Packet Scheduling Mechanism for Multimedia Services to Guarantee QoS in 3GPP LTE System

Pligyu Shin† and Kwangsue Chung††,

Department of Communication Engineering, Kwangwoon University, Korea

Summary
The packet scheduling mechanism is an important technical issue in the Long Term Evolution (LTE) system. Several packet scheduling mechanisms have been proposed to improve the Quality of Service (QoS) in the LTE wireless communication system. However, most of these mechanisms cannot satisfy various QoS requirements simultaneously. In this paper, we analyze the problems in existing packet scheduling mechanisms in terms of QoS provisioning to propose a new packet scheduling mechanism. The Adaptive Weight-based Scheduling (AWS) mechanism is designed to satisfy various QoS requirements of application services simultaneously. The AWS mechanism calculates several weighting factors which provide QoS provisioning by using various information. Moreover, the AWS mechanism adaptively adjusts the ratio among the weighting factors by using each QoS satisfaction ratio, and then calculates the priority metrics for each user. Therefore, the AWS mechanism provides a guarantee to meet simultaneously various QoS requirements of the application services.

Key words:
Packet scheduling, LTE, QoS, Multimedia streaming

1. Introduction

Recently, the growing demand for network services, such as Voice over IP (VoIP), web browsing, video telephony, and video streaming, with constraints on delay and bandwidth requirements, pose new challenges in the design of future generation cellular networks. The 3rd Generation Partnership Project (3GPP) has developed the Long Term Evolution (LTE) wireless network system standards to respond to these challenges [1,2]. LTE can expand wireless cell coverage, improve system capacity, and support a low transmission delay and high transmission rate. As a result, LTE has become a technology that enables users and network providers to reduce costs and improve the Quality of Service (QoS) [3].

Scheduling in LTE is closely related to the QoS guarantee for a wide range of application services. Scheduling is performed in the Medium Access Control (MAC) layer using various types of status information such as the Channel Quality Indicator (CQI) reporting information, which is wireless channel status information obtained from User Equipment (UE). The role of a scheduler is to satisfy the service requirements by the efficient use of wireless resources and the allocation of those resources to users [4-6]. Several scheduling mechanisms to guarantee the QoS requirements such as the maximum allowable delay and required transmission rate have been proposed. However, the existing scheduling mechanisms cannot satisfy various QoS requirements simultaneously.

In this paper, we propose an Adaptive Weight-based Scheduling (AWS) mechanism in order to satisfy various QoS requirements from different types of application services simultaneously. In order to satisfy various QoS requirements, AWS computes weighted values for calculating a priority metric using a variety of information such as Head Of Line (HOL) delay, throughput, receiving buffer occupancy, and loss rate. Further, a priority metric calculation method that adjusts the weight ratio adaptively is used on the basis of the satisfaction ratio of each QoS requirement. Through the AWS mechanism, a variety of QoS requirements for application services can be satisfied simultaneously.

This paper is organized as follows. In Section 2, we describe the LTE wireless network systems, the existing scheduling mechanisms, and their problems and solutions. Section 3 presents the detailed operation of the AWS mechanism. Section 4 describes the simulation environment for evaluating the performance of AWS and shows the results of this evaluation. Finally, Section 5 presents the conclusions.

2. Related Work

2.1 LTE Wireless Network System

3G cellular network systems have been developed using technologies such as High Speed Downlink Packet Access (HSDPA) and High Speed Uplink Packet Access (HSUPA) for the purpose of improving the data transmission rate and QoS. However, as the demand for fast and good-quality service has increased, a need has developed for technology that is more advanced than the 3G cellular network systems. Standards have been developed in the 3GPP to address this need. For LTE, a more advanced technology was required than that of the 3G cellular networks based on voice services in order to support various application services effectively. LTE

Manuscript received July 5, 2014
Manuscript revised July 20, 2014

resolved the problems that arose from various requirements by doubling the spectral efficiency of the 3G cellular systems, improving the coverage in terms of the transmission rate of the cell-boundary users, increasing the system capacity by enhancing the transmission rate, and considering the mobility of users. Further, various technologies were introduced to support QoS [6].

Table 1: Standardized QCI characteristics

QCI	Resource type	Priority	Packet delay budget (ms)	Packet loss rate	Example services
1	GBR	2	100	10^{-2}	Conversational voice
2	GBR	4	150	10^{-3}	Conversational video (live streaming)
3	GBR	3	50	10^{-3}	Real time gaming
4	GBR	5	300	10^{-6}	Non-conversational video (buffered streaming)
5	Non- GBR	1	100	10^{-6}	IMS signaling
6	Non- GBR	6	300	10^{-6}	Video (buffered streaming) TCP based (e.g. WWW, e-mail, chat, FTP)
7	Non- GBR	7	100	10^{-3}	Voice, video (live streaming), interactive gaming
8	Non- GBR	8	300	10^{-6}	Video (buffered streaming) TCP based (e.g. WWW, e-mail, chat, FTP)
9	Non- GBR	9			

In LTE, a graded QoS can be provided according to the type of application service. The QoS Class Identifier (QCI) is defined in order to provide a graded service in Dedicated Radio Bearer (DRB). The QCI is categorized into several classes according to various QoS requirements. QCI, as shown in Table 1, is largely divided into two categories, the Guaranteed Bit-Rate (GBR) and the Non-Guaranteed Bit-Rate (Non-GBR), and is defined by various QoS requirements such as the priority levels, the maximum allowable delay, and acceptable packet loss rate [7,8].

2.2 Existing Scheduling Mechanism

Various existing scheduling mechanisms have been proposed to improve QoS. The QoS of application services is provided according to QoS parameters for the minimum quality guarantee defined for each application service. The scheduling mechanism performed its operation on the basis of various QoS parameters.

The Token Bucket-based Scheduling (TBS) mechanism, which was proposed for guaranteeing the transmission rate of application services, performs scheduling by dividing tasks into two groups: the GBR traffic group and the non-GBR traffic group [9]. The TBS mechanism adds a token bucket model at the transmission queue of each user in the eNB, and sets up a token generation rate for the token buckets as each user's required transmission rate in order to guarantee the transmission rate.

Scheduling mechanisms for satisfying the maximum allowable delay involve a policy designed to transmit packets within a limited required delay. The policy for satisfying the maximum allowable delay is the most important requirement of a multimedia application service such as streaming and VoIP services. Various scheduling mechanisms have been proposed for satisfying the maximum allowable delay of the application services. The Modified-Largest Weight Delay First (M-LWDF) mechanism considers the channel status of a UE and the maximum allowable delay of an application service, adding a metric base to the Proportional Fair (PF) mechanism for the existing Largest Weight Delay First (LWDF) mechanism [10]. Non-realtime applications calculate metrics using the PF mechanism, while realtime applications such as a streaming and VoIP service perform scheduling using metrics calculated by applying the weights of M-LWDF. The Delay Prioritized Scheduling (DPS) mechanism selects the candidates of a transmission queue in the order of the Head Of Line (HOL) delay at each transmission queue, and then performs scheduling by allocating a Resource Block (RB) that has the best channel status so that data can be sent urgently to transmission queues that are close to the maximum allowable delay [11].

In terms of minimizing the packet loss rate, it is important to maintain a receive buffer status. The Buffer-Aware Traffic-Dependent (BATD) mechanism has been proposed, which is a scheduling mechanism that minimizes the packet loss rate caused by the receive buffer overflow, while maintaining a high transmission rate and fairness between users [12]. Further, similar to BATD, the Channel-Adapted Buffer-Aware (CABA) mechanism was proposed as a scheduling policy for reducing the packet loss rate due to the overflow while periodically monitoring the receive buffer status and adjusting the priority metrics according to the buffer status [13].

The existing scheduling mechanisms proposed for various QoS improvements led to various performance improvement results such as minimizing the transmission delay and maximizing the transmission rate and resource efficiency. However, if the existing scheduling mechanisms are applied to a network environment where there are various application services and various QoS requirements for different users needs to be guaranteed, performance improvements might be made, particularly

QoS requirements; however, this would bring about a limitation whereby the other QoS requirements would not be able to be considered concurrently. In order to improve the quality of service by resolving this limitation and satisfying various QoS requirements simultaneously, studies of an adaptive weight-based scheduling mechanism are needed, in which weighting factors are applied to each QoS requirement and the satisfaction of QoS requirements is considered for each application service.

3. Adaptive Weight-based Scheduling

In wireless network environments, numerous mechanisms have been proposed for performance enhancement, such as minimizing the transmission delay, maximizing the transmission rate, and minimizing the loss rate. However, most existing mechanisms cannot satisfy various QoS requirements simultaneously. In order to resolve this problem, we propose an Adaptive Weight-based Scheduling (AWS) mechanism that is designed to satisfy the QoS requirements of various application services in a cellular network environment. The proposed AWS mechanism implements a priority metric-based scheduling that calculates the weights for QoS requirements (such as the maximum allowable delay, the minimum required transmission rate, and the maximum allowable transmission loss), and performs an adaptive control of the weighting factor in accordance with the satisfaction ratio of the QoS requirements.

3.1 Overview of the AWS Mechanism

Fig. 1 shows the system model of the AWS mechanism that we propose as a scheme to satisfy the QoS requirements of various application services simultaneously. The proposed AWS mechanism collects information regarding the User Equipment (UE) channel state, transmission delay, transmission rate, buffer status, loss rate, etc. The collected information is used for the calculation of a priority metric with the scheduler within the evolved Node B (eNB).

The system model of the AWS mechanism has a structure that ensures interoperability among several modules such as: (i) the transmission queues responsible for the data transmission of every application service; (ii) the information collection module that receives the transmission queue status and that receives the buffer status and the channel states of UE; and (iii) the eNB scheduler that implements scheduling.

The system model accommodates data transmission queues that are classified with a Dedicated Radio Bearer (DRB) in accordance with the types of application services. Further, in order to satisfy the required transmission rate, each transmission queue is constructed with a token bucket. The information collection modules consist of a Transmission Queue Status Report (TQSR) module, a Receiving Buffer Status Report (RBSR) module, and a Channel Quality Indicator (CQI) module. The TQSR module is responsible for collecting information on the status of each transmission queue and token bucket and transmitting it to the eNB scheduler. The RBSR module then performs the function of periodically receiving the receiving buffer information at each UE and transmitting it to the eNB scheduler, and the CQI module then delivers the information collected on the UE channel state to the Adaptive Modulation and the Coding (AMC) modules, which regulate the transmission rate in accordance with the channel state, and to the eNB scheduler that uses the collected information for calculating the priority metric for scheduling.

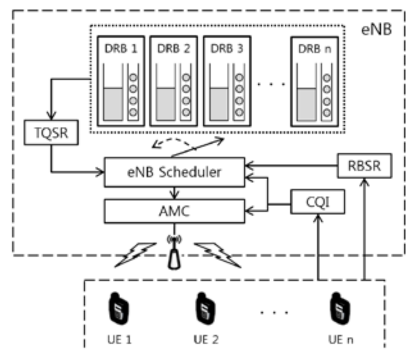


Fig. 1 System model of the AWS mechanism

3.2 Calculation of the Weight to Satisfy the Maximum Allowable Delay

In the AWS mechanism, the weight is calculated to satisfy the maximum allowable delay by using the information (transmitted by the TQSR module), about the Head Of Line (HOL) delay at each transmission queue. The HOL delay denotes the time taken for a data packet transferred from the application layer to arrive at the corresponding transmission queue and be transmitted. Fig. 2 shows the transmission queue model of eNB designed to obtain the HOL delay information [10].

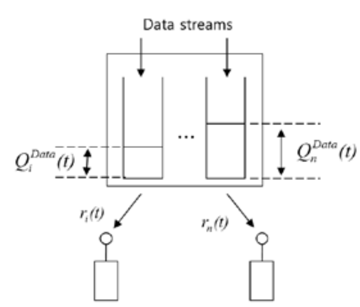


Fig. 2 Transmission queue model of eNB

$Q_{\mathrm{i}}^{Data}(t)$ represents the number of packets accumulated at the transmission queue of the i-th user at the time t, and denotes the transmission rate calculated on the basis of the wireless channel status of the UE transmitted through the CQI module. The HOL delay of the transmission queue is calculated using Eq. (1).

$$D_{HOL,i}^{Data}(t) = \frac{Q_i^{Data}(t)}{r_i(t)} \tag{1}$$

The HOL delay change depending on the size of data accumulated in the transmission queues and the transmission rate of the UE. The state of high accumulation of the data waiting to be transmitted in the transmission queues implies a high HOL delay and a high transmission delay due to a long waiting time. In contrast, the small size of accumulated data in the transmission queue implies a low HOL delay and a low transmission delay due to a short waiting time.

Scheduling is performed in order to reduce the transmission delay by using the characteristic whereby the HOL delay is directly related to the data transmission delay. If a transmission queue has a higher HOL delay than the other transmission queues, the scheduler is assigns a high priority metric to its transmission queue to reduce transmission delay.

The existing HOL delay-based scheduling scheme operates according to the measure of assigning a weighting factor that increases in proportion to the HOL delay. If this measure is used in the case of network congestion, only the given transmission queue can transfer data, which leads to the starvation of the other transmission queues, ultimately resulting in a degradation of QoS for all users connected to the same network. This problem can be solved by limiting the maximum value of the weight used for satisfying the maximum allowable delay to 1, as defined by Eq. (2). Hence, a fair allocation of weight can be ensured even in cases where all transmission queues have high HOL delay. The calculation method of the proposed scheme to determine the weight is defined by Eq. (3).

$$w_i^{Delay}(t) \le 1 \tag{2}$$

$$w_i^{Delay}(t) = \begin{cases} \left(\dfrac{D_{HOL,i}^{Data}(t)}{\tau_i^{Data}}\right)^2, & if\ D_{HOL,i}^{Data}(t) < 1 \\ 1, & otherwise \end{cases} \tag{3}$$

In AWS mechanisms, higher priority metric is assign to given transmission queue as its HOL delay approaches the maximum allowable delay . Therefore, the transmission queue concerned can be given the data transmission opportunity on a preferential basis.

3.3 Calculation of the Weighted to Satisfy a Required Transmission Rate

The weight value to satisfy the required transmission rate is calculated by using the information of the token bucket provided for each transmission queue within eNB, as transferred by the TQSR module. The token bucket method is widely used as a traffic shaping function, which either stabilizes traffic prone to large fluctuations in the data transmission rate or regulates the specified transmission rate. The token bucket method operates by eliminating the exact amount of a token from its bucket corresponding to the amount of the data transmitted by a transmission queue. Fig. 3 shows the operation of the token bucket method.

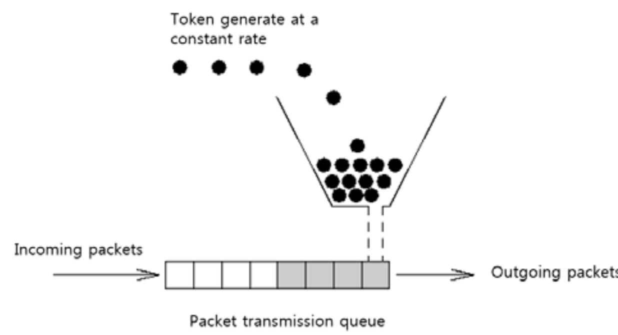


Fig. 3 Operation of the token bucket method

The characteristics of the token bucket method are such that the number of tokens remaining in the bucket decreases if the data transmission rate of the transmission queue exceeds the token generation rate; conversely, it keeps increasing if the data transmission rate becomes lower than the token generation rate. The weight value of the AWS mechanism to satisfy a required transmission rate is calculated according to this characteristic of changing the number of tokens remaining in the bucket depending on the data transmission behavior in the transmission queue. The required transmission rate of the application services in each transmission queue is determined to be the token generation rate. Therefore, the number of remaining tokens increases if the data transmission rate in the transmission queue does not meet the required transmission rate. Fig. 4 shows the transmission queue model of eNB based on the token bucket method [9].

In the transmission queue model based on the token bucket method, R_i^{Token}, the token generation rate of the bucket applied to the transmission queue of user i, is determined to be R_i^{Req} as defined by Eq. (4), R_i^{Req} is the required transmission rate of user i.

$$R_i^{Token} = R_i^{Req} \tag{4}$$

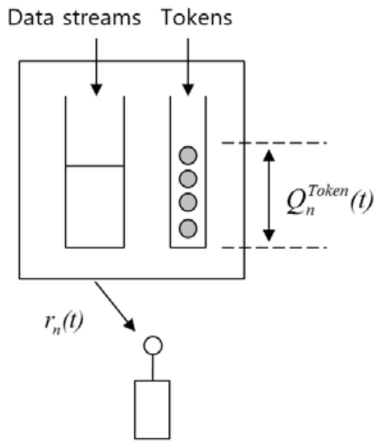


Fig. 4 Transmission queue model based on token bucket method

L_i^{Token}, which is the size of the token bucket, can be expressed as the multiplication of D_i^{Token} (token consumption time) and R_i^{Token} (token generation rate), as defined by Eq. (5).

$$L_i^{Token} = R_i^{Token} D_i^{Token} \tag{5}$$

The token consumption time, D_i^{Token}, is set to 1 second, and the token bucket size is set to 100 kbits, if the required transmission rate is 100 kbps. Then, the HOL delay of the token bucket is calculated using Eq. (6). Q_i^{Token} denotes the number of tokens remaining in the token bucket of the i-th user at the time t.

$$D_{HOL,i}^{Token}(t) = \frac{Q_i^{Token}(t)}{r_i(t)} \tag{6}$$

The HOL delay of the token bucket varies depending on the number of tokens remaining in the token bucket. If a large number of tokens are accumulated in the token bucket, the calculated HOL delay of the token bucket will be high, which would indicate that the data transmission rate of the transmission queue falls short of the required transmission rate. In this situation, the scheduler should provide a high priority metric to satisfy the required data transmission rate of the transmission queue concerned.

The next step involves the calculation of weight to satisfy the required transmission rate of the transmission queue by using the correlation between the HOL delay of the token bucket and the changes in the priority metric. Eq. (7) is the equation for calculating the weight to satisfy the required transmission rate in the proposed AWS mechanism.

$$w_i^{Token}(t) = \left(\frac{D_{HOL,i}^{Token}(t)}{D_i^{Token}}\right)^2 \tag{7}$$

3.4 Calculation of the Weighted to Satisfy the Maximum Allowable Loss Rate

The calculation of the weight to satisfy the maximum allowable loss rate is performed using information on the receiving buffer state and the loss rate. The scheduler should be taken into account to prevent overflow of the receiving buffer caused by excessive data transmission in eNB that is likely to occur through an abrupt increase in the data transmission rate during unstable traffic with Variable Bit Rate (VBR). Further, weights should be adjusted to reduce the packet loss of the application services (caused by the overflow of the transmitted data) by utilizing the loss rate information monitored periodically. Fig. 5 shows the receiving buffer model of UE.

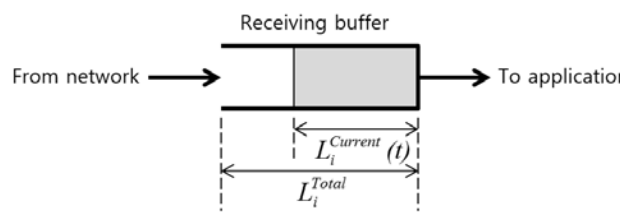


Fig. 5 Receiving buffer model of UE

In the receiving buffer model, L_i^{Total} denotes the total size of the receiving buffer and $L_i^{Current}(t)$ represents the amount of buffer used at time t. In order to prevent packet loss due to overflow, the weight should be adjusted to be low in inverse proportion to the buffer occupancy so as to allow the transmission of a small number of data packets. Conversely, in the case of low buffer occupancy, a higher weight should be assigned in order to allow the transmission of a large number of data packets. Furthermore, the weight of the loss rate are calculated to decrease as the mean loss rate approaches the maximum allowable loss rate. Eq. (8) shows the weight value of the loss rate to satisfy the maximum allowable loss rate of user i calculated on the basis of the information of the buffer status and the loss rate.

$$w_i^{Loss}(t) = \left(1 - \frac{L_i^{Current}(t)}{L_i^{Total}}\right)\left(1 - \frac{r_i^{Loss}(t)}{\delta_i^{Loss}}\right) \tag{8}$$

In Eq. (8), r_i^{Loss} denotes the mean loss rate at time t, and δ_i^{Loss} denotes the maximum allowable loss rate of the given application service. Eq. (8) demonstrate that decreasing weight is applied as the receiving buffer occupancy approaches the maximum value and the mean packet loss rate approximates the maximum allowable loss rate.

3.5 Calculation of the Priority Metric Applying an Adaptive Weighting Factor

The calculation of the priority metric using an adaptive weighting factor should be preceded by the calculation of the satisfaction ratio upon QoS requirements. The satisfaction ratio of the maximum allowable delay can be calculated according to the occurrence frequency of exceeding the maximum allowable delay by the HOL delay of the transmission queue. The satisfaction ratio of the minimum required transmission rate is calculated according to the occurrence frequency by which the mean transmission rate falls short of the required transmission rate. In addition, the satisfaction ratio of the maximum allowable loss rate is calculated according to the occurrence frequency of the UE receiving buffer overflow.

Eq. (9) represents the satisfaction ratio of the maximum allowable delay, and the calculation is carried out by utilizing the frequency at which the HOL delay exceeded τ_i^{Data}, the maximum allowable delay, during the time period T. Eq. (10) represents the satisfaction ratio of the minimum required transmission rate, and the calculation is carried out by utilizing the frequency at which the mean transmission rate falls short of the required transmission rate during the time period T. The satisfaction ratio of the maximum allowable loss rate can be calculated with Eq. (11) by utilizing the frequency of the overflow occurring in the receiving buffer of the UE during the time period T. The degree to which the given QoS requirements are satisfied increases when the values calculated approach 0.

$$x_i(t) = \frac{Count\ of\ \left(D_{HOL,i}^{Data} > \tau_i^{Data}\right)\ occurence}{T} \quad (9)$$

$$y_i(t) = \frac{Count\ of\ \left(\bar{R}_i(t) > R_i^{Req}\right)\ occurence}{T} \quad (10)$$

$$z_i(t) = \frac{Count\ of\ (Overflow)\ occurence}{T} \quad (11)$$

The satisfaction ratio of the QoS requirements is between 0 and 1. In the proposed AWS mechanism, we use an inverse logarithmic function, which manifests a higher increasing rate than an exponential function, in order to achieve a rapid increase in the weighting factor of between 0 and 1. The values to be used for the adaptive weighting factor calculation based on the satisfaction ratio of the QoS requirements can be obtained using Eq. (12).

$$\begin{cases} X_i(t) = -\ln(1 - x_i(t)) \\ Y_i(t) = -\ln(1 - y_i(t)) \\ Z_i(t) = -\ln(1 - z_i(t)) \end{cases} \quad (12)$$

The adaptive weighting factor to be used for the priority metric can be calculated on the basis of the satisfaction ratio of the QoS requirements, as defined in Eq. (13). Eq. (13) represents adaptive weighting factor to satisfy the maximum allowable delay, the required transmission rate, and the maximum allowable loss rate.

$$\begin{cases} \bar{X}_i(t) = \dfrac{X_i(t)}{X_i(t) + Y_i(t) + Z_i(t)} \\ \bar{Y}_i(t) = \dfrac{Y_i(t)}{X_i(t) + Y_i(t) + Z_i(t)} \\ \bar{Z}_i(t) = \dfrac{Z_i(t)}{X_i(t) + Y_i(t) + Z_i(t)} \end{cases} \quad (13)$$

In the proposed AWS mechanism, the priority metric of user i applying the weighting factor adapted to the satisfaction ratio of the QoS requirements is calculated using Eq. (14).

$$\begin{aligned} m_i^{AWS-adaptive}&(t) \\ = &\frac{m_i^{PF}(t)}{\#\ of\ QoS\ requirements\ of\ user\ i} \\ &\cdot \left\{B_i^{Delay}\bar{X}_i(t)w_i^{Delay}(t) + B_i^{Token}\bar{Y}_i(t)w_i^{Token}(t) \right. \\ &\left. + B_i^{Loss}\bar{Z}_i(t)w_i^{Loss}(t)\right\} \end{aligned} \quad (14)$$

m_i^{PF} is the priority metric of the PF scheduling mechanism. B_i^{Delay}, B_i^{Token}, and B_i^{Loss} denote the boolean values that indicate the existence of the QoS requirements for user i.

4. Experimental Results and Analysis

We analyze the performance of the proposed AWS mechanism using simulations. The performances of application services such as the average transmission rate, average transmission delay, and average packet loss rate are measured. Additionally, through measuring the QoS requirement satisfaction ratio for each application service, we can verify that the AWS mechanism performs an adaptive scheduling that satisfies various QoS requirements simultaneously.

4.1 Simulation Environments

To measure the performance of the AWS mechanism, we used the LTE-Sim simulator, which specializes in measuring scheduling mechanisms [14]. Table 2 shows the simulation parameters of the LTE wireless networks for measuring the performance of the AWS mechanism.

The simulations configuration uses a cell with a 500 m radius. The number of UEs used for the simulations changed from 5 to 30, and the simulation time was set to 100 seconds. The mapping information for the transmission rate based on the downlink SNR values was selected

Table 2: Simulation parameters for LTE wireless network

Parameter	Value
Number of cells	1 (Hexagonal)
Carrier frequency	2 GHz
Bandwidth	5 MHz
Number of RBs	25
Overhead	MAC and RLC : 5 bytes PDCP : 2 bytes CRC : 3 bytes
Cell radius	0.5 km
CQI reporting period	2 TTI (2 ms)
Scheduling duration	1 TTI (1 ms)
Simulation time	100 s

Table 3: Data rate mapping table according to downlink SNR value

Downlink SNR value (dB)	Modulation and coding	Transmission rate (kbps)
1.7	QPSK (1/2)	168
3.7	QPSK (2/3)	224
4.5	QPSK (3/4)	252
7.2	16QAM (1/2)	336
9.5	16QAM (2/3)	448
10.7	16QAM (3/4)	504
14.8	64QAM (2/3)	672
16.1	64QAM (3/4)	756

Table 4: QoS requirements of each application service

Type of application services	Maximum allowable delay	Required transmission rate	Maximum allowable loss rate
Video	100 ms	180 kbps	-
Audio	50 ms	-	10^{-2}
Control msg.	-	-	10^{-6}
Web	-	-	-

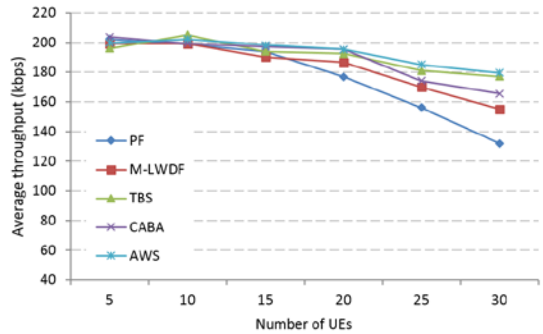
from Table 3.

The QoS requirements for each application service were set in order to measure the satisfaction ratios for various QoS requirements through simulations. Table 4 presents the QoS requirements for each type of application service.

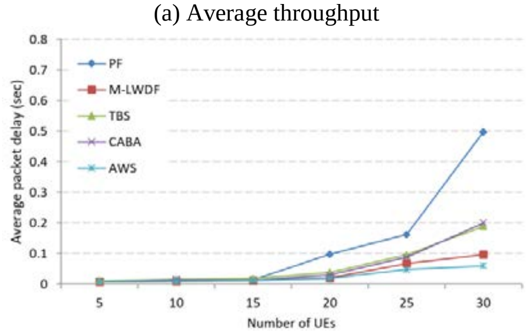
An H.264 video traffic model with a 192-kbps bit rate and VBR characteristic was used as the video traffic service for the simulations, and a G.729 audio traffic model with a 8.4-kbps bit rate and on-off function was used for an audio traffic service.

4.2 Performance Evaluation of Video Traffic Services

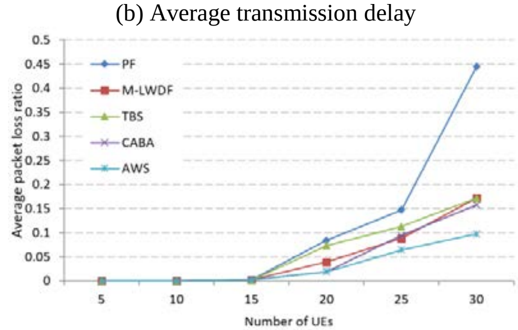
We evaluated the performance of the video traffic service of the proposed AWS mechanism and existing scheduling mechanisms. The average transmission rate, average transmission delay, and packet loss rate were compared. Fig. 6(a) shows the average transmission rates of the AWS mechanism and the existing scheduling mechanisms. Through the evaluation of the transmission rate performance of the video traffic service, we could see that, among the existing mechanisms, the PF mechanism showed



(a) Average throughput



(b) Average transmission delay



(c) Average packet loss rate

Fig. 6 Performance of video traffic services

the lowest performances because the PF mechanism does not consider the QoS requirements. Further, we could see that the TBS mechanism, which considers the QoS requirements for the required transmission rate, showed relatively high performances, and the proposed AWS mechanism showed the highest performances. In the case of increasing network loads caused by a large number of UEs, it was confirmed that the AWS mechanism had the highest performances because AWS adjusts the weight factor ratios adaptively by using QoS satisfaction ratios of the required transmission rate and satisfies the required transmission rate of the video traffic service of 170 kbps. This means that the AWS mechanism can serve the greatest number of UEs. Fig. 6(b) shows the average transmission delays of the AWS mechanism and the existing scheduling mechanisms. Among the existing scheduling mechanisms, the M-LWDF mechanism (in which weights are applied to the transmission queues using the HOL delay), showed relatively low transmission delays. The AWS mechanism showed the best performances. Moreover, when using the AWS mechanism, the number of UEs satisfying the required transmission delay, 100 ms, was the highest. Fig. 6(c) shows the average packet loss rate results for video traffic services. The CABA mechanism, which periodically monitors the receive buffer status of UEs to minimize the loss rate, showed relatively low loss rates, while the AWS mechanism showed the lowest loss rates. From the result, it was observed that the AWS mechanism showed the best performances compared to the existing scheduling mechanisms.

4.3 QoS Satisfaction Ratio of Video Traffic Services

Through simulation results of video traffic services using the AWS mechanism and existing scheduling mechanisms, we evaluated the satisfaction ratios for the QoS requirements for video traffic services.

Table 5 shows the QoS satisfaction ratios for the required transmission rate of video traffic services, 180 kbps, for the AWS mechanism and existing scheduling mechanisms according to the change in the number of UEs. Based on the QoS satisfaction ratio results for the required transmission rate of video traffic services, in a situation where the network load increases, it was possible for the AWS mechanism to satisfy the required transmission rate for 25 UEs, which was the highest. Further, the TBS had the highest QoS satisfaction among the existing scheduling mechanisms.

Table 6 shows the QoS satisfaction ratio for the maximum allowable delay, 100 ms, of video traffic services. Through simulations, the AWS, M-LWDF, and CABA mechanisms were able to satisfy the maximum allowable delay for 25 UEs. Additionally, when using 30 UEs, the AWS mechanism had the highest QoS satisfaction ratio for the maximum allowable delay, which was 0.958.

4.4 Performance Evaluation of Audio Traffic Services

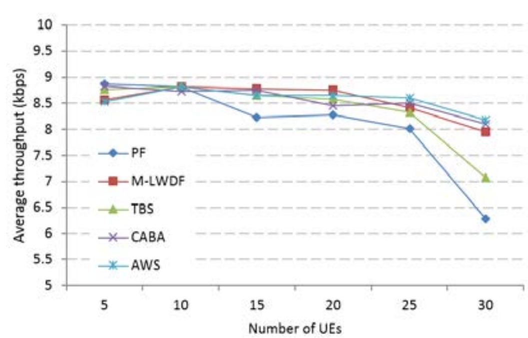
We evaluated the performance of the audio traffic service of the proposed AWS mechanism and existing scheduling mechanisms. The average transmission rate, average transmission delay, and packet loss rate were compared. Fig. 7(a) shows the simulation result for the average transmission rate of audio traffic services for the AWS and existing scheduling mechanisms. There was a tendency for the average transmission rate to decrease as the number of UEs increased. Fig. 7(b) shows the simulation results for the average transmission delay of audio traffic services. The AWS mechanism showed the best performances compared to the other scheduling mechanisms because the AWS adjusts the weight adaptively according to the satisfaction ratio of the QoS requirements. Fig. 7(c) shows the simulation results of the packet loss rates. Based on the packet loss rate results, we could see that the AWS mechanism had the lowest loss rates.

Table 5: QoS satisfaction ratio of required transmission rate for video traffic services

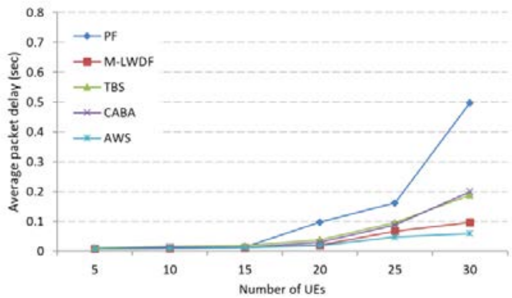
Number of UEs	PF	M-LWDF	TBS	CABA	AWS
5	1	1	1	1	1
10	1	1	1	1	1
15	1	1	1	1	1
20	0.959	1	1	1	1
25	0.847	0.923	0.983	0.945	1
30	0.716	0.841	0.961	0.898	0.974

Table 6: QoS satisfaction ratio of maximum allowable delay for video traffic services

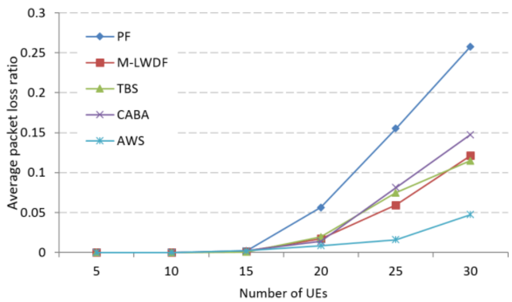
Number of UEs	PF	M-LWDF	TBS	CABA	AWS
5	1	1	1	1	1
10	1	1	1	1	1
15	1	1	1	1	1
20	0.65	1	1	1	1
25	0	1	0.8	1	1
30	0	0.546	0	0	0.958



(a) Average throughput



(b) Average transmission delay



(c) Average packet loss rate

Fig. 7 Performance of audio traffic services

4.5 QoS Satisfaction Ratio of Audio Traffic Services

Through simulation results of audio traffic services using the AWS mechanism and existing scheduling mechanisms, we evaluated the satisfaction ratios for the QoS requirements for audio traffic services. Table 7 shows the QoS satisfaction ratios for the maximum allowable delay of audio traffic services, 50ms, for the AWS mechanism and existing scheduling mechanisms. In a situation where the network load increased, the AWS mechanism was able to satisfy the maximum allowable delay for 25 UEs, which was the highest. Among the existing scheduling mechanisms, the M-LWDF showed the highest QoS satisfaction ratio.

Table 8 shows the QoS satisfaction ratio for the maximum allowable loss rate, 1%, for the AWS and existing mechanisms. The AWS mechanism, which calculates the weights of the loss rates and adjusts the weight adaptively according to the QoS satisfaction ratio, had the lowest loss rate, and a total of 20 UEs satisfied the maximum allowable loss rate. Further, among the existing scheduling mechanisms, the CABA mechanism, which minimizes the loss rate by periodically monitoring the receive buffer status, showed the best result.

Table 7: QoS satisfaction ratio of maximum allowable delay for audio traffic services

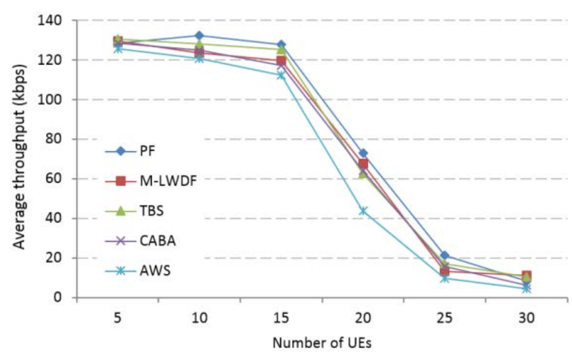
Number of UEs	**PF**	**M-LWDF**	**TBS**	**CABA**	**AWS**
5	1	1	1	1	1
10	1	1	1	1	1
15	1	1	1	1	1
20	0.5	1	1	1	1
25	0	0.65	0.8	0.245	1
30	0	0.06	0	0	0.806

Table 8: QoS satisfaction ratio of maximum allowable loss rate for audio traffic services

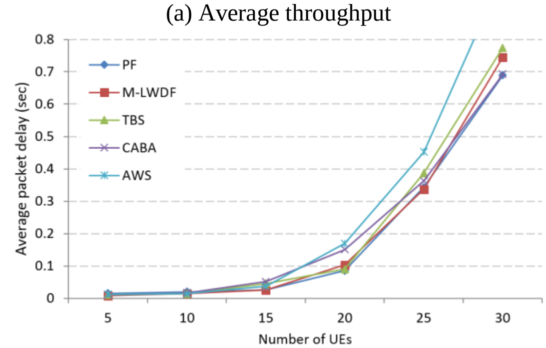
Number of UEs	**PF**	**M-LWDF**	**TBS**	**CABA**	**AWS**
5	1	1	1	1	1
10	1	1	1	1	1
15	1	1	1	1	1
20	0	0.277	0.05	0.612	1
25	0	0	0	0	0.379
30	0	0	0	0	0

4.6 Performance Evaluation of Web Traffic Services

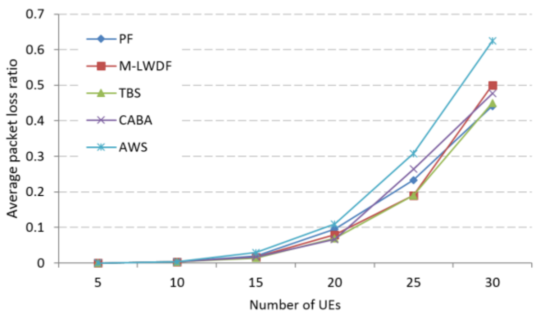
We evaluated the performance of the web traffic service of the proposed AWS mechanism and existing scheduling mechanisms. Fig. 8 shows the simulation results for the average transmission rate, average transmission delay, and packet loss rate of web traffic services for the AWS and existing scheduling mechanisms. The performance of the AWS mechanism for the web traffic services is similar to existing scheduling mechanisms. On the other hand, the AWS mechanism for the video and voice services shows higher performance than existing scheduling mechanisms.



(a) Average throughput



(b) Average transmission delay



(c) Average packet loss rate

Fig. 8 Performance of web traffic services

5. Conclusion

LTE has been developed as a technology that improves performance by expanding the coverage, increasing the system capacity, and supporting a low transmission delay and high transmission rate. With LTE, a variety of functionalities have been added to eNB in order to provide users with high performance and service quality. In particular, scheduling performs important roles in QoS performance improvements. Several scheduling mechanisms have been proposed to improve the QoS. However, the existing scheduling mechanisms cannot satisfy several QoS requirements simultaneously.

In this paper, we propose an AWS mechanism in order to satisfy various QoS requirements from different types of application services simultaneously. In order to satisfy various QoS requirements, the AWS computes weight values for calculating a priority metric using a variety of information. Further, a priority metric calculation that adjusts the weight ratio adaptively is used according to the satisfaction ratio of the QoS requirements. Simulation shows that the proposed AWS mechanism improves the performance compared to existing scheduling mechanisms in terms of the transmission rate, transmission delay, and loss rate. Further, when network loads increase, the AWS mechanism provides services that satisfy the QoS requirements, with a greater number of UEs.

Acknowledgments

This work was supported by the IT R&D program of MSIP/KEIT.[10044553 , Development of Multiple Unified Video Production System Technology based on Web Object]

References

[1] 3rd Generation Partnership Project (3GPP), "Technical Specification Group Radio Access Network - Requirements for Evolved UTRA (E-UTRA) and Evolved Universal Terrestrial Radio Access Network (E-UTRAN)," 3GPP TS 25.913, 2011.

[2] E. Dahlman, S. Parkvall, J. Skold, and P. Beming, "3G Evolution: HSPA and LTE for Mobile Broadband," Academic Press, August 2007.

[3] B. Al-Manthari, H. Hassanein, and N. Nasser, "Packet Scheduling in 3.5G High-speed Downlink Packet Access Networks: Breath and Depth," IEEE Network, vol. 21, no. 1, pp. 41-46, January-February 2007.

[4] T. Sesia, I. Toufik, and M. Baker, "LTE, The UMTS Long Term Evolution: From theory to Practice," John Wiley & Sons, 2009.

[5] E. Dahlman, S. Parkvall, J. Skold, and P. Beming, "3G Evolution HSPA and LTE for Mobile Broadband," Academic Press, 2008.

[6] H. Ekstrom, "QoS control in the 3GPP Evolved Packet System," IEEE Communication Magazine, vol. 47, pp. 76-83, February 2009.

[7] 3rd Generation Partnership Project (3GPP), "Technical Specification Group Services and System Aspects - Policy and charging control architecture (Release 9)," 3GPP TS 23.203, 2011.

[8] International Telecommunication Union (ITU), "Overall Network Operation, Telephone Service, Service Operation and human factors, ITU-T Recommendation," E.800, Annex B, August 2008.

[9] P. Shin and K. Chung, "Token Bucket-based Scheduling Mechanism for Multimedia Services in 3GPP LTE system," Journal of Korea Institute of Information Technology, vol. 11, no. 1, pp. 101-110, January 2013.

[10] M. Andrews, K. Kumaran, K. Ramanan, A. Stolyar, P. Whiting, and R. Vijayakumar, "Providing Quality of Service over a Shared Wireless link," IEEE Communication Magazine, vol. 39, no. 2, pp. 150-154, February 2001.

[11] K. Sandrasegaran, H. Ramli, and R. Basukala, "Delay-Prioritized Scheduling (DPS) for Real Time Traffic in 3GPP LTE system," in Proc. of the Wireless Communications and Networking Conference, pp. 1-6, April 2010.

[12] J. Huang and Z. Niu, "Buffer-aware and Traffic-dependent Packet Scheduling in Wireless OFDM Networks," in Proc. of the IEEE Wireless Communication and Networking Conference, pp. 1554-1558, March 2007.

[13] Y. Lin and G. Yue, "Channel-adapted and Buffer-aware Packet Scheduling in LTE Wireless Communication system," in Proc. of the Wireless Communications, Networking and Mobile Computing, pp.1-4, October 2008.

[14] G. Piro, "LTE-sim - The LTE simulator," http://telematics.poliba.it/LTE -Sim/

Pilgyu Shin received the B.S. and Ph.D. degrees in Electrical and Communication Engineering from Kwangwoon University in 2005 and 2013, respectively. During 2008-2011, he worked in Innowireless Co., Ltd. for develop testing equipment of mobile communication technology such as Mobile WiMAX and LTE. He now with CETECOM Korea for GCF/PTCRB certification project management.

Kwangsue Chung received his B.S. degree from Hanyang University, Seoul, Korea, M.S. degree from KAIST (Korea Advanced Institute of Science and Technology), Seoul, Korea, and Ph.D. degree from the University of Florida, Gainesville, Florida, USA, all from the respective Electrical Engineering Departments. Before joining Kwangwoon University in 1993, he spent 10 years with ETRI (Electronics and Telecommunications Research Institute) as a research staff. He was also an adjunct professor of KAIST from 1991 to 1992, and a visiting scholar at the University of California, Irvine from 2003 to 2004. His research interests include communication protocols and networks, QoS mechanism, and video streaming.