheduling, the conventional PF metrics are used for small cell scheduling to achieve low computational complexity.

Also, most QoS scheduling algorithms based on some procedure to measure variables such as priority, and HOL delay, and to calculate the metric using these variables require high complexity and scheduler processing burden. In uplink scheduling, it is particularly difficult to estimate some variables such as HOL delay [12–14]. In the proposed algorithm, the values of priority in the QCI table are transformed to scheduling metrics to achieve low-complexity QoS scheduling.

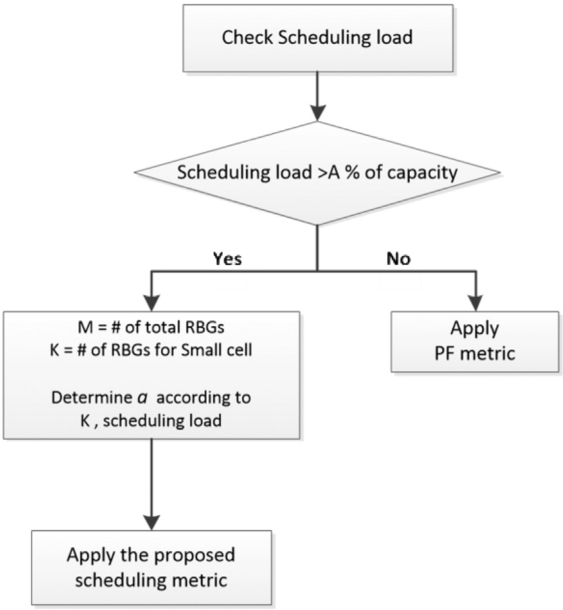

Fig. 1. Flow chart of proposed scheduling algorithm.

2. Proposed QoS scheduling algorithm

This paper proposes an effective scheduling algorithm reflecting on the amount of radio resources and traffic load of small cell BSs. Even though resources are scarce or traffic is overloaded, GBR traffic that requires a minimum data rate and quality should be supported. The proposed scheduling algorithm provides flexible scheduling weight for GBR traffics according to the amount of radio resources and traffic load. Fig. 1 shows the flow chart of the proposed algorithm.

The scheduling metric of the ith UE for the jth RBG is formulated by

$$m_{i,j} = \begin{cases} \dfrac{d_{i,j}}{\overline{R}_{i,j}}, \textit{Traffic load} < A\% \textit{ of capacity} & \text{(a)} \\ \left(\dfrac{M}{K}\right)^{\frac{(9-priority_{i,j})}{\beta}} \cdot \dfrac{d_{i,j}}{\overline{R}_{i,j}}, & \\ \quad \textit{Traffic load} \geq A\% \textit{ of capacity}, & \text{(b)} \end{cases} \tag{1}$$

where d_{ij} is the achievable data rate of the ith UE and $\overline{R}_{i,j}$ is the past average throughput of the ith UE. Also, M is the number of total resource block groups and K is the number of resource block groups for small cell. $Priority_{i,j}$ is an integer value that is given in the QCI table (priority $1, 2, \ldots, 9$) [4]. β is the scale parameter depending on K and traffic load.

When the radio resources are sufficient to support traffic load for scheduling, it is possible to guarantee QoS of GBR traffic while also offering good-quality non GBR traffic. However,

Table 1
QCI table [4].

QCI	Resource type	Priority	Packet delay budget	Packet error loss rate
1	GBR	2	100 ms	10^{-2}
2		4	150 ms	10^{-3}
3		3	50 ms	10^{-3}
4		5	300 ms	10^{-6}
5	Non-GBR	1	100 ms	10^{-6}
6		6	300 ms	10^{-6}
7		7	100 ms	10^{-3}
8		8	300 ms	10^{-6}
9		9	300 ms	10^{-6}

user competition for resource assignment increases as traffic load increases, and competition becomes more intense for high traffic load. In that case, QoS guarantee of GBR traffic becomes much more difficult. Scheduling metrics are changed based on the amount of traffic load and allocated radio resources, expressed by 'A% of network capacity' in the proposed algorithm. After setting the criterion for A% of network capacity, the proposed scheduling metric including higher weight for GBR traffic is applied if traffic load is greater than A%. Otherwise, the conventional PF metric is applied to achieve low computational complexity for scheduling. Next, we calculate (1)(b) with a scale factor β depending on traffic load and the amount of available resources. In (1)(b), M is the number of total RBGs determined according to the network bandwidth, and K indicates the number of available RBGs for the small cell. The priority values are the same as those listed in Table 1. As mentioned above, GBR and non-GBR traffic have less scheduling opportunity when the available resources for small cells decrease. Thus, the proposed scheduling algorithm increases the weighting factor assigned to users with a high priority in order to guarantee GBR traffic as the resources are reduced.

The values of priority in Table 1 are properly converted into a scheduling metric by scaling with the value of β. Through a system-level simulation, we obtained a value of β that shows the smallest average delay of GBR traffic. In addition, the value of β depends on the traffic load, which we divide the values over several groups. That is, the range of traffic load has a similar value to β that shows the best performance in terms of average delay for a group. QoS performance of GBR traffic is therefore improved using this algorithm. The proposed scheduling algorithm properly guarantees QoS of GBR traffic using the scaling parameter β, determined by the amount of available resources and traffic load, while the previous QoS scheduling algorithms (e.g., PF, MLWDF) do not consider the network condition of small cells. Moreover, the proposed scheduling method is effective for small cell networks since the values of scheduling metrics are easily computed to reduce the scheduling burden. Especially in uplink, an additional algorithm for delay estimation is essential in other QoS scheduling algorithms such as MLWDF and EXP/PF [9,10]. The proposed scheduling algorithm does not need additional estimation since it uses priority values in QCI table.

3. Performance evaluation

In this section, we analyze the performance of the proposed scheduling algorithm using System Level Simulation (SLS) and demonstrate the superiority of our algorithm compared with PF and MLWDF scheduling algorithms. In the simulation, users who are randomly distributed on a small cell share the radio resources, and there is no interference between macro and small cells due to an OD partitioning network configuration. Radio resources are assigned to the user by the RBG unit, and both Adaptive Modulation and Coding (AMC) and Hybrid ARQ are supported in data transmission. If the computed BLER after transmission is higher than the target BLER (10%), information including the assigned RBG index, received time, Modulation Coding Scheme (MCS) level and Effective SINR (ESINR) is stored in a circular buffer for HARQ. After 8 ms of HARQ Round Trip Time (RTT), data kept in the buffer are retransmitted [15]. Scheduling is performed every 1 ms, and detailed parameters are shown in Table 2.

Table 2
Simulation parameters.

Parameter	Value
System bandwidth	10 MHz
Number of RBs	50 RBs
Number of RBs per RBG	3 RBs
Number of occupied subcarrier	601 subcarriers
TTI (Transmission Time Interval)	1 ms
Carrier frequency	2 GHz
Velocity	3 km/h
Cell radius	200 m
Number of UEs	10
UE/BS transmit power	23 dBm/30 dBm
BS antenna gain	5 dBi

3.1. Downlink performance results

First, we evaluate the proposed scheduling algorithm in terms of average delay according to the amount of available resources for a small cell, K. Figs. 2 and 3 show the average delay performance. As the available resources are reduced, the average delay increases regardless of traffic type. In case of GBR traffic, the proposed scheduling algorithm improves delay performance as K decreases, with no noticeable difference among schedulers when the resources are sufficient. In return for guaranteeing QoS of GBR traffic, the delay in non-GBR traffic worsens when the proposed scheduling algorithm is applied. The reason for this is that the proposed algorithm supports GBR traffic more than non-GBR traffic in limited

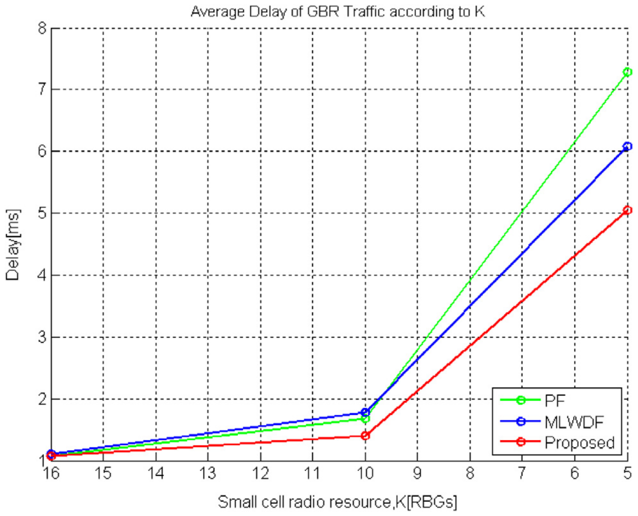

Fig. 2. Average delay of GBR traffic (downlink).

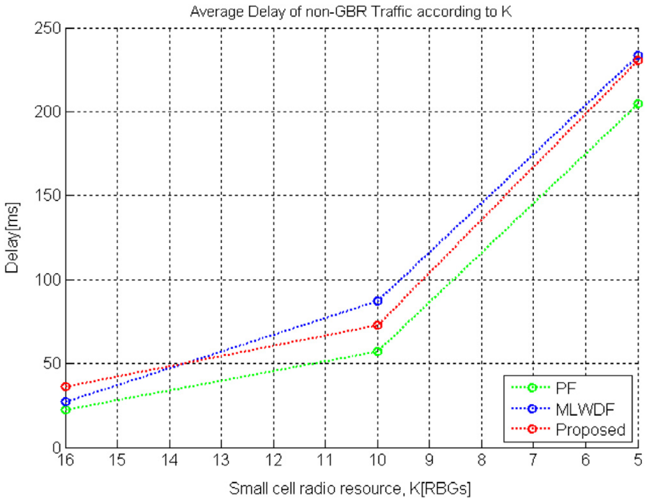

Fig. 3. Average delay of non-GBR traffic (downlink).

resource constraints, resulting in non-GBR traffic latency. On the other hand, in PF scheduling, there is no difference between GBR and non-GBR since the PF scheduler does not account for QoS guarantees. In Fig. 4, the delay performance of non-GBR traffic shows similar tendency to that of PF since the priority of non-GBR traffic is low and the scheduling metric in Eq. (1)(b) is similar to the PF (see Figs. 5–7 and 9).

It is natural that for Packet Loss Rate (PLR) increases as the number of RBGs, K, decreases. Fig. 4 shows that the proposed scheduling algorithm improves PLR since the GBR traffic can utilize more radio resources than those of PF and MLWDF. When the small cell utilizes a part of whole resources ($5 \leq K \leq 15$), the proposed scheduling algorithm shows best PLR performance for GBR traffic. When the number of RGBs is 16, the three algorithms show similar PLR performances because the radio resources are sufficient. When the resources are scarce ($K \leq 3$), the PLR performances of the three algorithms are also similar due to insufficient absolute radio resources. For non-GBR traffic, when the proposed scheduling algorithm is applied, there is some loss in PLR performance at the cost of QoS support for GBR traffic. The PF scheduler does not distinguish GBR and non-GBR traffic.

3.2. Uplink performance result

When the traffic load is constant, a longer delay in average transmission time is expected as K decreases. The PF scheduler has a longer delay than MLWDF and the proposed scheduler because it does not consider characteristics of GBR traffic (e.g., delay, priority). On the other hand, unlike the MLWDF scheduler which accounts only for delay, the proposed scheduler reduces average delay time by guaranteeing GBR traffic with priority and β which is set to have a minimized delay. Since the proposed scheduler provides a higher QoS for GBR traffic by increasing the weight, it achieves better performance in terms of delay compared to the MLWDF scheduler even if the delay performance of non-GBR traffic worsens.

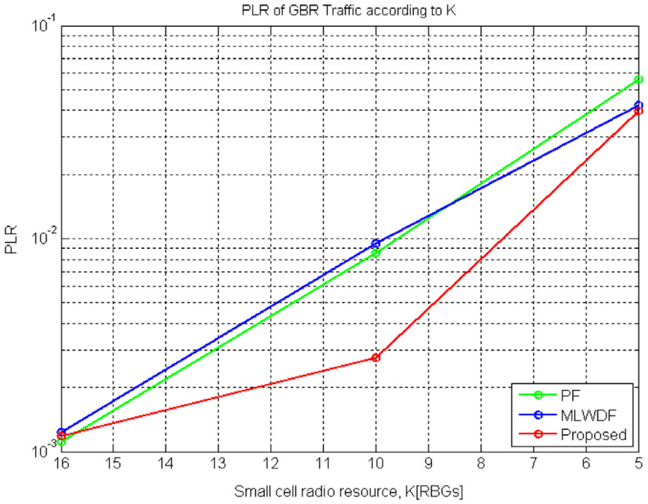

Fig. 4. PLR performance of GBR traffic (downlink).

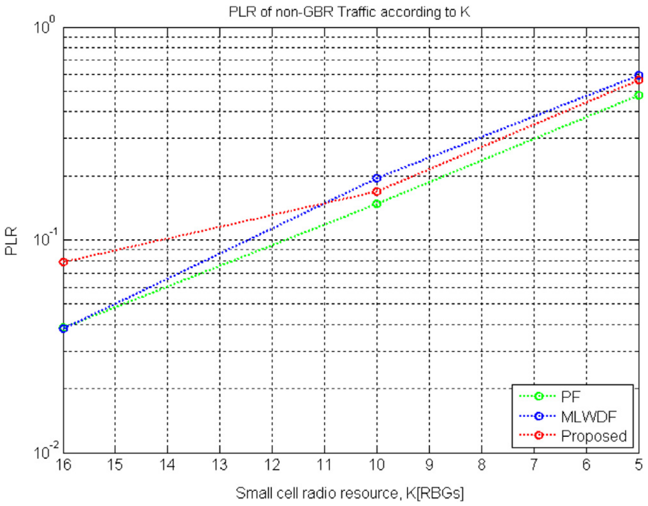

Fig. 5. PLR performance of non-GBR traffic (downlink).

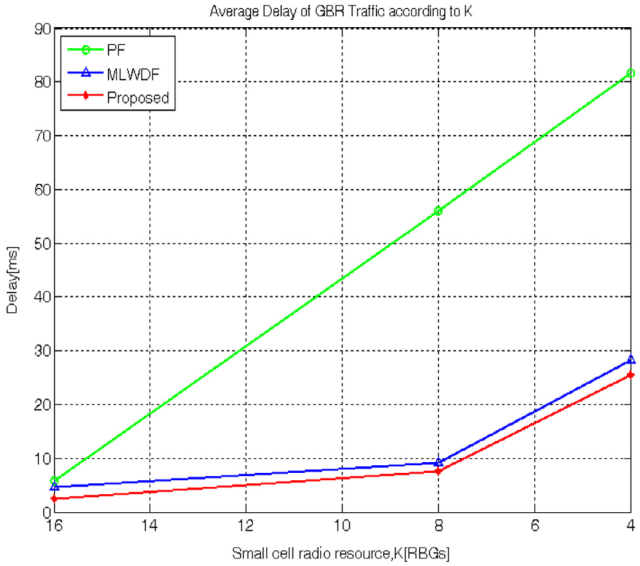

Fig. 6. Average delay of GBR traffic (uplink).

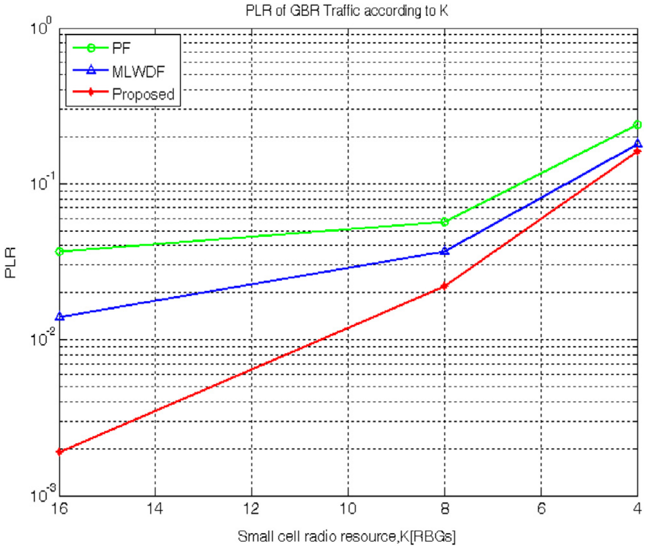

Fig. 8. PLR performance of GBR traffic (uplink).

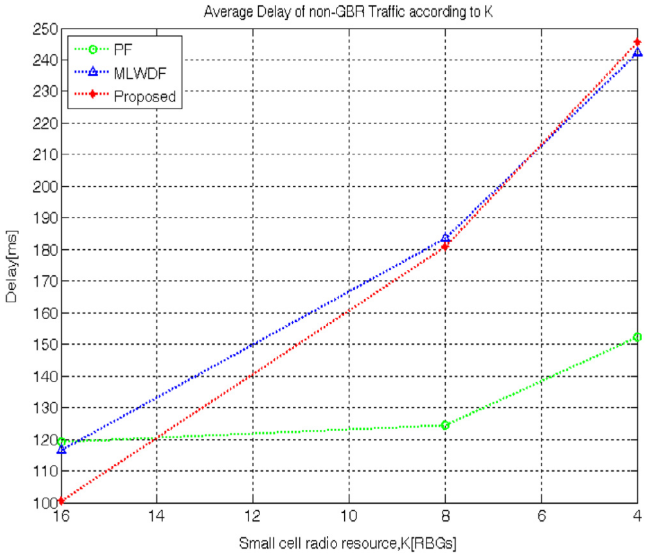

Fig. 7. Average delay of non-GBR traffic (uplink).

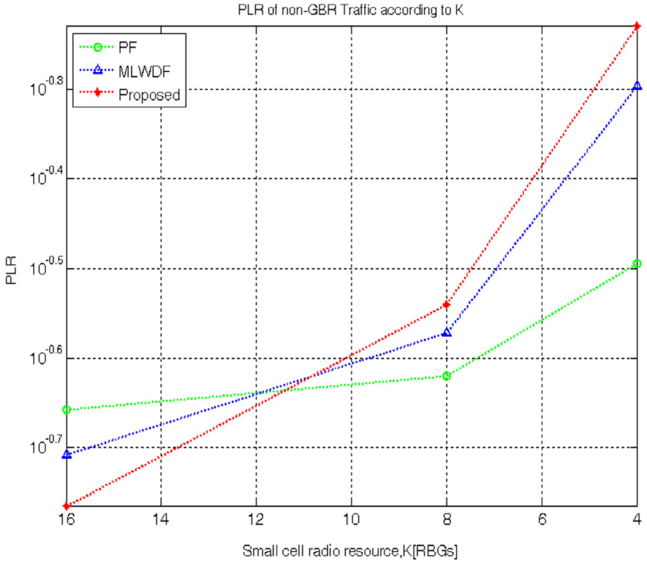

Fig. 9. PLR performance of non-GBR traffic (uplink).

Increasing the delay in average transmission time results in a high probability that packet delivery time will exceed the delay budget, which degrades PLR performance. As explained in the downlink case, the performances of delay and PLR worsen when the amount of available resources is reduced. Even if delay and PLR are degraded, however, the proposed scheduling algorithm provides better PLR performance for GBR traffic than those of PF and MLWDF schedulers, as shown in Fig. 8.

4. Conclusions

In this paper, we propose a novel QoS-guaranteed scheduling algorithm that adjusts the scheduling metrics for GBR traffic according to the amount of available resources and traffic load of small cell BSs. When OD resource partitioning network is configured in a heterogeneous network, the amount of resources for small cells is variable. Although the resources are reduced, GBR traffic (e.g., VoIP, videos, games) should be guaranteed. Traffic load is considered as well as available radio resources. We formulated a scheduling metric that considers available resources, traffic load, and priority of services in the QCI table. The values of priority are properly converted to the scheduling metric using a scaling factor, β, which is chosen to provide the minimum delay in average transmission time. Using a system-level simulation, we demonstrate that the proposed scheduling algorithm improves delay and PLR performance for GBR traffic effectively.

Acknowledgments

This work was supported by Institute for Information & communications Technology Promotion (IITP) grant funded by the Korea government (MSIT) (No. 2014-0-00282, Development of 5G Mobile Communication Technologies for Hyper-connected smart services).

Conflict of interest

The authors declare that there is no conflict of interest in this paper.

References

[1] A. Damnjanovic, et al., A survey on 3GPP heterogeneous networks, IEEE Wirel. Commun. 18 (3) (2011) 10–21.

[2] Matheus C. Nogueira, et al., QoS aware schedulers for multi-users on OFDMA downlink: Optimal and heuristic, IEEE LATINCOM, Nov. 2016, pp. 1–6.

[3] D. Fooladivanda, C. Rosenberg, Joint resource allocation and user association for heterogeneous wireless cellular networks, IEEE Trans. Wireless Commun. 12 (1) (2013) 248–257.

[4] 3GPP TS 23.203 v13.0.0., Policy and charging control architecture, June 2014.

[5] S. Schwarz, et al., Low Complexity Approximate Maximum Throughput Scheduling for LTE, IEEE ASILOMAR, Pacific Grove, CA, USA, Nov. 2010, pp. 1563–1569.

[6] F. Capozzi, et al., Downlink packet scheduling in LTE cellular networks: Key design issues and a survey, IEEE Commun. Surv. Tutor. 15 (2012) 678–700.

[7] P. Bender, et al., CDMA/HDR: A Bandwidth-efficient high-speed wireless data service for nomadic users, IEEE Commun. Mag. 38 (2000) 70–77.

[8] Seungwan Ryu, et al., Wireless packet scheduling algorithm for ofdma system based on time-utility and channel state, ETRI J. 27 (6) (2005) 777–787.

[9] M. Andrews, et al., Providing quality of service over a shared wireless link, IEEE Commun. Mag. 39 (2) (2001) 150–154.

[10] R. Basukala, et al., Performance analysis of EXP/PF and M-LWDF in downlink 3GPP LTE system, IEEE AH-ICI, Kathmandu, Nepal, Nov. 3–5, 2009, pp. 1–5.

[11] G. Piro, et al., Two level downlink scheduling for real-time multimedia services in LTE networks, IEEE Trans. Multimedia 13 (5) (2011) 1052–1065.

[12] Sung Oh Kwon, Neung Hyung Lee, Uplink QoS Scheduling for LTE System, IEEE VTC2011, Yokohama, Japan, 2011, pp. 1–5.

[13] A. Baid, et al., Delay estimation and fast iterative scheduling policies for LTE uplink, IEEE WiOpt, Paderborn, Germany, May 2012, pp. 89–96.

[14] H. Safa, et al., A QoS-Aware Uplink Scheduling Paradigm for LTE Networks, IEEE AINA, Barcelona, Spain, Mar. 25–28, May 2011, pp. 1097–1104.

[15] Ye Ok Jang, et al., Effect of interference mitigation technique and performance analysis for small cell in homogeneous networks, J. KICS 39A (10) (2014) 937–945.