Dynamic slot-based carrier scheduling scheme for downlink multimedia traffic over LTE advanced networks with carrier aggregation

Ramachandran VIJAYARANI*, Lakshmanan NITHYANANDAN
Department of Electronics and Communication Engineering, Pondicherry Engineering College, Puducherry, India

Abstract: Carrier aggregation (CA) in long-term evolution advanced (LTE-A) networks is a promising tool for providing bandwidth extension for high data rate communication. In CA, the efficient scheduling of carriers has imposed a new challenge in the design of delay-sensitive carrier scheduling for mobile multimedia applications. In this paper, a new dynamic slot-based carrier scheduling scheme has been proposed for efficient resource allocation to support the quality of service (QoS) of mobile traffic over CA transmission. The key idea of this scheduling is the allocation of slots for each component carrier based on the queue state information, taking into consideration the priority based on the delay requirements of individual mobile multimedia traffic. Two different cases of the scheduler with and without priority are implemented to facilitate the performance comparison. The simulation results show that the proposed scheme provides better spectrum utilization, throughput, and delay performance than the existing scheduling scheme and improves delay-sensitive mobile traffic QoS.

Key words: Carrier aggregation, carrier scheduling, multimedia traffic, quality of service, queue state information, priority

1. Introduction

For future information and communication technologies, the major challenges are the enormous growth of user demands for ubiquitous access and high data rate multimedia traffic. In practice, it is unendurable to impose fixed delay requirements for various delay-sensitive real-time multimedia traffic volumes due to the time-varying nature of the underlying wireless channel. Altogether, these phenomena influence QoS as perceived by end users. Real-time delay-sensitive applications, such as voice over internet protocol (VoIP) and multimedia teleconferencing videos, have very distinct QoS requirements. Practically, delay and system throughput need to be leveraged in order to implement carrier aggregation (CA) to support mobile multimedia traffic applications [1]. Data delivery with a low latency is required to preserve the interactive nature of the video applications.

Aggregation of frequency spectrum is one of the viable techniques for a long-term evolution advanced (LTE-A) network to enhance the peak data rates [2]. Consequently, a new CA scheme has been proposed by 3GPP to achieve bandwidth extension up to 100 MHz to enhance the peak data rates [3]. In CA, the wider bandwidth between evolved universal terrestrial radio access network NodeB (eNB) and the user equipment (UE) is attained with more than one aggregated component carrier (CC), either contiguous or noncontiguous spectrum bands [4]. In Release-10, CA is employed to allow UE to use multiple carriers for high data rate

---

*Correspondence: vijee_er@pec.edu

communication at the cost of energy. When CA is implemented, an efficient carrier scheduling (CS) scheme is essential to achieve efficient and fair utilization of the wireless spectrum. This has imposed a new challenge in the design of CS in downlink transmission for mobile multimedia applications. Hence, a well-designed CS scheme is essential to meet the reliable transmission and low cost of deployment in LTE-A systems. Different carrier scheduling algorithms [5–11] have been proposed, which schedule users depending on the available bandwidth to enhance the throughput of the system. Unfortunately, all the existing CA-based CS schemes barely take the delay requirements of real-time mobile multimedia applications into consideration.

In this paper, a new dynamic slot-based CS (DSCS) scheme is proposed, with the aim of enhancing the delay performance of multimedia traffic in multicarrier downlink networks. In the proposed scheme, CS is achieved through the slot allocation method with and without priority consideration, which schedules CCs required for packet transmission for each user in a particular duration based on the status of the queue. This slot-based allocation does not need any information about the flows' states. Furthermore, to provide QoS for delay-sensitive traffic applications, priority-based packet scheduling is incorporated, which classifies the incoming traffic into different priority levels according to their delay tolerance [12], and the resource blocks (RBs) are assigned accordingly.

This proposed DSCS is a straightforward CS scheme to manage network traffic based on priority over multiple CCs. In this scheme, each user receives the data from all CCs simultaneously and continuously. However, this aggregation of multiple CCs maximizes the spectral efficiency under the given opportunity of the resource scheduling strategy. Meanwhile, it also helps to achieve saturated resource utilization. In other words, by using a slotted scheme, no resource could be wasted as long as $\sum_{l=1}^{L} Q_l(t) \neq 0$. In this paper, spectral efficiency, throughput, and average delay performance of various network traffic are investigated for the proposed scheme with and without priority and the results are compared with those of an existing dynamic aggregation carrier (DAC)/ideally balanced system (IBS) CS scheme.

The rest of this paper is organized as follows. Section 2 reviews some related work. Section 3 presents the system model. Section 4 explains the proposed scheme and problem definition. Section 5 describes the analysis of the proposed scheme in detail. Section 6 presents the simulation results and discussions. Finally, conclusions are drawn in Section 7.

2. Related work

Independent link adaptation performed per CC facilitates transmission optimization on different CCs according to the known channel conditions. Individual CCs with different transmission powers could provide different levels of coverage [5]. Besides, to achieve high data rates and wider coverage, the capability of aggregating multiple carriers provides the flexibility to deal with required bandwidth extensions [6]. In [7], multiple CC allocation and adaptive adjustment of transmission parameters for different CCs were jointly considered. Since all CCs are assigned to users, trunking efficiency and packet scheduling gain are maximized [8]. CS could be done either jointly or independently on CCs [5]. To achieve higher resource utilization, a CS scheme, termed as separated random user scheduling and separated burst level scheduling, is proposed with acceptable complexity [9]. In [10], a DAC scheme is proposed to provide better resource balance in wireless networks for elastic traffic using IBS; however, this method is impractical due to the change of primary CCs to idle CCs. Moreover, in this method, the packets are served by primary CCs as long as the queue of any secondary CC becomes empty, which increases the average sojourn time. The dynamic carrier aggregation scheme using IBS in [11] allows full

resource utilization for uplink transmission while reducing UE energy consumption. An RB schedule with the utility proportional fairness approach in UE is proposed in [13].

While the aforementioned literature focuses on the contributions of CS in different aspects, they never concentrate on the issue of simultaneous consideration of various network traffic types, such as real-time traffic (RT) and elastic traffic (ET). In addition, delay-sensitive multimedia applications, such as VoIP and interactive video, have strict requirements on transport delay, jitter, packet loss, and bandwidth availability. CS with respect to received packets is designed to support QoS for various traffic types and achieves high system throughput [14]. In [15], the authors modeled a problem as an infinite-horizon average cost Markov decision process based on the dynamics of channel state information and queue state information (QSI) to perform adaptive user scheduling to minimize the average delay of a network with heterogeneous base stations.

Therefore, it is essential to prioritize network traffic to minimize congestion risk in the end-to-end service path in order to deliver high-quality voice or video, as well as the preferential treatment required for such critical applications.

3. System model

Consider a CA scheme based on downlink orthogonal frequency division multiplexing, where B eNBs with M_t transmitting antennas communicate with K active UEs equipped with M_r receiving antennas; each UE can transmit or receive on multiple CCs in the LTE-A network, because LTE-A supports up to 8×8 multiple input multiple output (MIMO) in the downlink and up to 4×4 MIMO in uplink [3]. The eNB can employ L CCs (maximum value of L is 5 in LTE-A) at the same frequency band as each CC with an identical bandwidth W to transmit the data. With CA, UE can aggregate multiple CCs simultaneously. Each CC is divided into R resource blocks and each spans M subcarriers in the frequency domain and one frame in the time domain. The number of RBs in a CC is distributed to multiple UEs by its resource scheduler (RS). However, multiple aggregated CCs connected to a single UE increase signal-processing complexity and power consumption. The transmission power at eNB over a single CC_l to communicate with the kth UE, denoted by $P_{l,k}$, is evenly distributed over the whole bandwidth, where l is the number of CC links ($1 \leq l \leq L$). Hence, the instantaneous signal-to-interference-plus-noise ratio of UE k on RB r at CC_l can be expressed [11] as

$$\gamma_{l,k} = \frac{P_{l,k}.a_{l,k}.\|H_{k,l,r}\|^2}{\sum\limits_{i \in B} \sum\limits_{k_i} P_{l,k_i} \beta_{l,k_i} \left\|H_{k,l,r}^{k_i}\right\|^2 + \delta_N^2}, \tag{1}$$

where $a_{l,k}$ denotes the power attenuation due to path loss and shadow fading of UE k with CC_l; $H_{k,l,r}$ represents the independent fading channel gain on RB r at CC_l of UE k; $i \in B$ is the index of the interference link of eNB in the set of neighboring eNBs B; P_{l,k_i} is the transmission power with interference to UE k_i; β_{l,k_i} denotes the path loss attenuation factor from interfering eNB to k_i; and δ_N^2 is the received noise power of additive white Gaussian noise. In this paper, it is assumed that the eNB has perfect knowledge of the channel. Hence, the maximum achievable data rate of the kth UE on RB r at CC_l of the serving eNB is given as:

$$R_{l,k} = W \log_2(1 + \gamma_{l,k}) \tag{2}$$

Each CC has a separate queue for data storage, which is denoted as $S_l, (l = 1, 2, ..., L)$. Assume that the length of the queue at each CC is Q_l and each queue has a finite capacity of packets. $q_l (0 \leq q_l \leq Q_l)$ denotes the

number of packets that are stored in the queue. The value of q_l is different for each CC and depends on packet service, which indicates the status of the queue and is fed back to the FC for slot allocation.

3.1. DSCS scheme

An efficient scheduling scheme must become an active participant in the perfect delivery. The scheduling should be aware of the type of service to ensure that valuable network resources are shared efficiently. The proposed scheme is involved in packet scheduling and CS in order to achieve the required QoS performance for real-time multimedia applications. In this section, the proposed DSCS without priority and with priority is discussed.

3.2. DSCS without priority

The proposed DSCS model without priority is shown in Figure 1. At every TTI, queue state information is fed back to FC and the slot allocation is performed for incoming traffic without considering their priority. The resource blocks are assigned to the packets through RS_l. When $Q_l(t) = 0$, the queue is empty and slot is allotted for the corresponding CC. When $Q_l(t) > q_l(t)$, the queue is nonempty and filled with $q_l(t)$ packets in the queue. When QSI $Q_l(t) \leq q_l(t)$ for a CC indicates that the queue is full, overflow occurs. Hence, the slot is terminated and no allocation is carried further to the corresponding CC_l for a period τ. The flowchart of the slot allocation with respect to QSI is described in Figure 2. Thus, in a slot t, the number of packets transmitted to S_l based on QSI can be written as
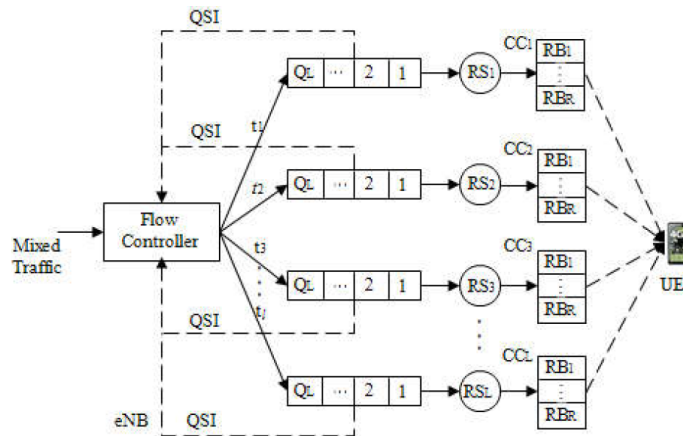

Figure 1. Proposed slot-based carrier scheduling scheme without priority.

$$S_l(t) = \begin{cases} N, for \ Q_l(t) = 0, \quad Q_l(t) > q_l(t) \\ 0, for \ Q_l(t) \leq q_l(t) \end{cases} \qquad \text{where} \quad l = 1, 2, ..., L, \qquad (3)$$

where N is the number of packets required to schedule for a link l and is a constant.

1. Priority-based DSCS scheme

The dynamic CS with priority model has two processing stages. The first stage is slot-based CS with QSI and the second stage is priority-based packet scheduling.

In slot-based scheduling, the slot has been allocated for the corresponding CC_l and the packets are forwarded based on QSI. The next stage is prioritized CS, where the incoming packets are identified based on
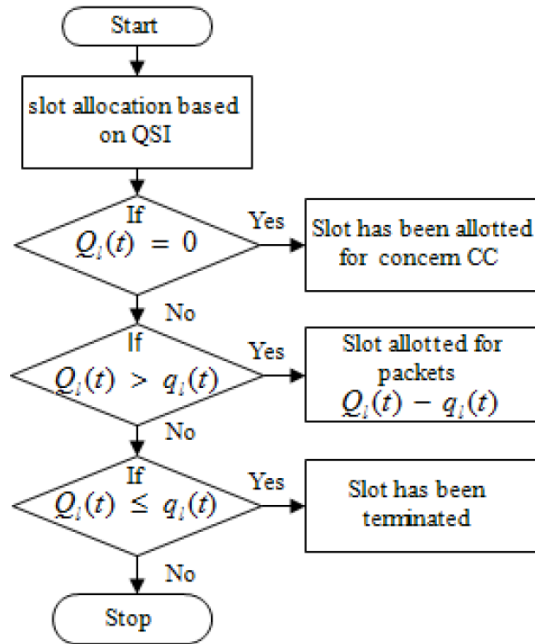
Figure 2. Flow chart for queue scheduler.

their priority, ensuring acceptable delay for each application. The proposed dynamic CS, with different traffic, is illustrated in Figure 3 and Figure 4 illustrates a CS module of a single CC with priority. In this scheduling, separate queues are deployed for multimedia service to manage the heterogeneous traffic. However, each queue can hold homogeneous traffic. In each CC, the arriving packets are classified as primary traffic (PT), RT, and ET by the queue classifier. To obtain the priority level of each service type, the QoS objectives need to be specified. The priority levels assigned for queuing discipline are according to the delay tolerance of different services [16]. The QoS requirement for packet delay is given as

$$D_{\min} \geq D_t \geq D_{\max}, \tag{4}$$

where $D_{\max}$ is the maximum allowable packet delay to achieve the QoS. $D_{\min}$ is the minimum allowable packet delay and D_t is the delay threshold of traffic that depends on the queue waiting time of packet.

The priority levels of packets based on delay tolerance level in a slot t with N transmitted packets is thus classified as

$$N(t) = \begin{cases} N^P, if \ D_{\max} > D_t, \quad D_{\min} > D_t \\ N^R, if \ D_{\max} > D_t, \quad D_{\min} < D_t \\ N^E, if \ D_{\max} < D_t, \quad D_{\min} < D_t \end{cases} \tag{5}$$

The uppermost priority is assigned to the delay-intolerant traffic, PT N^P, which requires almost zero delay. The next highest and lowest priority is assigned to RT and ET packets, respectively. The classified packets are further delivered into the corresponding queues: the RT queue (RQ), which buffers RT packets N^R, and the ET queue (EQ) for elastic traffic N^E.

Subsequently, each RS schedules the appropriate RBs for the packets from the prioritized queue and uniformly transmits over the channel. When PT arrives to a CC currently in the idle state, it starts its transmission and RBs are allocated for PT packets. However, when PT packets arrive at a busy channel currently occupied by another ET/RT service, those services are interrupted. Therefore, the channel has to
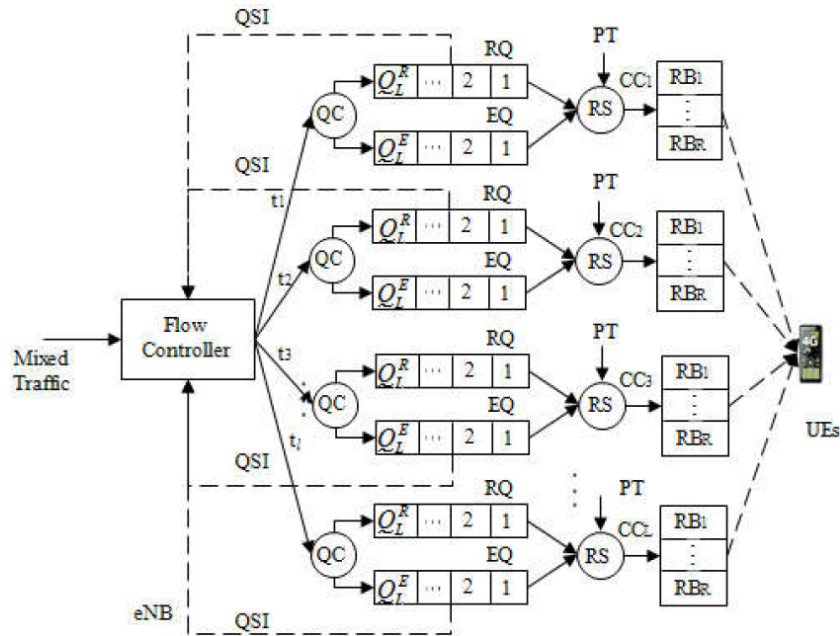
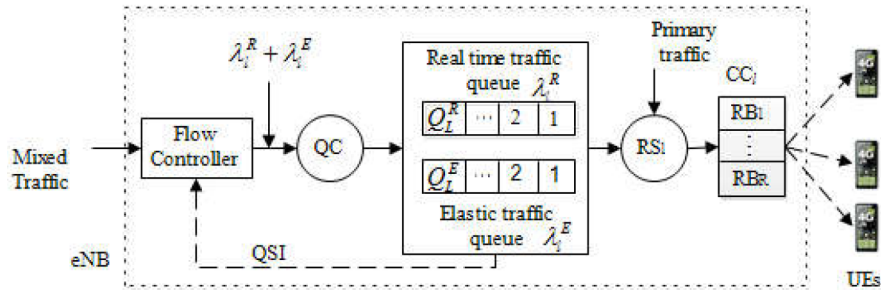**Figure 3.** Proposed slot-based carrier scheduling scheme with priority.

Figure 4. Scheduling module of a single CC.

be vacated immediately and the corresponding CCs will be allocated to the PT packets. The channel access strategy in [11] is utilized for the prioritized queuing system. The CC serves its nonempty queue based on their priority level. A queue of any CC becomes empty and QSI is fed back to FC, which attributes the packets to every TTI. Hence, UE is served through multiple CCs with the aggregation technique. The CCs are sorted according to their QSI and the packets are allocated circularly according to a time slot fashion. Therefore, the proposed DSCS scheme achieves the effective utilization of total capacity of all CCs. The working of priority based scheduling is highlighted in Figure 5.

4. Analysis of the proposed DSCS scheduling scheme

4.1. Scheduler queuing model

Let F denote the set of all the arrival flows at eNB and assume that the time is discretized. The number of packets injected into the queue of CC link l at time slot t is denoted as $A_l(t)$. Assume that $\{A_l(t)\} \in F$ and $A_l(t) = \{A_{l,k}(t), k \in K, l \in L\}$, where $A_{l,k}(t)$ is the arrival process for user k from link l. The number of
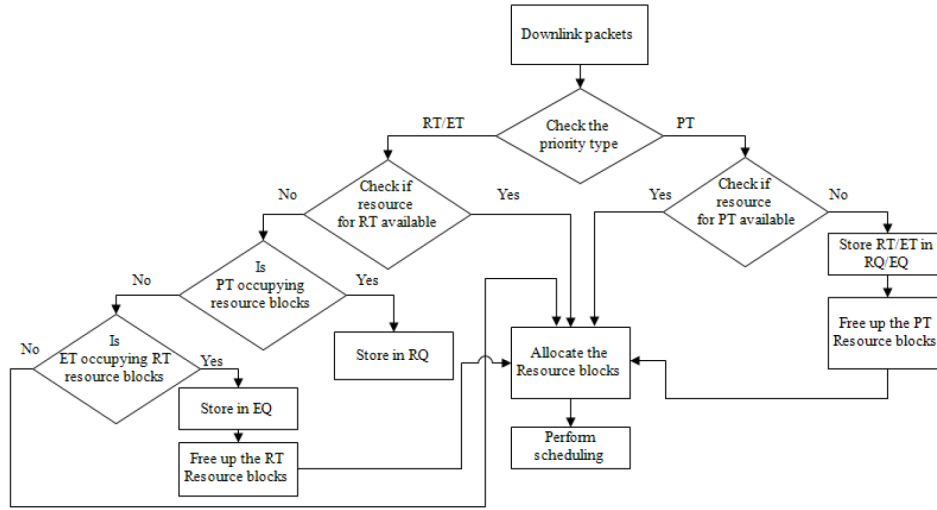
Figure 5. Flow chart for priority-based packet scheduling.

arriving packets in a time slot is uniformly bounded with probability [16] as

$$A_l(t + \tau) - A_l(t) \leq A_{\max}, \forall A_l \in F, t \geq 0, \tag{6}$$

where $A_l(t)$ is the total number of packets that arrived on the link l during the first t time slots, and $A_{\max}$ is a constant. In each time slot, the total departure by the end of the slot t is denoted as $G_l(t) = \mu_{l,k}\tau$. $Q_l(t)$ is the dynamic value of the queue length maintained at CC_l. Since QSI is fed back to the FC, it is inferred that the forwarded packets are successfully transmitted over the link and the queue turns to empty immediately. Thus, the dynamics of queue at time slot t can be described in Eq. (7):

$$Q_l(t) = Q_l(0) + A_l(t) - G_l(t) \tag{7}$$

The FC allocates the packets such that $Q_l(t) \geq 0$ always holds. At each TTI, $Q_l(t)$ is fed back to the FC to make the decision for slot allocation. Therefore, the slot to be allocated through FC with QSI CC_l is written as

$$Q_l(t + \tau) = \min \{[Q_l(t) - G_l(t)] + A_l(t)\} \tag{8}$$

The duration of slot τ for each queue can be expressed as

$$\tau = A_l(t) + G_l(t) \leq Q_l(t) \tag{9}$$

Likewise for the CS, the QSI of $\min\{Q_l^R, Q_l^E\}$ is fed back to the FC to avoid overflow in the queue, where Q_l^R and Q_l^E indicate the queue length of RT service and ET service, respectively.

4.2. Performance measures of scheduling model

Let $\lambda_{l,k}$ denote the average arrival rate to each queue of link l of user k. The traffic load of the network to user k is defined as $\rho_{l,k} = \lambda_{l,k}/\mu_{l,k}$. Let $S(t) = \{S_{l,k}(t), k \in K\}$ denote the size of the packet being transmitted at the tth scheduling slot to UE k. The average packet transmission rate is $\mu_{l,k} = R_{l,k}/S_{l,k}$.

Likewise, the average arrival rates of PT, RT, and ET traffic in each CC link l are $\lambda_{l,k}^P$, $\lambda_{l,k}^R$, and $\lambda_{l,k}^E$, respectively, and the corresponding service rate per channel for each traffic is denoted as $\mu_{l,k}^P$, $\mu_{l,k}^R$, and $\mu_{l,k}^E$, respectively.

Using Little's law, the end-to-end delay of RT services D_l^R is obtained [17] as

$$D_l^R(t) = T_l^{R(q)}(t) + T_l^{R(s)}(t), \tag{10}$$

where $T_l^{R(q)}$ is the waiting time of RT packets in the queue and $T_l^{R(s)}$ denotes the transmission time of RT packets progressing in the system. Therefore,

$$D_l^R(t) = \frac{Q_l^R(t)}{\lambda_l^R} + \frac{N_l^{R(s)}(t)}{\lambda_l^R}, \tag{11}$$

where the number of ongoing RT packets in the system is denoted as $N_l^{R(s)}$. Similarly, the end-to-end delay of ET packets can be expressed as

$$D_l^E(t) = T_l^{E(q)}(t) + T_l^{E(s)}(t) = \frac{Q_l^E(t)}{\lambda_l^E} + \frac{N_l^{E(s)}(t)}{\lambda_l^E}, \tag{12}$$

where $T_l^{E(q)}$ is the waiting time of ET packets in the queue. $T_l^{E(s)}$ denotes the transmission time of progressing ET packets in the system and $N_l^{E(s)}$ is the number of ongoing ET packets in the system.

From Eq. (11), the average end-to-end delay for RT services $\bar{D}_k^R$ of user k is given by

$$\bar{D}_k^R = \sum_{t=0}^{T-1} E\left\{D_{l,k}^R\right\} \tag{13}$$

Likewise, from Eq. (12), the average end-to-end delay for ET services $\bar{D}_k^E$ of user k is given by

$$\bar{D}_k^E = \sum_{t=0}^{T-1} E\left\{D_{l,k}^E\right\} \tag{14}$$

The average sojourn time is a total period a packet spends from its arrival to complete departure. It is defined as

$$E_{l,k}\{E_{a,l}[T_{a,k}]\}, \tag{15}$$

where $T_{a,k}$ is the sojourn time of arriving packets (a,k) in CC l to user k.

The spectrum utilization of the system is the average number of utilized channels over the total number of channels [17]. In $A_l \in F$, a total number of $w(l)$ channels out of W are utilized, where T is the time taken for user k. The spectrum utilization for the downlink transmission of UE k can be expressed as

$$U_k = \sum_{a \in F} \frac{w(l)}{W.T} \tag{16}$$

5. Simulation and discussions

The performance comparison of the proposed scheme for various traffic loads with and without priority is presented in this section. To perform the simulation, a single cell scenario with a fixed eNB at the center of the cell is assumed. The users are moving in a vehicular propagation environment in a bounded region of radius equal to 1 km with a speed of interval 3 km/h. The user arrivals are assumed with the Poisson process and they are randomly distributed in various distances from the serving eNB. There are 20 users for each traffic type in the cellular network with uniform distribution.

VoIP is considered PT, video streaming is considered RT, and file transfer protocol applications are considered ET. An on/off Markov model is considered for the voice traffic (PT). The on and off periods are exponentially distributed with an average value of 300 ms each. During the off periods, there are no voice packets generated. During the on periods, voice packets of each one-way voice flow are generated at a rate of 32 kb/s with a packet size of 160 bytes. Trace-based applications are used as video flows (RT). They send packets based on realistic video trace files (H.264). For elastic traffic, the greedy CBR model is assumed. There are ET packets available in the outgoing interface queue. Each ET packet has a length of 1000 bytes.

In this simulation, the spectral efficiency, delay, and throughput of the proposed scheme are investigated. The simulation parameters are based on LTE specifications and presented in the Table.

Table. Simulation parameters.

Parameters	Values
Cellular layout	3-sectorized hexagonal grid
Minimum distance between UE and cell	35 m
Number of CCs	4
Bandwidth of each CC	10 MHz
Number of RBs per CC	50
Carrier frequency (fc)	2 GHz
Total eNB transmission power	40 dBm
Thermal noise spectral density	–174 dBm/Hz
Fast fading model	Rayleigh
UE speed of interval	3 km/h
Antenna pattern	4×4
Noise figure of UE	9 dB
TTI	1 ms

From Figure 6, it is observed that spectrum utilization is improved in the proposed prioritized dynamic CS strategy. When the system is operated without a queue, the spectrum utilization is lower in contrast with the system with prioritized queues. With higher $\rho_{l,k}$, the improvement of spectrum utilization is more significant. In the proposed scheme with priority, as the separate queues are configured for low and high priority for different traffic types, more packets can be accommodated in the queue, which allows more user requests; hence the average number of CCs utilized over the available number of CCs is improved. The spectrum utilization of the proposed scheme is increased by 20% compared with the scheme that configured without priority queue, and is 60% higher than the DAC-IBS scheme. In the case of the DAC-IBS scheme, all the CCs are utilized to serve the flows only with nonempty queues. In order to improve spectrum utilization, it is required to reduce the duration for which the channel stays idle. When a channel becomes idle, it is more likely that the other RT/ET packets will use it, which enhances the utilization of RBs effectively. It is noted that the prioritized queue can fulfil this

requirement; therefore, more user packets are accommodated. The reason for nonzero spectrum utilization at $\rho_{l,k} = 0$ is because of the baseline utilization due to user services.

Figures 7–9 show the average sojourn time under different traffic loads. The average sojourn time varies for different types of traffic and schemes, due to the introduction of a priority-based queue. In Figure 7, as the primary traffic arrives, other services are interrupted and RS allocates RBs to PT packets and ensures almost zero delay; hence the minimum average sojourn time is attained over other kinds of traffic. In the case of ET due to the least priority, which encounters more interruptions from PT/RT services, the interrupted elastic packets are kept in EQ. Hence, it achieves a larger sojourn time than other traffic types because of the high processing delay (Figure 9). However, the performance gain will not be affected for ET, which can tolerate a high delay in transmission.
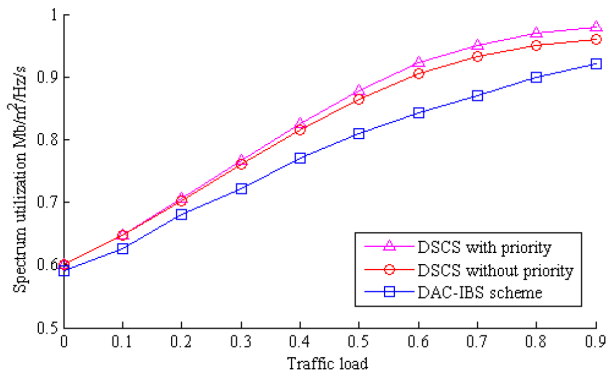

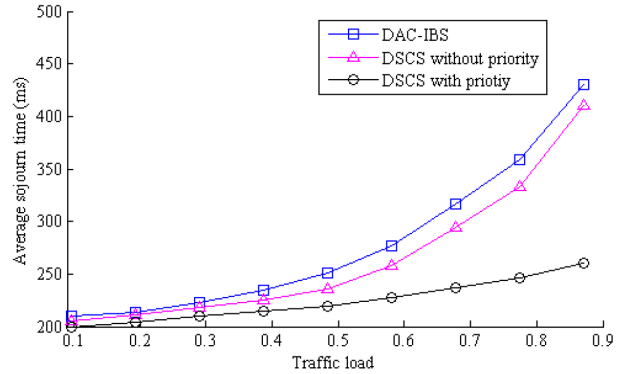**Figure 6.** Spectrum utilization comparison of proposed scheme.

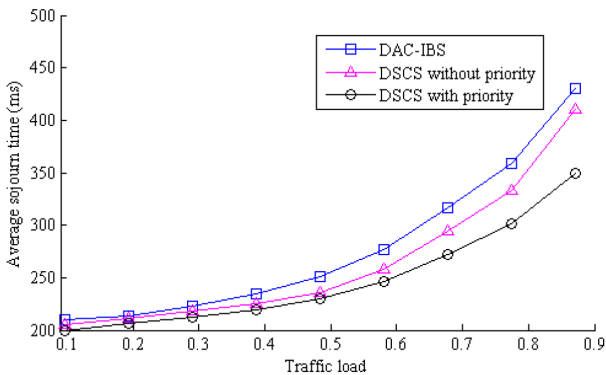**Figure 7.** Average sojourn time comparison of primary traffic.

Figure 8. Average sojourn time comparison of real-time traffic.
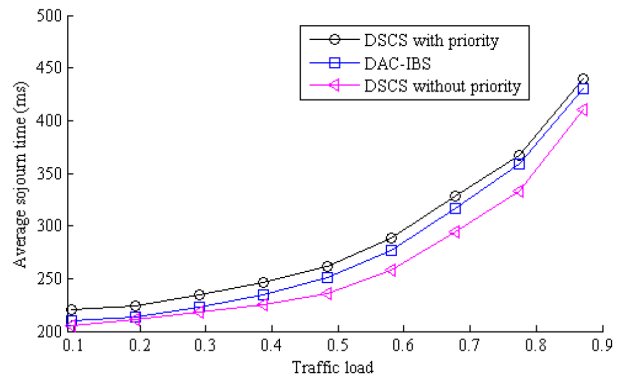

Figure 9. Average sojourn time comparison of elastic traffic.

The average delay performance of the primary user traffic over different traffic loads is compared with that of other schemes in Figure 10. The average packet delay of all the services increases as the traffic load increases. Compared with other schemes, the proposed scheme with priority provides less delay for delay-sensitive traffic PT than for other kinds of traffic on the networks. Since PT packets gain uppermost priority, RBs are allocated immediately to traffic and hence provide service with almost minimal delay. When PT packets arrive at a busy channel currently occupied by an RT/ET user with assembled CC, their service is interrupted and the channel has to vacate immediately; thus, carriers are offered to PT user. After PT departure, the CC becomes idle, and the opportunity to access those RBs from the corresponding CC will be shared between aRT and ET services

waiting in the queue by utilizing the priority. At $\rho_{l,k} = 0.9$, the proposed DSCS scheme with priority achieves 20% less delay than the scheme without priority and 40% less delay than DAC-IBS scheme for primary traffic.

From Figure 11, it can be found that the average delay for RT users with priority is better than that of other schemes. Since the proposed priority-based scheme offers the available carriers immediately to PT users based on priority, RT packets need to wait in RQ until the PT packets are processed completely. When an idle CC appears, the priority of accessing those RBs is assigned to the service waiting in RQ. Hence, there will be a delay for RT packets to transfer. Similarly, in Figure 12, the average delay for ET is larger than in other schemes and due to ET's high delay tolerance the service is interrupted whenever PT/RT service arrives, hence causing an increase in average delay.
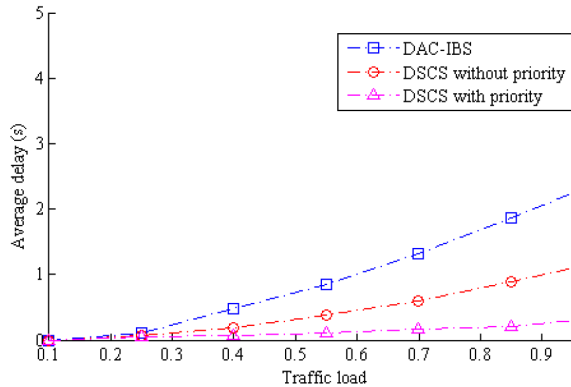

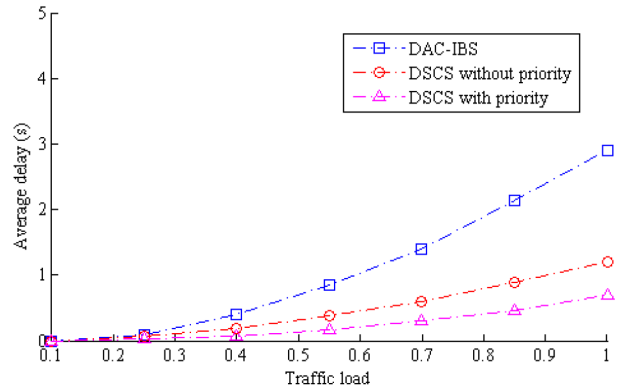**Figure 10.** Average delay performance comparison of primary traffic.

Figure 11. Average delay performance comparison of real-time traffic.

Figure 13 presents the system throughput achieved according to the number of users of different scheduling schemes. It is noted that system throughput increases with the number of users because of the rise in the allocated number of RBs. For the proposed DSCS with a priority scheme, throughput increases and it is higher than in DSCS without priority and DAC-IBS scheme. This performance, in terms of the throughput of the system, is explained by the adaptation of the proposed scheduling scheme based on priority by exploiting resources according to the various network traffic types. In addition, it considers the status of the queue when making scheduling decisions and classifying the users based on traffic characteristics, providing services accordingly for achieving high throughput.
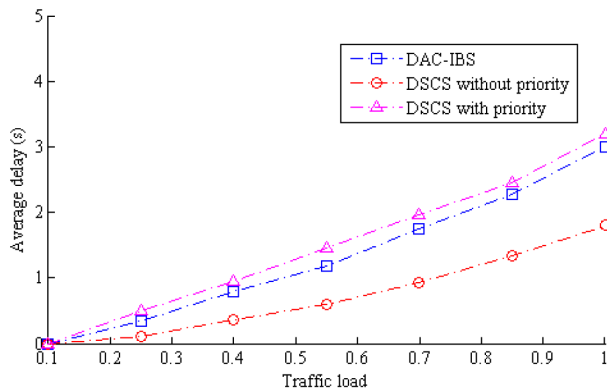

Figure 12. Average delay performance comparison of elastic traffic.
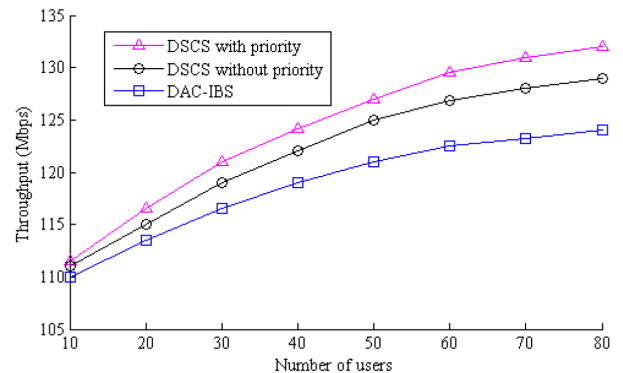
Figure 13. Throughput comparison of different scheduling schemes.

Figure 14 shows a typical realization of throughput versus time t of the proposed DSCS priority-based scheme for different mobile traffic. It is observed that the throughput of primary traffic outperforms other multimedia traffic. Due to being highest priority, the other services are interrupted and the available RBs are assigned to PT packets. In addition, increasing the traffic load leads to increasing contention over the limited RBs assigned to the incoming packets, which affects their throughput with respect to time. Moreover, with least priority, ET services may forcibly terminate because of the high queuing delay, which reduces the throughput of the elastic user in the case of prioritized scheduling.
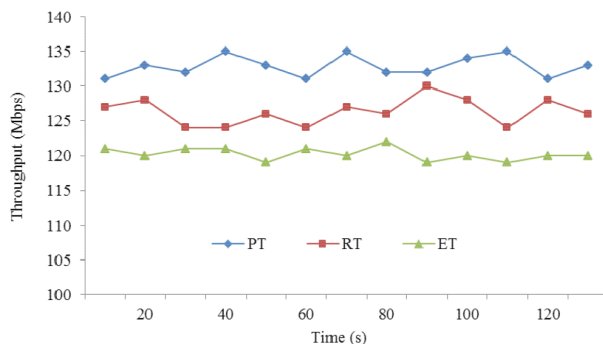

Figure 14. Throughput comparison of prioritized DSCS with various traffic types.

6. Conclusion

In this paper, a DSCS strategy is proposed to improve the spectrum utilization and QoS of mobile multimedia traffic over downlink LTE-A networks with CA. The prioritized scheduling of different traffic types has been achieved by introducing a queuing technique and utilizing QSI for better slot allocation. The proposed scheme incorporates the priority of mobile traffic that efficiently and dynamically allocates radio resources to users, enabling LTE-A systems to achieve an improvement in real-time service provisioning.

Simulation results have shown that the proposed DSCS scheduling scheme outperforms existing schemes in terms of spectrum utilization, throughput, and average delay, and also achieves the guaranteed QoS requirements of delay-critical applications like voice and video streaming, and extends support to other network traffic with various delay requirements.

References

[1] Biswas J, Liu RP, Ni W, Collings IB, Jha SK. Joint channel and delay aware user scheduling for multiuser MIMO system over LTE-A network. In: Proceedings of the 21st International Symposium on Quality of Service (IWQoS); 2013; IEEE/ACM, pp. 1-8.

[2] Park CS, Sundstrom L, Wallen A, Khayrallah A. Carrier aggregation for LTE-advanced: design challenges of terminals. IEEE Commun Mag 2013; 51: 76-84.

[3] Akyildiz IF, Gutierrez-Estevez DM, Reyes EC. The evolution to 4G cellular systems: LTE-Advanced. Physical Commun 2010; 3: 217-244.

[4] Shen Z, Papasakellariou A, Montojo J, Gerstenberger D, Xu F. Overview of 3GPP LTE-Advanced carrier aggregation for 4G wireless communications. IEEE Commun Mag 2012; 50:122-130.

[5] Pedersen K, Frederiksen F, Rosa C, Nguyen H, Garcia L, Wang Y. Carrier aggregation for LTE-advanced: functionality and performance aspects. IEEE Commun Mag 2011; 49: 89-95.

[6] Chung YL, Tsai Z. A quantized water-filling packet scheduling scheme for downlink transmissions in LTE-advanced systems with carrier aggregation. In: Proceedings of the International Conference on Software, Telecommunications and Computer Network (SoftCOM); 2010; pp. 275-279.

[7] Wang H, Rosa C, Pedersen KI. Performance analysis of downlink inter-band carrier aggregation in LTE–advanced. In: Proceedings of the Vehicular Technology Conference (VTC Fall); 2011; IEEE. pp. 1-5.

[8] Wang Y, Pedersen K, Mogensen P, Sorensen T. Carrier load balancing methods with bursty traffic for LTE-advanced systems. In: Proceedings of the 20th International Symposium on Personal, Indoor and Mobile Radio Communication; 2009; IEEE. pp. 22-26.

[9] Zhang L, Zheng K, Wang W, Huang L. Performance analysis on carrier scheduling schemes in the long-term evolution-advanced system with carrier aggregation. IET Commun 2011; 5: 612-619.

[10] Zheng K, Liu F, Xiang W, Xin X. Dynamic downlink aggregation carrier scheduling scheme for wireless networks. IET Commun 2014; 8: 114-123.

[11] Liu F, Zheng K, Xiang W, Zhao H. Design and performance analysis of an energy-efficient uplink carrier aggregation scheme. IEEE J Sel Area Comm 2014; 32: 197-207.

[12] Chung YL, Jang LJ, Tsai Z. An efficient downlink packet scheduling algorithm in LTE-advanced systems with carrier aggregation. In: Proceedings of the 2011 IEEE Consumer Communications Network Conference (CCNC); 2011; IEEE. pp. 632-636.

[13] Shajaiah H, Abdelhadi A, Clancy TC. Towards an application-aware resource scheduling with carrier aggregation in cellular systems. IEEE Commun Letters 2016; 20: 129-132.

[14] Wu K, Liao W. On service differentiation for multimedia traffic in multi-hop wireless networks. IEEE T Wirel Commun 2009; 8: 2464-2472.

[15] Cheng A, Li J, Yu Y, Jin H. Delay-sensitive user scheduling and power control in heterogeneous networks. IET Networks 2015; 4: 175-184.

[16] Li Q, Negi R. Maximal scheduling in wireless networks with priorities. IEEE T Wirel Commun 2012; 11: 3701-3713.

[17] Balapuwaduge I, Jiao L, Pla V, Li FY. Channel assembling with priority-based queues in cognitive radio networks: strategies and performance evaluation. IEEE T Wirel Commun 2014; 13: 630-645.