IE Interfaces
Vol. 17, Special Edition, pp. 111-121, December 2004.

Time Utility and Channel State based Wireless Downlink Packet Scheduling Algorithm for OFDMA System

Seungwan Ryu[1,2†] · Hyun Hwa Seo[2] · Soo Jung Chung[2] · Soon Yong Lim[2] · Seikwon Park[1]

[1]Department of Information Systems, Chung-Ang University, Ansung, 456-756
[2]Broadband Mobile MAC Research Ream, Mobile Telecom. Research Division, ETRI, Daejon, Korea, 305-350

OFDMA 무선 시스템에서의 시간-효용과 채널 상태 기반의 하향 링크 패킷 스케줄링

류승완[1,2] · 서현화[2] · 정수정[2] · 임순용[2] · 박세권[1]

[1]중앙대학교 정보시스템학과 / [2]한국전자통신연구원 광대역무선 MAC 연구팀

In this paper, we propose an urgency and efficiency based wireless packet scheduling (UEPS) algorithm that is able to schedule real time (RT) and non-real time (NRT) traffics at the same time. The proposed UEPS algorithm is designed to support wireless downlink packet scheduling in the OFDMA system which is a strong candidate wireless system for the next generation mobile communications. The UEPS algorithm uses the time-utility function as a scheduling urgency factor and the relative status of the current channel to the average one as an efficiency indicator of radio resource usage. The design goal of the UEPS algorithm is to maximize throughput of NRT traffics with satisfying QoS requirements of RT traffics. The simulation study shows that the proposed UEPS algorithm is able to give better throughput performance than existing wireless packet scheduling algorithms such as proportional fair (PF) and modified-largest weighted delay first (M-LWDF) while satisfying QoS requirements of RT traffics such as the average delay and the packet loss rate under various traffic loads.

Keywords: wireless networks, packet scheduling, OFDMA, downlink, time-utility function, urgency, efficiency, quality of service (QoS)

1. Introduction

Scheduling discipline in packet switching is a rule which determines packet processing order at every scheduling instant (i.e., each timeslot) based on status of the server and packets waiting at buffers to maximize system performance. However, packet scheduling methods used in wired realm can not be applied directly to wireless environment because of its inherent characteristics and restrictions including location dependent error, time-varying channel capacity stems from unstable wireless channel status, and bursty error caused from these factors.

In this paper, we propose an efficient wireless downlink packet scheduling algorithm for OFDMA system which is a strong candidate wireless system for the next generation mobile communications.

This research was partially supported by Chung-Ang University Research Grant in 2004.
† Corresponding author : Professor Seungwan Ryu, department of Information Systems, Chung-Ang University, Ansung, 456-756/An invited researcher at Broadband Mobile MAC Research Team, ETRI, Daejon, 305-350, E-mail : rush2384@cau.ac.kr

There are many existing packet scheduling algorithms such as Proportional Fair (PF) (R. Padovani *et al.*, 2000) and Modified-Largest weighted delay first (M-LWDF) (K. Ramanan *et al.*, 2001) designed mainly to support non-real-time (NRT) data services in CDMA-1x-EVDO (HDR) system, a third generation mobile communication system. In contrast, we propose a wireless packet scheduling algorithm that can support real-time (RT) and NRT data traffics at the same time.

In wireless communications the inherent characteristics and transmission requirements of RT traffics are differ from those of NRT data traffic. In general, RT traffics such as voice and video streaming traffics require a low and bounded delay but can tolerate some information loss.

In contrast, NRT data traffics require low information loss but the delay require- ments are less stringent compared to the RT traffics. Because of such characteristics the RT traffic is often called a delay-sensitive traffic and the NRT traffic is called a loss-sensitive traffic.

In order to schedule packets of RT and NRT traffics at the same time, the proposed algorithm uses two scheduling factors, the urgency of scheduling and the efficiency of radio resource usage, by taking not only the inherent characteristics and restrictions of wireless environment but also QoS requirements of each traffic into account for. In the proposed utility and efficiency based packet scheduling (UEPS) algorithm, the time-utility function is used to represent the urgency factor while the channel state is used to indicate the efficiency of usage of the radio resource. The idea behind the UEPS algorithm is to maximize throughput of NRT traffics as long as QoS requirements of RT traffics such as the packet delay and the loss rate requirements are satisfied. Then the UEPS algorithm transmits packets of RT and NRT traffics based on their scheduling priorities obtained from the urgency and the efficiency factors.

This article is organized as follows. In the next section, we briefly survey existing wireless packet scheduling algorithms. In section 3, we introduce the OFDMA wireless system model used in this study. In section 4, we propose the UEPS algorithm that is able to support RT and NRT traffics at the same time. In section 5, we evaluate performance of the UEPS algorithm and compare it with existing algorithms via simulation study. Finally, a summary of this study is given followed by further study issues.

2. Related Works

In this section, we introduce two existing representative wireless packet scheduling algorithms, Proportional Fair (PF) (R. Padovani *et al.*, 2000) and M-LWDF(K. Ramanan *et al.*, 2001), mainly designed to support data services in CDMA-1x-EVDO (HDR) system. The CDMA-1x-EVDO (HDR) system is a wireless system of the third generation partnership project 2 (3GPP2) that is able to NRT data services as well as traditional voice telephony services. For NRT data services, the scheduler selects only one channel among many channels held in between a base station and mobile users in each timeslot. The design goal of scheduling algorithms in this system is to maximize throughput of NRT data services.

Design objective of PF algorithm is to maximize long-term throughput of a user equipment (UE) whose current channel status (i.e., achievable data rate) is better compare to the average throughput. Suppose that $R_i(t)$ and $T_i(t)$ are the current achievable data rate and the estimate of average throughput of user i at timeslot t, where $i \in I = \{1, ..., M\}$. Then the PF algorithm works as follows:

- Scheduling: The user with the highest ratio of $R_i(t)/T_i(t)$ among users will receive transmission from a base station (BS) at each scheduling time. Ties are broken randomly.
- Update the average throughput of each user i,

$$T_i(t+1) = (1-1/t_c)T_i(t) + (1/t_c)R_i(t)\Delta_i \quad (1)$$

where $\Delta_i = 1$ if user i is chosen to transmit, otherwise $\Delta_i = 0$ and t_c is a low pass filtering parameter related to the maximum amount of time (i.e., the number of time-slots) for which an individual user can be starved, and $t_c = 1000$ is recommended to use.

Since the PF algorithm is designed to support only data services in CDMA-1x-EVDO system, it can not support RT services such as voice and RT-video streaming services.

M-LWDF algorithm was proposed to support not only NRT data services but also almost realtime services such as video streaming services in CDMA-1x-EVDO (HDR) system (K. Ramanan *et al.*, 2001). Design objective is to maintain delay of each traffic smaller than a predefined threshold value with probability. Delay and throughput requirements are $\Pr W_i > \tau_i \le \delta_i$ and $T_i > t_i$ respectively, where W_i is the head-of-line (HOL) delay, τ_i is the maximum allowable delay threshold, δ_i is a maximum allowable probability of exceeding τ_i, and T_i is a predefined minimum throughput threshold. In each time-slot t, a user i^* is selected according to the scheduling priority as follow

$$i^* = \arg\max_{i \in I} \gamma_i W_i(t) R_i(t) \tag{2}$$

where $\gamma_i = a_i / \overline{R}_i(t)$ is an arbitrary constant, $a_i = -(\log \delta_i)/\tau_i$, and $\overline{R}_i(t)$ is the average channel rate with respect to flow i. By setting an appropriate value to each parameter γ_i, the delay requirement can be satisfied. However, it is difficult to find optimal γ_i value for each traffic class $i \in I$.

3. System Model

3.1 An Overview of OFDMA Wireless System

OFDMA, also referred to as multiuser-OFDM, is being considered as a modulation and multiple access method for the next generation (or 4th generation) wireless networks. OFDMA is an extension of Orthogonal Frequency Division Multiplexing(OFDM), which is currently the modulation of choice for high speed data access systems such as IEEE 802.11a/g wireless LAN (WiFi) and IEEE 802.16a/ d/e wireless broadband access systems (WiMAX). In current OFDM systems, only a single user can transmit on all of the subcarriers at any given time, and time division or frequency division multiple access is employed to support multiple users. The major drawback of this static multiple access scheme is the fact that different users see the wireless channel differently is not being utilized. OFDMA, on the other hand, allows multiple users to transmit simultaneously on the different subcarriers per OFDM symbol.

In this study we consider an OFDMA system with 20MHz of bandwidth. The OFDMA system is being developed at ETRI as a strong candidate wireless system for the beyond 3G or 4G mobile system. In this paper, we propose a wireless downlink packet scheduling algorithm that will be implemented in the medium access control (MAC) layer at a base station. It is assumed that there are 1,536 subcarriers, and all subcarriers are shared by all users in a cell in terms of sub-channels, a subset of the subcarriers. We assume that there are 12 sub-channels and each sub-channel is a group of 128 subcarriers. It is also assumed that all subcarriers are used for data transmission for simplification, and subcarriers in each sub-channel are selected by a pre-determined random pattern. The modulation and coding scheme is determined by the prescribed adaptive modulation code (AMC) table based on the instantaneous signal-interference-ratio (SIR) of each sub-channel. A summary of system parameters is shown in <Table 1>.

Table 1. A summary of system parameters

Parameters	Value
System	OFDMA
Downlink channel bandwidth	20 MHz
OFDM symbol duration	100µs
Total number of subcarriers	1,536
Number of subcarriers per sub-channel	128
Number of sub-channels	12
Frame period	20 ms
Slot period	1 ms

3.2 Structure of the Packet Scheduler in a Base Station

The proposed packet scheduling system in a base station (BS) consists of three blocks: a packet classifier (PC), a buffer management block (BMB), and a packet scheduler (PS) as shown in <Figure

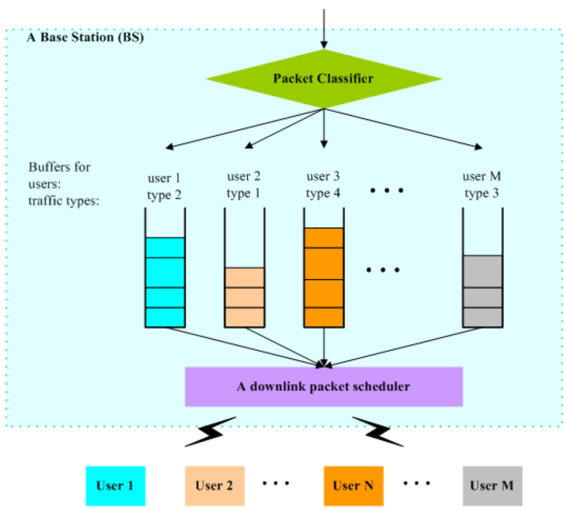

Figure 1. The structure of the proposed packet scheduler.

1>. The packet classifier classifies incoming packets according to their types and QoS profiles, and sends them to buffers in BMB. The BMB maintains QoS statistics such as the arrival time and delay deadline of each packet, the number of packets, and the head-of-line (HOL) delay in each buffer. Finally, the PS transmits packets to users according to the scheduling priority obtained using QoS statistics and channel status reported by user equipments.

4. The Urgency and Efficiency based Packet Scheduling (UEPS) Algorithm

Two scheduling factors, the urgency of scheduling and the efficiency of radio resource usage, are used at the same time to schedule RT and NRT traffic packets in the proposed wireless downlink packet scheduling for OFDMA system. The time-utility function (TUF) is used to represent the urgency of scheduling while the channel state is used to indicate efficiency of radio resource usage.

4.1 The Urgency of Scheduling

Since utility is decreasing in delay, i.e., the longer the delay, the lower the utility, urgency of scheduling can be expressed as a function in delay (J. Wang *et al.*, 2004). In this paper, a time-utility function(TUF) is used to indicate the urgency of scheduling. The time-utility means degree of satisfaction that an user experiences due to transmission delay of a packet. Generic TUFs of RT and NRT traffics are shown in <Figure 2>. A TUF of a delay-sensitive RT traffic can be expressed as a hard and discontinuous function in delay since utility of an RT traffic drops abruptly when its delay reaches deadline. In other words, utility of an RT traffic is the same value as long as the packet is transmitted within its deadline, but becomes 0 when the packet is transmitted after the deadline. On the other hand, the TUF of an NRT traffic is a continuously decreasing function in delay, in that utility of an NRT traffic decreases slowly as delay increases.

Figure 2. Generic time-utility functions of RT (left) and NRT (middle and right) traffics.

4.1.1 Relaxation of the hard TUF of RT traffics

Since the TUF of an RT traffic is a hard and discontinuous function in delay, the unit change of the utility, $|U'_{RT}(t)|$, can not be obtained at its delay deadline. In order to address this problem, the TUF of an RT traffic can be relaxed to a continuous z-shaped function which has similar properties of the original hard discontinuous function. An example of z-shaped function relaxation of TUF of RT traffic is shown in <Figure 3>. In the relaxed z-shaped function, the delay deadline is relaxed from a time instant to a time interval $[M_1, M_2]$, where $M_1 <$ deadline $\leq M_2$. In the OFDMA system shown in <Table 1>, the length of the time interval is 1ms, where M_1=deadline-1 ms and M_2=deadline. Then the first derivative of the utility function ($|U'_i(t)|$) can be obtained in this time interval since the utility drops continuously and linearly during this time interval $[M_1, M_2]$. This time interval is called the marginal scheduling time interval(MSTI).

A z-shaped function relaxation of TUF of an RT traffic can be easily achieved analytically using an s-shaped function having close relation with z-shaped function. A good example of such an s-shaped function is the sigmoid function, $f_{sigmoid}(t, a, c) = 1/(1+e^{-a(t-c)})$, where a and c are parameters that determine slope and location of the inflection point of the function. Then, by using the sigmoid function, a z-shaped function $U_{RT}(t)$ and its unit change at time t, $|U'_{RT}(t)|$, are obtained as follow.

$$U'_{RT}(t) = 1 - f_{sigmoid}(t, a, c) = \frac{e^{-a(t-c)}}{1+e^{-a(t-c)}} \quad (3)$$

$$|U'_{RT}(t)| = -ae^{-a(t-c)}/(1+e^{-a(t-c)})^2 \quad (4)$$

In this relaxation example, the delay deadline of the TUF of an RT traffic becomes the inflection point of the relaxed z-shape function, c. The unit change of utility of an RT traffic at time c, i.e., $|U'_{RT}(t=c)| = a/4$, and this value represents the urgency level of an HOL packet of an RT traffic at its deadline. If there are many different types of RT traffics such as voice and real-time video streaming traffics, those traffics can be differentiated by assigning different values for each RT traffic type. For example, assigning a higher scheduling priority to a voice traffic over an RT video traffic can be easily achieved by setting the sloping parameter (a) of the voice traffic at the inflection point larger than that of an RT video traffic, i.e., $a_{voice} > a_{video}$.

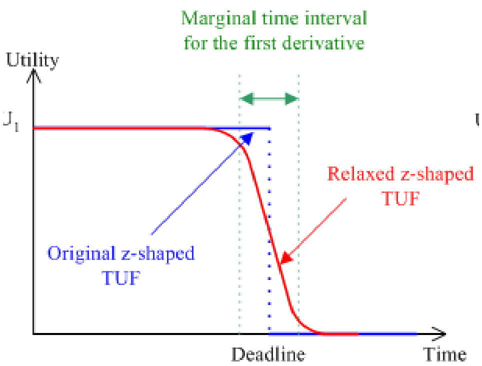

Figure 3. An example of z-shaped relaxation of a TUF of an RT traffic.

4.1.2 TUFs of NRT traffics

Since TUFs of NRT traffics are monotonic decreasing functions in time (delay), an analytic model can be easily obtained using related monotonic increasing functions. For example, a truncated exponential function, $f(a_i, t, D_i) = \exp(a_i t)$, can be a candidate TUFs for NRT traffics, where a_i is an arbitrary parameter and $0 \le t \le D_i$ is delay deadline of NRT traffic i. Consider a normalized truncated exponential function, $f(a_i, t, D_i) = \exp(a_i t)/\exp(D_i)$, which is normalized by the maximum time, D_i, so that it can have smoother slope. Then, a possible TUF of an NRT traffic i is $f_{NRT_i}(t) = 1 - f(a_i, t, D_i) = 1 - \exp(a_i t)/\exp(D_i)$, and the urgency is $|U'_{NRT_i}(t)| = a_i \exp(a_i t)/\exp(D_i)$. The urgency factor of NRT traffics can be simply differentiated by assigning appropriate parameters a_i values for each NRT traffic type. Suppose that there are two NRT traffics, i and j, for instance, and i should have higher urgency value over j when $R_i(t) = R_j(t)$, $\forall i, j$. Then a necessary condition for differentiating those two NRT traffics is.

$$|U'_{NRT_i}(t)| = \frac{a_i \exp(a_i t)}{\exp(D_i)} > |U'_{NRT_j}(t)| = \frac{a_j \exp(a_j t)}{\exp(D_j)} \quad (5)$$

4.1.3 Urgency factors among RT and NRT traffics

The urgency factor of each traffic class, $|U'_i(t)|$, is used to determine the scheduling priority among HOL packets of different traffic classes waiting in the BMB. Then assignment of the urgency factor among traffic classes is dependent on designer's preference. For example, if a designer wants to give a higher scheduling priority to an RT traffic over other traffic classes, he or she just needs to set $|U'_{RT-voice}(t)| > |U'_i(t)|$, where i=RT video, or NRT data traffics. In this study, we set the urgency factors of all traffic classes in the order of RT voice, RT video, NRT traffics by assigning urgency factors as follow.

$$\begin{aligned} &|U'_{RT-voice}(t)| > |U'_{RT-video}(t)| \\ &> |U'_{NRT-1}(t)| > |U'_{NRT-2}(t)| \end{aligned} \quad (6)$$

4.2 Efficiency of Radio Resource Usage

Efficiency in wireless communications is related to usage of scare radio resources, i.e., the limited number of radio channels or limited bandwidth. Thus the channel state of available radio channels can be used as an efficiency indicator. For example, the current channel state ($R_i(t)$), the average channel state ($\overline{R}_i(t)$) or the ratio of the current channel state to the average ($R_i(t)/\overline{R}_i(t)$) can be used as an efficiency indicator. In this study, a moving average of the channel state of each user $i \in I$, $\overline{R}_i(t+1) = (1-1/W)\overline{R}_i(t) + (1/W)R_i(t)$, is used for the average channel state, where W is the window size used in calculation of the moving average of the channel state. Note that $\overline{R}_i(t)$ used in our paper is different from the average throughput of user i, $T_i(t)$, in past t_c timeslots (1) used in PF algorithm (R. Padovani *et al.*, 2000) . Therefore the higher the user's instantaneous channel quality relative to its average value, the higher the chance of a user to transmit data with a rate near to its peak value.

4.3 The Proposed UEPS Algorithm

We propose an urgency and efficiency based wireless packet scheduling (UEPS) algorithm that takes two scheduling factors, the urgency of scheduling and the efficiency of scarce radio resource usage, into account for determining scheduling priorities among different traffic classes. In detail, the first derivative of TUF of traffics, $|U'_i(t)|$, indicates the urgency of scheduling while the relative status of the current channel to the average value, $R_i(t)/\overline{R}_i(t)$, indicates the efficiency of radio resource usage. The scheduling priority function of each user is a product of the urgency factor, $|U'_i(t)|$, and the efficiency factor, $R_i(t)/\overline{R}_i(t)$. Then the scheduler selects 12 users according to their scheduling priority as follow.

$$i^* = \arg\max_{i \in I} |U_i'(t)| \frac{R_i(t)}{\overline{R}_i(t)} \quad (7)$$

The proposed algorithm is designed to maximize throughput of NRT traffics with satisfying QoS requirements of RT traffics. In other words, since the downlink between a BS and user equipments (UEs) is the last link to destinations (UEs), the UEPS algorithm tries to maximize throughput of NRT data services while supporting maximum allowable QoS requirements of RT traffics such as the maximum allowable delay and loss rates. In detail, when there is enough delay margin, the length of time from the current scheduling time to the delay deadline, for RT traffics assuming that channel states of all users are the same, the UEPS scheduler transmits NRT traffics first to maximize throughput of the system.

For example, suppose that an RT traffic which has 40ms of a delay tolerance arrived at time t=5ms. Then, at time t=10ms, the delay margin of such traffic is 35ms because its delay deadline is 5+40 =45ms. In contrast, the scheduler gives an RT traffic a higher scheduling priority over NRT traffics when its delay approaches deadline, i.e., its delay margin approaches 0.

4.3.1 Implementation issues

The scheduler can transmit up to N users' data using N subchannels in each timeslot. In this study, we assume there are 12 subchannels for data transmission as shown in <Table 1>. Therefore the UEPS scheduler transmits up to 12 users' data including RT and NRT traffics in each timeslot. This is the main difference of the UEPS algorithm designed for OFDMA system from the existing wireless packet scheduling algorithms designed for CDMA2000-EVDO (HDR) system where only one user is selected in each timeslot for data transmission.

However, when there are urgent RT traffics, i.e., RT traffics that already reach their delay deadline, to be transmitted more than 12 users in a timeslot due to heavy RT traffic load, some of those traffic can not be transmitted. In this case the corresponding HOL packets of RT traffics will be lost since their urgency factor, $|U_i'(t)|$, will drop to 0 from the next timeslot. Such forced packet lost of RT traffics can be reduced by consideration of the urgency factor several timeslots earlier than their delay deadline. For example, if the UEPS scheduler begins to considering the urgency factor of a RT traffic n timeslots ahead from its delay deadline as expressed in below equation (8), the forced packet lost can be reduced at the expense of throughput reduction of NRT traffics. <Figure 5> shows the concept of introduction of the marginal scheduling time interval (MSTI) to the UEPS algorithm to prevent forced packet lost due to heavy RT traffic load.

$$|U_i'(t)| = \max\{|U_i'(t)|, |U_i'(t-1)|, \cdots, |U_i'(t+(n-1))|\} \quad (8)$$

where n is an arbitrary integer value. Note that this is the same as extension of the length of the marginal time interval $[M_1, M_2]$ in <Figure 3> from 1 unit scheduling timeslot (i.e., 1ms) to n timeslots (i.e., n ms). In other words, scheduling of a HOL RT traffic packet begins at time after $t_d - t_0 - n + 1$ms from its arrival time (t_0) as shown in <Figure 5>.

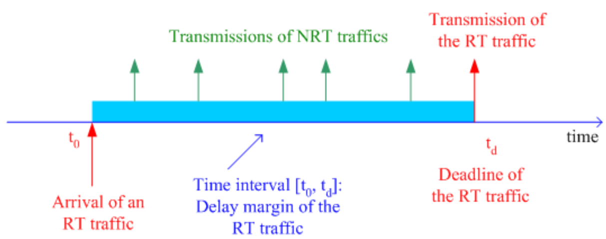

Figure 4. The concept of UEPS algorithm for scheduling traffics consisting of RT and NRT traffics.

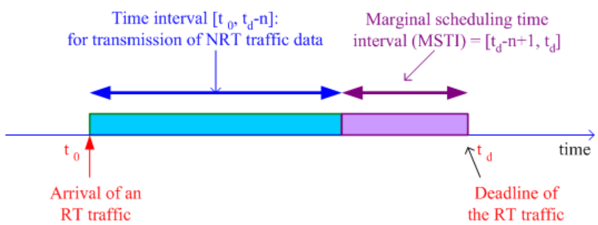

Figure 5. The concept of the marginal scheduling time interval (MSTI) in UEPS algorithm.

5. Performance Evaluation

5.1 Simulation Model and Traffic Parameters

5.1.1 Traffic types

In the simulation study it is assumed that there are four different traffic types, and each user generates one of four traffics. Four types of services are:

- **Realtime (RT) voice:** RT voice is designed to support RT traffics such as voice on IP (VoIP) that periodically generate packets of fixed size. Assuming that silence suppression is used, voice traffic is modeled by a 2-state Markov (ON/OFF) model. The length of the ON and OFF periods follow the exponential distribution with mean of one second and 1.35 seconds respectively. <Figure 6> shows a generic real-time VoIP traffic model.

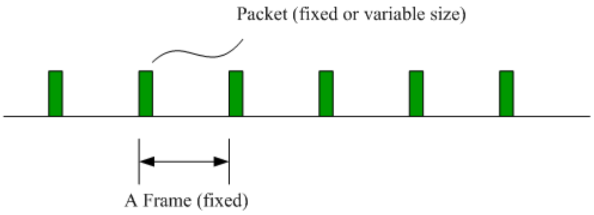

Figure 6. A real-time VoIP traffic model.

- **Realtime (RT) video:** RT video supports RT traffics such as RT video steaming service that periodically generate packets of variable sizes. We uses 3GPP streaming video traffic for this type of traffic (3GPP2/TSG-C.R1002). A traffic model and characteristics of an RT video traffic are shown in <Figure 7> and <Table 2> respectively.

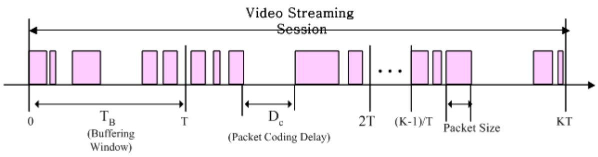

Figure 7. A real-time video traffic model.

Table 2. A summary of characteristics of a real-time video traffic model.

Component	Distribution	Parameters
Main object size (S_M)	Truncated Normal	Mean: 10710 bytes, STD.: 25032 bytes Min.:100 bytes, Max: 2Mbytes
Embedded Object size (S_E)	Truncated Normal	Mean: 7758 bytes, STD.: 12168 bytes Min.:50 bytes, Max: 2Mbytes
Number of embedded objects/page (N_d)	Truncated Pareto	Mean: 5.64, STD.: 53
Reading time(D_{pc})	Exponential	Mean=30 seconds
Parsing time (T_p)	Exponential	Mean=0.13 seconds

- **Non-real time (NRT) Data Service type 1:** NRT data service type 1 supports NRT data traffics such as FTP and web browsing that require wide bandwidth and variable sized bursty data. In this study, we use the web traffic (WWW) model proposed to have a session consisting of several web pages which contains multiple packets or datagrams as shown in <Figure 8>. Characteristics of WWW traffic model are summarized in <Table 3>.

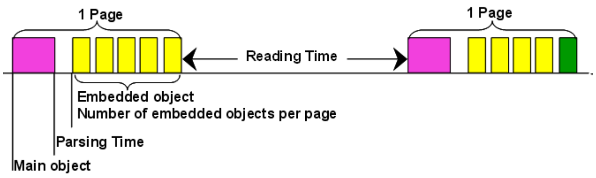

Figure 8. A web (WWW) traffic model.

NRT Data Service type 2 (Best Effort (BE)): BE supports NRT data traffics such as emailing services. We assume that messages arrival to the mailboxes is modeled by Poisson process.

5.1.2 System parameters

We consider a hexagonal cell structure consisting of a reference cell and 6 surrounding cells with 1 km of radius. We assume that all cells use omni-directional antenna. Mobile stations are uniformly distributed in a cell, and move with velocity of uniform distribution in a random direction. The BS transmission power is 12W which evenly distributed to all 12 sub-channels. We use 3GPP path loss model (3GPP, 2000). A summary of simulation parameters for system model is shown in <Table 4>.

Table 3. A summary of characteristics of WWW traffic model.

Characteristics	Distribution	Parameters
Inter-arrival time between frames	Deterministic	100ms
Number of packets/frame	Deterministic	8
Packet size	Truncated Pareto (Mean:50, Max:125 (bytes))	K=20 bytes, α=1.2
Inter-arrival time between packets	Truncated Pareto (Mean:6, Max:12.5 (ms))	K=2.5ms, α=1.2

5.2 Performance Metrics

We evaluate and compare performance of the proposed UEPS, PF and M-LWDF algorithms in terms of three different performance metrics such as the packet loss rate, the average packet delay, and the average throughput via simulation study.

For the delay-sensitive RT traffics, the average packet delay is mainly used to evaluate performance. Although RT traffic is tolerant to packet loss, it has maximum allowable packet loss rate. For example, packet loss rate of the RT voice should be less than 3% (T. Janevski, 2003). Therefore, performance of RT traffics is also evaluated in terms of the packet loss rate. For RT traffics, performance of UEPS algorithm is compared with that of M-LWDF algorithm in terms of the average delay and the packet loss rate. QoS performance requirements of RT voice and video traffics (T. Janevski, 2003) are

- RT Voice: delay < 40ms, loss rate < 3%
- RT Video: delay <150 ms, loss rate < 1%

For the loss-sensitive NRT traffics, the average throughput is used to evaluate performance of the UEPS, PF and M-LWDF algorithms. Parameters of PF and M-LWDF algorithms are

- PF: t_c=1000,
- M-LWDF: $W_{\max}$=40ms, δ=0.03 for RT voice and $W_{\max}$=150ms, δ=0.01 for RT video

Table 4. A summary of simulation parameters for system model.

Parameters	Value
User distribution	Uniform
Number of cells/layout	7/hexagonal
Beam pattern	Omni-directional
Radius of a cell	1km
Velocity of a MS	Uniform: 3 ~ 100 km/second
Path loss model	$L=128.+37.6\log_{10}R$
BS total Tx power	12 W

5.3 Performance Evaluation

In order to evaluate performance of UEPS algorithm and compare it with those of PF and M-LWDF algorithms, various traffic loads are generated from the light to the heavy traffic load. Since the proposed UEPS scheduler selects 12 users at each timeslot, the number of users arrived in each timeslot is used as the offered traffic load. In the simulation study, the offered traffic load (λ) distributes (2, 20) users/ timeslot, i.e., λ=2/12~ 20/12=0.167~1.67.

5.3.1 Performance of RT traffics

As a preliminary research work on performance evaluation of the UEPS, PF and M-LWDF algorithms, we first determine the length of the marginal scheduling time interval (MSTI), n, for RT traffics in UEPS algorithm. In this paper, because of page limitation we only show how to determine n_{voice} for RT voice traffic. We evaluate packet loss rates of RT voice and video traffics extensively with different n values. We give a selected simulation results in which there MSTI values, i.e., n_{voice}=10, 20 and 30, for RT voice traffic, while the MSTI for RT-video (n_{video}) is fixed to 30. As shown in <Figure 9>, when n_{voice}=10 the packet loss rate of the voice traffic stays below the maximum allowable packet loss rate (3%) with giving a lower packet loss rate of the video traffic than other n_{voice} values. In particular, the packet loss rate of the RT video traffic stays below the maximum allowable value (1%) up to 16 users/ slot when n_{voice}=10. In addition, UEPS algorithm results in better performance with n_{voice}= 10 and n_{video}=30 than M-LWDF algorithm in terms of packet loss rates of RT voice and video traffics.

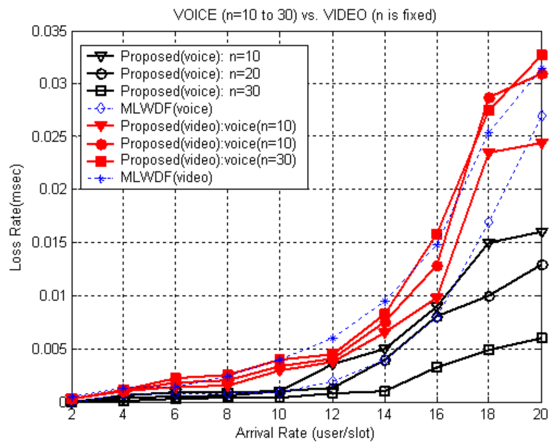

Figure 9. Average loss rates of RT traffics when for different marginal time intervals.

<Figure 10> shows average delays of RT voice and video traffics under various traffic loads. From the above preliminary experimental results on determining the MSTI value for RT traffics, we select two MSTI values, n_{voice}=10 and n_{video}=30. This means that scheduling of HOL packets of RT voice and video traffics begins after 30ms and 140ms from their arrival time, and thus minimum delays of RT voice and video traffics are 30ms and 140ms respectively. With these MSTI values the UEPS algorithm gives average delays of RT voice

and video traffics lower than the maximum allowable delay requirements, i.e., <40 ms and <150 ms, but higher than those of M-LWDF under all traffic loads. <Figure 11> shows packet loss rates of UEPS and M-LWDF for the RT voice and video traffics under various traffic loads. For the RT voice traffic, the packet loss rate of UEPS under all traffic loads is acceptable, i.e., less than 3%, whereas that of M-LWDF grows rapidly as the traffic load increases, and finally become larger than the maximum allowable loss rate requirement (3%). For RT video traffic, the packet loss rate of UEPS is acceptable until the traffic load reaches up to 14 users/timeslot (i.e., λ=1.167). However, the packet loss rate of M-LWDF is higher than that of UEPS under all traffic loads.

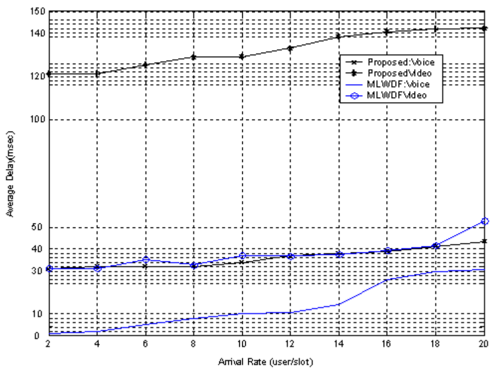

Figure 10. Average delays of RT voice and Video traffics under different traffic loads.

5.3.2 Throughput of NRT traffics

To evaluate performance of the UEPS algorithm and compare it with PF and M-LWDF algorithm we generate two different traffic environments. First, we evaluate throughput of UEPS, PF and M-LWDF algorithms under a traffic environment where only two NRT traffics, WWW and email traffics, are generated. Throughput performance of UEPS, PF and M-LWDF algorithms in this traffic environment under various traffic loads are shown in <Figure 12>. For WWW traffic, throughput of UEPS is higher than those of PF and M-LWDF under all traffic loads except the light traffic load. In particular, the higher the traffic load, the higher the throughput of UEPS than those of other algorithms. For email traffic, throughputs of UEPS and PF are almost the same, and higher than that of M-LWDF under all traffic loads.

Next, we evaluate throughput of UEPS, PF and M-LWDF algorithms for NRT traffics under a traffic environment where four traffics, RT voice, RT video, WWW and email traffics, are generated. Throughput performance of UEPS, PF and M-LWDF algorithms in this traffic environment under various traffic loads is shown in <Figure 13>. For NRT traffics, UEPS algorithm gives lower throughput than PF algorithm in this traffic environment. However, for both of WWW and email traffics, the UEPS algorithm shows higher throughput performance than M-LWDF algorithm under most of traffic load with supporting RT and NRT traffics at the same time.

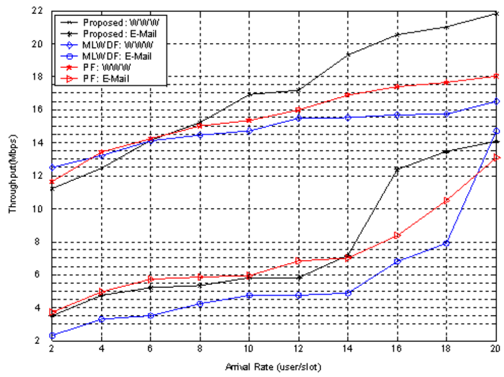

Figure 11. Packet loss rates of RT voice and Video traffics under different traffic loads.

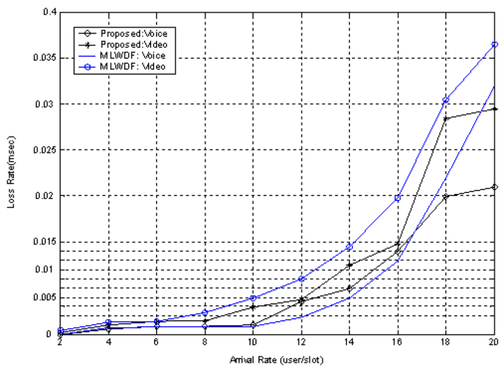

Figure 12. Throughput of NRT traffics (WWW and e-mail) under different traffic loads.

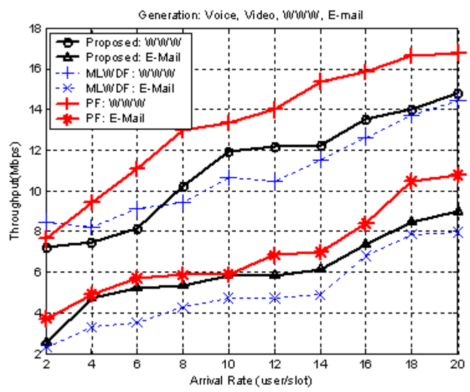

Figure 13. Throughput of NRT traffics (WWW and e-mail) under different traffic loads.

6. Conclusions and Further Study Issues

In this paper, we deigned a novel wireless packet scheduling algorithm, the UEPS algorithm, that is able to schedule RT and NRT traffics simultaneously by taking urgency of scheduling and efficiency of radio resource usage into account for. The UEPS algorithm uses the time-utility function as an urgency factor and the relative status of the current channel to the average one as an efficiency factor. The main design goal of the UEPS algorithm is to maximize throughput of NRT traffics with satisfying QoS requirements of RT traffics. Simulation study shows that the proposed UEPS algorithm shows better throughput performance than PF and M-LWDF with satisfying QoS requirements of RT traffics under various traffic loads.

However, there are further study issues. First, it is necessary to study the UEPS algorithm under more general wireless packet scheduling situations. One possible case is that a user requests download of RT traffics and NRT traffics to a BS at the same time. Now we are exploiting application of the UEPS algorithm in such a general scheduling situation. For practical implementation, the hard and discontinuous TUFs of RT traffics are relaxed to a continuous z-shaped functions as shown in <Figure 3>. Then we introduce the concept of MSTI to obtain the first derivative of the utility function. We extend the length of MSTI from an unit timeslot (i.e., 1 ms) to a longer ones (i.e., 10ms for RT voice and 30 ms for RT video) to reduce the forced packet loss of RT traffics stems from heavy RT traffic load. However, in order to meet QoS requirements of RT traffics adaptively to the various traffic situations such as different traffic load levels and different traffic mix between RT and NRT traffics, it is also needed to adjust the length of MSTI adaptively to various traffic situations. We are also trying to find how to set the length of $[M_1, M_2]$ for each RT voice and video traffics under various traffic situations.

References

R. Padovani A. Jalali and R. Pankaj.(2000), Data throughput of cdma hdr a high efficiency-high data rate personal communication wireless system. In Proceedings of VTC2000-Spring, 1854-1858.

K. Ramanan A. Stolyar P.(2001), Whiting M. Andrews, K. Kumaran and R. Vijayakumar. Providing quality of service over a shared wireless link. IEEE Communications magazine, 39(2), 150-154.

J. Wang *et al*.(2004), Time-utility function-driven switched ethernet: Packet scheduling algorithm, implementation, and feasibility analysis. IEEE Trans. Parallel and Distributed Systems, 15(2), 119-133.

3GPP. Physical layer aspects of UTRA high speed downlink packet access (release 2000).

T. Janevski. (2003), Traffic Analysis and Design of Wireless IP Networks. Norwood, MA, Artech House.

류 승 완
고려대학교 산업공학과 학사
고려대학교 대학원 산업시스템공학과 석사 (Operations Research 전공)
뉴욕주립대 (SUNY at Buffalo) 박사 (Operations Research 전공)
현재: 중앙대학교 정보시스템학과 (조교수) / 한국전자통신연구원 광대역무선 MAC 연구팀 (초빙연구원)
관심분야: 차세대 무선 MAC protocols, 3세대 이후(beyond 3G) 또는 4세대 이동통신 서비스 및 시스템 architecture

서 현 화
고려대학교 전산과학과 학사
고려대학교 전산과학과 석사
현재: 한국전자통신연구원 광대역무선 MAC 연구팀
관심분야: 차세대 무선 LAN MAC Protocols, Random access protocols


정 수 정
고려대학교 전자공학과 학사
한국과학기술원 (KAIST) 전자공학과 석사
현재: 한국전자통신연구원 광대역무선 MAC 연구팀
관심분야: 통신이론, 이동 및 무선 통신


임 순 용
인하대학교 전자공학과 학사
인하대학교 전자공학과 석사
현재: 한국전자통신연구원 광대역무선 MAC 연구팀
관심분야: 고속 무선전송 기술


박 세 권
서울대학교 산업공학과 학사
서울대학교 산업공학과 석사
Texas A&M University 산업공학과 박사
현재: 중앙대학교 정보시스템학과 (교수/학과장)
관심분야: 통신 네트워크 설계 및 성능 분석, Software engineering