Available online at www.sciencedirect.com
ScienceDirect
Procedia Computer Science 59 (2015) 115 – 122
Procedia Computer Science

International Conference on Computer Science and Computational Intelligence (ICCSCI 2015)

Throughput-aware Resource Allocation for QoS Classes in LTE Networks

Nasim Ferdosian[a,*], Mohamed Othman[a], Borhanuddin Mohd Ali[b], Kweh Yeah Lun[a]

[a]*Department of Communication Technology and Network, Universiti Putra Malaysia, Serdang, Selangor D.E 43400, Malaysia*
[b]*Department of Computer and Communication Systems, Universiti Putra Malaysia, Serdang, Selangor D.E 43400, Malaysia*

Abstract

In LTE systems, multicast services must be delivered efficiently in response to the need for strong QoS support. However, each class of quality services has its own requirements to be satisfied. These quality constraints limit the scheduling flexibility, and the LTE downlink resource allocating algorithms need to assimilate these constraints while trying to maximize system performance in terms of fairness and throughput. This paper addresses this fundamental problem of LTE downlink scheduling by adopting the time-domain Knapsack algorithm over the traffic overload patterns. We fine tune the Knapsack algorithm, to overcome this problem and improve system performance objectives. We demonstrate that more efficient performance can be achieved in terms of fairness index and system throughput, which is evaluated using simulation results.

Peer-review under responsibility of organizing committee of the International Conference on Computer Science and Computational Intelligence (ICCSCI 2015)

Keywords: Long Term Evolution, downlink scheduling, Quality of Service, Knapsack algorithm, fairness

1. Introduction

The Long Term Evolution (LTE) cellular communication system has emerged as a fast-growing prevalent technology, delivering a diversity of mobile broadband services, in the communication market. The LTE specifications have been standardized to utilize Orthogonal Frequency Division Multiple Access (OFDMA) as the transmission scheme, commissioned to carry out the downlink communication[1]. The OFDMA transmission scheme in comparing with the old one (Code Division Multiple Access) provides a key advantage of flexibility for resource allocation decision makers in exploiting frequency diversity[2].

An LTE scheduler is expected to allocate radio resources efficiently to support a high variety of services and maximize system throughput. However, it is a crucial problem to accomplish all targets

* Corresponding author. Tel.: +603 8947 1707; Fax: +603 8946 6576
E-mail addresses: n.ferdosian@gmail.com (Nasim Ferdosian)., mothman@upm.edu.my (Mohamed Othman).

Peer-review under responsibility of organizing committee of the International Conference on Computer Science and Computational Intelligence (ICCSCI 2015)
doi:10.1016/j.procs.2015.07.344

at the same time. Each factor can be supplied at the cost of reducing another one[3]. For example, in response to the strong QoS support, many QoS-aware solutions were introduced which have been revealed to be unsuitable for dealing with throughput and fairness requirements[4].

A flexible QoS-oriented scheduler, divided into Time Domain (TD) and Frequency Domain (FD), was introduced in[5] for real-time video traffic. The proposed algorithm considers arrival rate and head of line packet delay as influential QoS factors for multiuser resource distribution. The authors in[6] described a self-optimization method to the LTE network scheduler in response to the active changes of network conditions and traffic over time and proposed an optimized-service aware (OSA) scheduler. To simplify the complexity of the resource allocation procedure, it has been partitioned into three separate stages: QoS classes identified classification, time domain and frequency domain scheduling. At the first step each bearer is classified into individual QoS class based on its CQI factors. Then the TD scheduler prioritizes the classified bearers according to their QoS data rate requirements and categorizes them into separate prioritized candidate bearers: GBR and Non-GBR. GBR bearers typically carry real-time applications which are sensitive to delay and need to be served with a guaranteed bitrate[7]. OSA algorithm sorts each GBR bearer according to the Head of Line (HOL) packet delay in the buffer of the related bearer.

The ranking function of traditional scheduling algorithms which are only based on the queue's priority, ignoring other metrics, would impose a lack of sufficient intellect over the resource allocation process[8]. In response to this challenging problem, the authors in[9] introduced Knapsack scheduling algorithm with emphasis on overload states. This class-based resource allocation algorithm supports QoS constraints by ordering the bearers using a ranking function calculated based on the multiple metrics, including GBR/Non-GBR class priority, bearer queue status, packet loss and delay. However, since the main volume of the LTE network traffic is real-time services, growing in an explosive manner, especially video and VoIP[10], the fairness issue among these services forms a major challenge as well as QoS support in current networks.

In this work, we propose an opportunistic approach to treat all three desired LTE targets aggregately as a single scheduling problem. Our proposed strategy can be viewed as a theoretical complementary of the work[9]. We formulate the optimization problem of resource apportion and make the ranking function to implement these optimal policies efficiently. The performance of the scheme is evaluated by comparing with Knapsack and Priority only algorithm as reference scheduling schemes. The inter-class and intra-class fairness assessment is done in terms of throughput[11] for VoIP traffic, and the impact of throughput-aware ranking function is evaluated with the help of simulation results.

2. System Model

In the present paper, OFDMA based 3GPP-LTE Downlink system with an E-UTRAN NodeB (eNB) (the base station in LTe networks) is considered. Each OFDMA frame of the LTE radio channel is constructed in the time and frequency domains. It contains ten 1ms sub-frames in time domain and a sub-channel of 12 consecutive same size sub-carriers in frequency domain. A single sub-carrier covers 15 kHz and the sub-channel subsequently 180 kHz of the spectrum. The basic resource unit for mapping the radio resources to active users is named Resource Block (RB). Each RB spans over a 0.5 ms time extent and one sub-channel[12,13]. The resource scheduling process is performed every Transfer Time Interval (TTI) which lasts one ms. We assume 5MHz bandwidth, consisting of 25 RBs in time-frequency domain and 330 active users in a single cell scenario. Here, the set of all users is defined by U ($U = \{1, 2, ..., u\}$) and the set of all RBs by RB ($RB = \{1, 2, ..., rb\}$). U is a collection of the users with different kind and number of bearers (single VoIP, single data, multiple data and VoIP-data) over the mixed-interval-time of normal and overload status. In each overload time interval, 50 users with single data session are added to the existed traffic and they are eliminated at the start of normal status.

Similarly to the algorithms reviewed in[14] it is supposed that the resource scheduling procedure is decoupled between time domain scheduler and the frequency domain scheduler separately, in the way that time domain scheduler selects an optimal set of bearers to be served and the required number of RBs for each bearer. Then in the next phase, frequency domain scheduler defines the most suitable RBs to be assigned to each bearer from the selected set of the prioritized bearers [15]. This separation can make it easier to optimize each step of scheduling independently and result in significant reduction of computational complexity [16]. The focus of this paper is the phase in which the time domain packet scheduling is involved.

The well-known Maximum Throughput (MT) scheme [17] is a well devised algorithm to maximize total cell throughput by maximizing the objective function $\sum d_i(t)$, where $d_i(t)$ is the achievable data rate of user i at time slot t. It selects the user that can achieve the biggest amount of throughput over a given RB in the current TTI. The MT approach relies on the instantaneous channel conditions of each user to measure the expected data rate of each user and accordingly maximize system overall throughput. However, this approach is insufficient to support multimedia applications with strict loss and delay requirements. Therefore, several resource allocation algorithms that satisfy requirements of these new applications have been developed [18,19,20].

Here, the scheduling architecture works on a service bearer level where, a bearer is considered to convey a service data flow. The bearers have their own service characteristics, depend on the application coming from, and are grouped into two divergent Guaranteed Bit Rate (GBR) and Non-Guaranteed Bit Rate (Non-GBR) groups. The 3GPP has standardized the corresponding service characteristics of the bearers into 9 Quality Channel Indicators (QCI) to ensure that the bearers belong to the same class of services, receive the same minimum level of service quality [21]. The standardized QCI classes along with their corresponding performance characteristics has been shown in Table 1.

Table 1. Standardized QoS characteristics in QCI classes [21]

QCI	Bearer type	Priority	Packet delay budget (ms)	Packet error loss rate	Example services
1	GBR	2	100	10	Conversational Voice
2		4	150	10	Conversational Video (Live Streaming)
3		3	50	10	Real Time Gaming
4		5	300	10	Non-Conversational Video (Buffered Streaming)
5	Non-GBR	1	100	10	IMS Signalling
6		6	300	10	Video (buffered streaming) TCP-based (e.g., www, e-mail, chat, ftp, p2p file sharing, progressive video, etc.)
7		7	100	10	Voice, Video (Live Streaming) Interactive Gaming
8		8	300	10	Video (buffered streaming) TCP-based (e.g., www, e-mail, chat, ftp, p2p file sharing, progressive video, etc.)
9		9	300	10	

Every QCI has a particular packet forwarding treatment according to their predefined specifics; therefore, all bearers assigned to an especial QCI must follow a common rate and scheduling policy while fulfilling the QoS class-based constraints. To this end, the Knapsack scheduling algorithm has been proposed as an opportunistic scheduling. It contains two main phases of resource allocation. In the first phase, the GBR bearers are assigned resources to allocate and ensure their guaranteed bitrate. In the next phase, the not used up resources are distributed among Non-GBR bearers and the queued packets of the GBR bearers to get resources further up the already provided guaranteed bitrate. The second step list of candidate bearers is sorted according to their rank value. Then, a subset of the highest ranked bearers is selected. It is expected to select the optimal set of the bearers. This selected subset is considered as an optimal solution to the knapsack problem of resource scheduling, expecting to maximize the sum of the values of the scheduled bearers without exceeding the total capacity of the system resources. The knapsack algorithm supports QoS class-based application constraints. However, the optimal scheduling policies need to be optimal under the fairness and maximum performance throughput requirements as well.

3. Problem Formulation

To comply these three targets, we will describe an alternative QoS class-based scheduling problem that would insure that all bearers get at least a certain fraction of the total system performance. This problem is formulated as a multi-objective optimization problem which can be stated by the following expressions.

Objective functions:

$$\max \sum_{i=1}^{n} \sum_{rb \in RB_i} r_{i,rb}, \tag{1}$$

$$\forall i \in N : \liminf_{t \to \infty} r_i(t) \geq \phi_i \bar{r}, \tag{2}$$

and

$$\forall i \in N : \min(l_i) \; and \; \min(d_i), \tag{3}$$

subject to:

$$\forall i, j \in N, i \neq j : RB_i \cap RB_j = \emptyset, \tag{4}$$

$$\forall i \in N, k \in K : d_{i_k} < D_k, \tag{5}$$

$$\forall i \in N, k \in K : l_{i_k} < L_k, \tag{6}$$

where i ($i \in N = \{1, 2, ..., n\}$) denotes the index of the bearer which is assigned to the QCI class k ($k = \{1, 2, ..., 9\}$) transmitting data over the RB_i resource blocks. RB_i is the set of resource blocks dedicated to the bearer i and is presented by the set $RB_i \subset RB = \{1, 2, ..., rb\}$. Let $r_{i,rb}$ be the achieved throughput by bearer i over the rbth resource block, ϕ_i is the minimum fraction of the total average throughput, required by bearer i with $\phi_i \geq 0$ and $\sum_{i=1}^{n} \phi_i \leq 1$, $r_i(t)$ denotes the average throughput of bearer i up to time t and $\bar{r}$ is the average total throughput, D_k and L_k are maximal packet delay target and maximal packet loss rate target from the corresponding QCI class k respectively, d_{i_k} and l_{i_k} are the measured packet delay and loss for bearer i from QCI class k respectively. Here, ϕ_i indicates the fairness target in terms of throughput.

In this context, the optimization objectives, including maximizing system throughput, fairness, minimizing delay and packet loss rate as respectively expressed by Equations (1),(2) and (3) should be achieved. The sets of all the assigned resource blocks are disjoint as indicated in Equation (4), implying that no resource block can be granted to more than one bearer. Delay and packet loss rate must be less than the standardized packet delay budget and packet error loss rate thresholds of the corresponding QCI class (Equations 5 and 6).

4. Proposed Scheme

In this context, a utility function is devised to mathematically compute the bearer rank value based on the influential parameters of the desired performance targets, allowing the resource decision makers to put the emphasis on the specific factors of the network performance. The main idea of the proposed ranking function was inspired by the normalized ranking function in[9] which is a combination of four individual ranking functions of QoS metrics (delay, loss, queue depth and priority). Each ranking function outputs a weighted rank value, bounded in $[0, p_i]$, and is calculated as follows:

$$rf(x_i, p_i) = p_i . \tanh(x_i), \tag{7}$$

where p_i is the adjustable weight for each QoS metric assigned by the operator and x_i is the normalized value of QoS metric i calculated as follows:

$$x_i = \frac{measured\ value\ of\ metric\ i}{QoS\ \ constraint\ of\ metric\ i}. \tag{8}$$

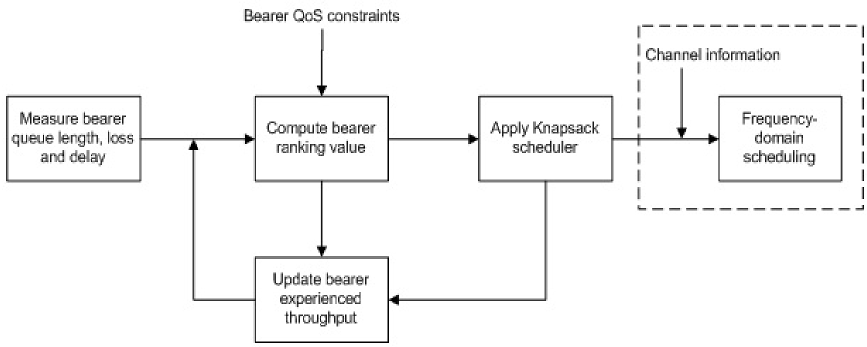

Fig. 1. Block diagram of the proposed scheduling scheme

In the normalized ranking function the QCI label, indicating the QoS constraints, dedicated to bearers, is the main factor that determines the transmission priority of a particular bearer. However, providing fairness and high throughput performance is a challenging issue in case of scheduling strategy unaware of the experienced data rate. We overcome this conflicting issue by considering a measure of throughput normalized by the quantity of past data rate experienced by each user; making this, it is possible to average the resources evenly between the users and consequently providing fairness along with higher overall throughput. Consequently, the overall rank for a given bearer will be calculated as follows:

$$\sum rf(x_i, p_i), \ \forall i \in \{delay, loss, queuedepth, priority, throughput\}, \tag{9}$$

where specifically, in case of throughput parameter, the Time Domain Proportional Fair metric will be used as the normalized throughput value. This metric is expressed as:

$$x_{throughput} = \frac{EstimatedThroughput}{PastAverageThroughput}. \tag{10}$$

The supportable throughput is estimated by the link adaptation, utilizing CQI value. The past average throughput which is the data rate history of each user. In practice, the distribution of throughput is unknown, and hence we need to keep the throughput experienced by the involved users to control its distribution. Every TTI the experienced throughput will be updated when the bearers are resource allocated. Block diagram of the proposed scheduling policy is depicted in 1.

5. Results and Discussion

In previous section, a throughput-aware resource allocation strategy was proposed, resulting in a fair service coverage offering for different quality classes of traffic. To perform our resource allocation model we applied the same scenario as defined in[9]. The reference scenario is a single-cell multicast system where resource allocation is performed at each TTI in time domain, each one lasting 1 ms. The fair resource division performance of the proposed scheduling solution is evaluated in two aspects of intra-class fairness and inter-class fairness under the predetermined normal and overload traffic conditions.

The average obtained throughput by individual QCI classes are used as inter-class fairness measure by different scheduling strategies. Table 2 shows the average throughput of different QCI classes for the proposed scheme in compare with priority only scheduler and knapsack algorithm along with the improved percentage. Distinct QoS demands of the bearers coming from different QCI classes results in disparate amount of the received throughput for these classes. The comparison of the average

Table 2. Average Throughput (Mbps) per QCI class

Scheduler	QCI1	QCI2	QCI3	QCI4	QCI6	QCI7	QCI8	QCI9
Throughput-aware	4.93	3.09	4.88	2.94	1.32	1.49	1.50	1.31
Knapsack	5.42	2.51	5.49	2 .36	1.54	1.44	1.52	1.19
Improved percentage(%)	-9.04	23.11	-11.11	24.58	-14.28	3.47	-1.31	10.08
Priority	5.44	2.41	5.02	2.73	1.83	1.51	1.96	0.65
Improved percentage(%)	-9.37	28.21	-2.79	7.69	-27.87	-1.32	-23.47	101.54

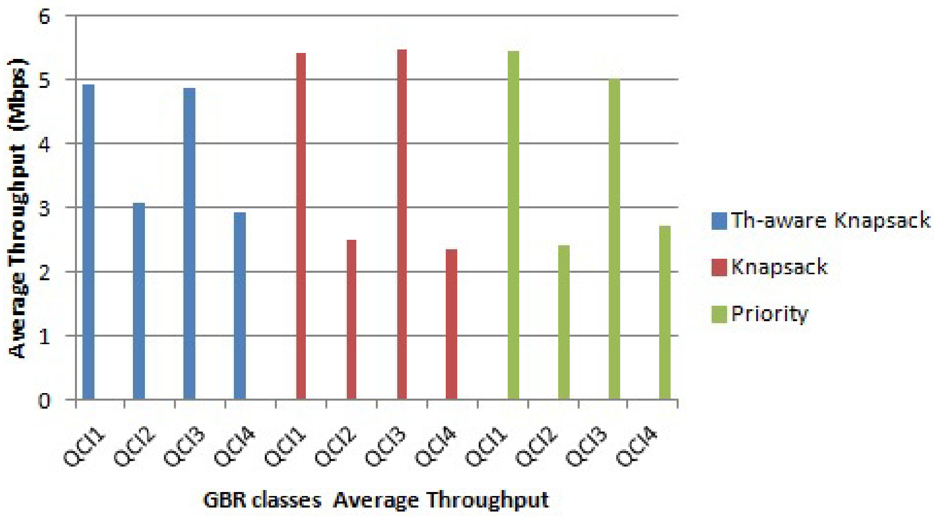

Fig. 2. GBR inter-class fairness versus schedulers

throughput for GBR and Non-GBR groups of bearer is depicted separately in Figs. 2 and 3. As can be seen in these figures the proposed scheme tried to make the level of satisfaction among individual users in terms of average throughput almost near to each other.

As much as, the effect of past experienced throughput awareness is more prominent for prioritizing the bearers with the same QoS characteristics, the measure of fairness for the VoIP traffic is important in assessing how fair the system scheduling solution performs in aspect of intra-class fairness. The intra-class fairness is evaluated by determining the average of the per user throughput in long term of simulation time. These simulation results are shown in Fig 4. As can be seen from this figure, the references strategies improve the performance of high priority bearers by sacrificing the transmission performance of the other bearers results in several fluctuations in graph. Notice that the downfalls of the graph implies the starvation of VoIP bearers by using the reference resource allocation schemes. Hence, the smaller and less number of throughput fluctuations indicate the higher level of fairness among the VoIP bearers by applying the throughput-aware knapsack algorithm.

6. Conclusion

This study paper has focused on downlink resource scheduling with QoS and fairness constraints for different quality classes in LTE networks. We describes the resource allocation as a multi-objective optimization problem, covering three main performance targets of LTE scheduling. The desired solution

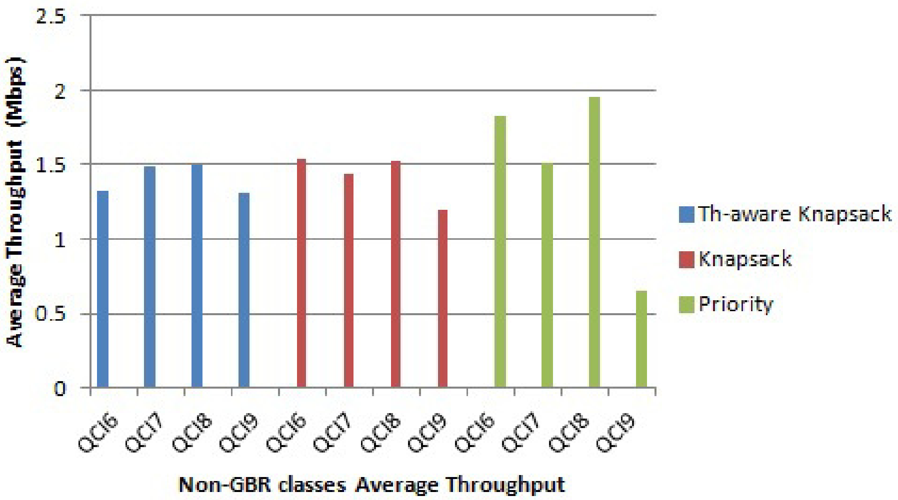

Fig. 3. Non-GBR inter-class fairness versus schedulers

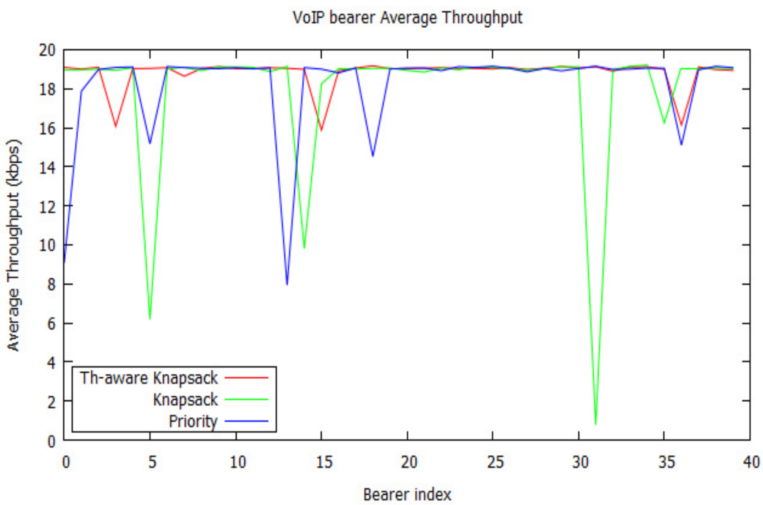

Fig. 4. Intra-class fairness (VoIP bearers)

which is selecting and scheduling the best candidate bearers was provided by using a throughput-aware knapsack algorithm to maximize the desired performance targets. with respect to average throughput measurements we can conclude that by modifying the Knapsack algorithm to use a throughput-aware ranking function the system performance in terms of total throughput can be enhanced in several classes of QCI to the close levels. On the other hand, as much as the VoIP traffic is the major volume of the existed wireless communication traffic, we assessed the proposed scheduling solution by comparing the level of the long-term obtained throughput by the VoIP bearers. The simulation results showed that the level of fairness was improved as the particular effect of added normalized throughput metric. The future work can focus in using other optimal solutions to provide a fair resource distribution in uplink system.

Acknowledgment

The authors are thankful to Dr. Michael Brehm and Prof. Dr. Ravi Prakash for their valuable contribution from the University of Texas at Dallas. Also, this work has been supported by the Malaysian Ministry of Education under the Fundamental Research Grant Scheme FRGS/2/2014 /ICT03/UPM/02/3.

References

1. Dahlman, E., Parkvall, S., Skold, J. 4G: LTE/LTE-advanced for mobile broadband. Academic press. 2013.
2. Ali-Yahiya, T. Understanding LTE and its Performance. Springer Science and Business Media, 2011.
3. Ali-Yahiya, T., Alagha, K. Downlink fairness-aware adaptive resource allocation approach for LTE networks. International Journal of Network Management. 2011; 21(4): 269-283.
4. Capozzi F., Piro G., Grieco L. A., Boggia G., Camarda P. Downlink Packet Scheduling in LTE Cellular Networks: Key Design Issues and a Survey. IEEE Commun. Surveys and Tutorials. 2012; 15(2): 678 - 700.
5. Nonchev S., Valkama M. QoS-oriented packet scheduling for efficient video support in OFDMA-based packet radio systems. In: Multiple Access Communications. Springer Berlin Heidelberg. 2011; 168-180.
6. Zahariev N., Zaki Y., Li X., Carmelita G., Weerawardane T., Timm-Giel A. "Optimized service aware LTE MAC scheduler with comparison against other well known schedulers." In Wired/Wireless Internet Communication, Springer Berlin Heidelberg, 2012; 323-331.
7. Ekstrom, H. QoS control in the 3GPP evolved packet system. Communications Magazine, IEEE. 2009; 47(2): 76-83.
8. Lima, F. R. M., Maciel, T. F., Freitas, W. C., Cavalcanti, F. R. P. Resource assignment for rate maximization with QoS guarantees in multiservice wireless systems. IEEE Transactions on Vehicular Technology. 2012; 61(3): 1318-1332.
9. Brehm, M., Prakash, R. Overload-state downlink resource allocation in LTE MAC layer. Wireless networks. 2013; 19(5): 913-931.
10. Piro, G., Grieco, L. A., Boggia, G., Fortuna, R., Camarda, P. Two-level downlink scheduling for real-time multimedia services in LTE networks. IEEE Transactions on Multimedia. 2011; 13(5): 1052-1065.
11. Othman, M., Ferdosian, N., Rasul, T. Rated Window and Packet Size Differentiation Methods for Per-Rate TCP Fairness Over IEEE 802.11. Arabian Journal for Science and Engineering. 2015; 1-11.
12. Evolved Universal Terrestrial Radio Access (E-UTRA) and Evolved Universal Terrestrial Radio Access Network (E-UTRAN); Overall description; Stage 2, TS 36.300, 3rd Generation Partnership Project (3GPP), June 2012.
13. Evolved Universal Terrestrial Radio Access (E-UTRA); Physical Channels and Modulation, TS 36.211 3rd Generation Partnership Project (3GPP), June 2012.
14. Ferdosian, N., Othman, M. Two-Level QoS-Oriented Downlink Packet Schedulers in LTE Networks: A Review. In Proceedings of the First International Conference on Advanced Data and Information Engineering. Springer Singapore. 2014; 597-604.
15. Araniti, G., Condoluci, M., Militano, L., Iera, A. Adaptive resource allocation to multicast services in LTE systems. IEEE Transactions on Broadcasting. 2013; 59(4): 658-664.
16. Kwan, R., Leung, C. A survey of scheduling and interference mitigation in LTE. Journal of Electrical and Computer Engineering. 2010; 1-10.
17. Kela, P., Puttonen, J., Kolehmainen, N., Ristaniemi, T., Henttonen, T., Moisio, M. Dynamic packet scheduling performance in UTRA long term evolution downlink. 3rd International Symposium on Wireless Pervasive Computing. 2008; 308-313.
18. Tony, T. Performance Analysis for QoS-Aware Two-Layer Scheduling in LTE Networks. International Journal of Emerging Trends and Technology in Computer Science (IJETTCS). 2013; 2(2): 79-84.
19. Lai, W. K., Tang, C. L. QoS-aware downlink packet scheduling for LTE networks. Computer Networks. 2013; 57(7): 1689-1698.
20. Mushtaq, M. S., Fowler, S., Mellouk, A., Augustin, B. QoE/QoS-aware LTE downlink scheduler for VoIP with power saving. Journal of Network and Computer Applications. 2014.
21. 3GPP. Tech. Specif. Policy and charging control architecture, 3GPP TS 23.203 version 9.4.0 Release 9, (2010-03)