이동무선 네트워크에서 하향링크 전송을 위한 시간비율 기반 스케쥴링 기법에 관한 연구*

†백천현** · Samit Soni***

A Time-fraction Based Scheduling Method for Downlink Transmissions in Wireless Network*

Chunhyun Paik** · Samit Soni***

■ Abstract ■

This paper deals with a mathematical approach for finding the time-fractions for the time-fraction based scheduling method in multimedia wireless networks. By introducing a constraint that regulates the performance fairness amongst users, we present a systematic method for harmonizing both the system and user performance. Numerical results show that the time-fraction based scheduling method reinforced with our scheme is very effective especially in multimedia environments.

Keywords : Transmission Scheduling, Time-fraction Assignment, Wireless Communications

1. Introduction

Multimedia data services are expected to grow rapidly over the next few years and are likely to become a major source of traffic in mobile communication systems. Given the fact that radio resources are scarce and limited, a major challenge of transmission scheduling is to ach-

논문접수일 : 2007년 01월 11일 논문게재확정일 : 2007년 05월 08일
* This work was supported by Korea Research Foundation Grant (KRF-2004-041-B00142).
** 동의대학교 산업경영공학과
*** Infosys Consulting, USA.
† 교신저자

ieve a high usage efficiency of radio resources while simultaneously facilitating a mixture of diverse services with wide-ranging QoS (quality of service) requirements. One of the unique features of wireless systems is the time-varying channel capacity owing to the time-varying radio propagation environment. A significant increase in system throughput can be obtained in this propagation environment if, instead of allocating radio resources sequentially like in a round-robin system, the transmission scheduling is done based on channel conditions. Since in general, users with better channel conditions could be supported with higher data rates, transmission scheduling that is only based on the maximization of system throughput, may cause a monopolization of radio resources by those users.

Many studies in the literature have focused on the development of transmission scheduling methods (TSMs) for sharing the radio resources fairly while keeping the system throughput as high as possible. References [1~3] presented the Proportional Fairness (PF) transmission scheduling method for CDMA high data rate (HDR) systems [4, 5]. The PF method is very widely used owing to its simplicity and capability of providing an appealing trade-off between the system throughput and fairness in the radio resource utilization. Under the PF method, the radio resources are assigned based on the current and average channel conditions of users, not considering the diverse QoS levels of users. This would cause some difficulty in supporting a mixture of users having different QoS requirements and is considered as a major limitation of this method. To overcome this limitation, some variants of the PF method have been developed [6-9]. The authors of [10, 11] have compared the PF method with other traditional simple methods like round robin and maximum carrier to interference. The M-LWDF (modified largest weighted delay first) transmission scheduling method was proposed to guarantee a QoS requirement on delay constraint [12]. In reference [13] the authors suggest a transmission scheduling method maximizing the minimum weighted throughput of users.

Another representative transmission scheduling method worth noting in the literature is the one based on time-fractions assigned to each user [14~16]. Given a time-fraction for each user, Liu et al. [14, 15] and Marbach et al. [16] have proposed a time-fraction based transmission scheduling method (TF-TSM) for maximizing the system performance and monetary revenue, respectively. In TF-TSM, the time-fractions can be considered as the portion of resources entitled to the users. Liu et al. [14] have developed an efficient algorithm that optimally assigns radio resources to users such that the long-term proportion of radio resources assigned to each of users is matched to the time-fractions allocated to them.

Since a practical transmission scheduling method should exploit the short-term variations of channel conditions while maintaining some degree of long-term QoS requirements and fairness, the values of the time-fractions in TF-TSM should be carefully regulated according to the changes in performance and fairness. This means that the time-fraction assignments (TFAs) play a key role in success of TF-TSM. To the best of our knowledge, there is no other study that deals with a systematic way of time-fraction assignments other than some intuitive methods given in [14]. In this paper, we develop an efficient TF-TSM by integrating the method given in [14] with the new TFA scheme that we develop.

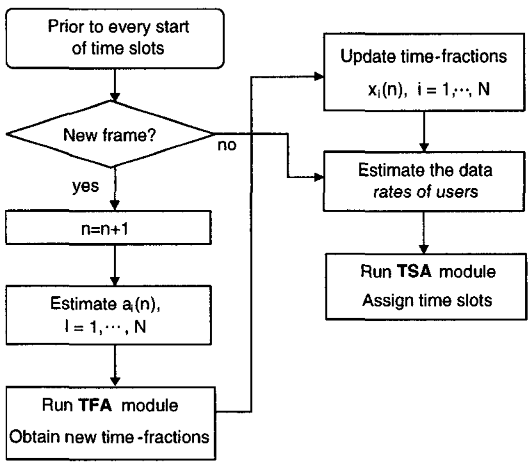
The rest of the paper is organized as follows. In Section 2, we first describe the procedure TF-TSM, and provide a mathematical model for TFA problem, along with its optimal solution. Some issues on TFA parameters are discussed in Section 3. In Section 4, we provide extensive computational experiments under simplified but realistic assumptions, to demonstrate the effectiveness of the TF-TSM with the proposed TFA method. Conclusions are made in Section 5.

2. TF-TSM and a Mathematical Approach for TFA

Consider a base station (BS) serving N users. These users may be classified into several different service classes based on the required performance criterion, specifically packet latency [1]. Assume that a packet scheduler resides in the BS and all users have data queued at the scheduler that needs to be transmitted to them at any given time. We further assume that time is partitioned into equal width intervals (Δ) called time slots. We consider downlink communications in which the BS can transmit to a single user during any time slot. Further, the data rates from the BS to the users may be different from each other and can vary at each time slot.

Our TF-TSM is composed of two sub-modules <Figure 1> : TSA (time slot assignment) and TFA (time-fraction assignment). Prior to starting every time slot, the possible data rate for each user is estimated based on the time-varying channel conditions. With the estimated data rates, the sub-module TSA allocates the next time slot to the user achieving the highest system performance (specifically, throughput). This is done while fulfilling the time-fraction constraint imposed on each of users. This time-fraction constraint for a user specifies that the proportion of the number of time slots assigned to the user during the TFA-cycle composing of f time slots. For the TSA in our TF-TSM, the existing TSA algorithm developed by Liu et al. [14] is adopted. It is assumed that f is sufficiently large so that the TSA algorithm converges. Refer to [14, 15] for the details of the sub-module TSA.

At the end of TFA-cycle, the sub-module TFA updates the current time-fraction values, taking into account the gaps between the QoS requirements of the users and the realization up to the current TFA-cycle. With these updated time-fractions, the sub-module TSA will allocate time slots to users for the next TFA-cycle.



〈Figure 1〉 Summary of the TF-TSM

By *performance fairness* we mean that each user should be guaranteed to get enough system resources for achieving their QoS requirements, which is different from the fairness used in [1, 3, 13] giving users an equal opportunity for getting system resources (*utilization fairness*). Note that if the performance requirements of users are similar to each other, the performance fairness could be achieved by considering only the utilization fairness among users.

Now, we deal with a mathematical approach of finding the time-fractions of users achieving the utmost efficiency of TF-TSM by providing a superior balance between performance fairness and utilization fairness. Let $a_i(n)$ be the estimated expected data rate of user i, $i=1,\cdots,N$, for the n^{th} TFA-cycle. With $L_i(n)$ denoting the expected average transmission delay per packet of user i during the n^{th} TFA-cycle, we have

$$L_i(n) = \frac{f \cdot \Delta}{1} \times \frac{l_i}{a_i(n) \cdot m_i(n) \cdot \Delta}$$
$$= \frac{l_i}{a_i(n)} \times \frac{f}{m_i(n)} = \frac{l_i}{a_i(n) \cdot x_i(n)} \quad (1)$$

where $m_i(n)$ and l_i denote the number of time slots assigned to user i for the n^{th} TFA-cycle and the packet length of user i, respectively. Let $x_i(n)$ $(=m_i(n)/f)$ be the time-fraction assigned to user i for the n^{th} TFA-cycle. The observed packet latency realization ($\widetilde{W}_i(n)$) over n TFA-cycles of user i are then defined by $\sum_{k=1}^{n} \widetilde{L}_i(n)/n$, where $\widetilde{L}_i(n)$ is the realized transmission delay for the n^{th} TFA-cycle.

Assuming that the packet lengths of all users are the same1), we introduce a constraint for performance (packet latency) fairness among users as

$$\frac{\max_i\{L_i(n)\}}{\min_i\{L_i(n)\}} \le \gamma(n) \Rightarrow$$
$$\frac{\max_i\{a_i(n) \cdot x_i(n)\}}{\min_i\{a_i(n) \cdot x_i(n)\}} \le \gamma(n) \quad (2)$$

where $\gamma(n)(\geq 1)$ is the control parameter for performance fairness for the n^{th} TFA-cycle.

Given $\gamma(n)$ and $a_i(n), i=1,\cdots,N$, the TFA problem (P) for the time-fractions of the n^{th} TFA-cycle is given as

$$\text{(P)} \quad \max \quad Z(n) = \sum_{i=1}^{N} a_i(n) x_i(n) \quad (3)$$
$$s.t. \quad \sum_{i=1}^{N} x_i(n) \le 1 \quad (4)$$
$$\frac{\max_i\{a_i(n)\, x_i(n)\}}{\min_i\{a_i(n)\, x_i(n)\}} \le \gamma(n) \quad (5)$$
$$x_i(n) \ge 0, 1, \cdots, N.$$

The objective of problem (P) is to find the time-fractions maximizing the system throughput while satisfying the performance fairness constraint (5). The current form of problem (P) is a nonlinear programming problem with nonlinear constraints. The problem can be solved using some state-of-the-art non linear solvers. However the solution time would not be adequate for problems demanding the almost-real time solution speeds, as in our TFA problem. Based on the special structure of the problem (P), we now derive an explicit form of the optimal solution. For compact exposition, in the remaining part of this section, the notation per-

1) In the real-life system, the users belonging to the different service classes usually have the dissimilar packet lengths as well as the diverse QoS target level [9, 17]. In this study, however, the packet lengths of all user are assumed to be equal in order to take a mathematical approach as in [18, 19], and its extension to the general model incorporating the diverse packet lengths will be left as the future study

taining TFA-cycle will be abbreviated if no confusion arises.

Theorem 1. The TFA problem (P) has the optimal solution of the form

$$x_i = \begin{cases} \frac{a_q}{a_i} x_q, & i \in m \\ \gamma \frac{a_q}{a_i} x_q, & i \in M \end{cases} \tag{6}$$

where $q = \arg_i \min\{a_i x_i\}$, $m = \{i : a_i x_i = a_q x_q\}$, $M = \{i : a_i x_i > a_q x_q\}$,

and $$x_q = \left[\sum_{i \in m} \frac{a_q}{a_i} + \gamma \sum_{i \in M} \frac{a_q}{a_i} \right]^{-1} \tag{7}$$

Proof : Consider a feasible solution $x = (x_1, \cdots, x_N)$of TFA problem (P). From constraint (5), $a_i x_i \le \gamma \cdot a_q x_q$, $i = 1, \cdots, N$ and $a_p x_p \le \gamma \cdot a_q x_q$ where $p = \arg_i \max\{a_i x_i\}$ In the case of $a_p x_p \le \gamma \cdot a_q x_q$, the current solution is not optimal because the objective function would improve by at least $(\gamma \cdot a_q x_q - a_p x_p)$ without violating the constraint (5) through adjusting the variables x_p and x_q such that $a_p x_p = \gamma \cdot a_q x_q$. This means the optimal solution should satisfy the following

$$a_p x_p = \gamma \cdot a_q x_q \tag{8}$$

Now, set $a_i x_i = \gamma a_q x_q$, $i \in M$ and note that the solution given in (6) satisfies constraints (4) if the variable x_q is given as (7) because

$$\sum_{i=1}^{n} x_i = \sum_{i \in m} x_i + \sum_{i \in M} x_i = \left[\sum_{i \in m} \frac{a_q}{a_i} + \gamma \sum_{i \in M} \frac{a_q}{a_i} \right] x_q = 1.$$

Moreover the solution gives the greatest value of objective function among the feasible solution satisfying equation (8) because the variables x_i, $i \in M$ take the maximum values from their possible values. This proves that the solution is indeed optimal. □

Remarks :

(1) To get an optimal solution, we need the set m. Assume that $a_1 \le a_2 \le \cdots \le a_N$ and L is the smallest number satisfying $Z_1 \le \cdots \le Z_L > Z_{L+1}$, where Z_l is the value of objective function when $m = \{1, 2, \cdots, l\}$. Then the set $m = \{1, \cdots, L\}$. Once the set m is given, we can get an optimal solution as follows : First, calculate x_i, $\forall i (\ne q)$ by setting $x_q = 1$ and using equation (6). By normalizing x_i, $\forall i$ such that $\sum_{i \in N} x_i = 1$ we obtain an optimal solution.

(2) Our TFA is appropriate for real-time applications because an optimal solution can be obtained through simple $(N-1)$ comparisons of the objective functions even in the worst cases.

(3) The optimal objective function (Z^*) has the simple form of $Z^* = (|m| + \gamma |M|) a_q x_q$.

We now provide some results for two extreme cases of the performance fairness constraints.

Lemma 1. If $\gamma = 1$ in the problem (P), the optimal solutions and objective function become

$$x_i = \frac{1/a_i}{\sum_{j=1}^{N} 1/a_j}, i = 1, \cdots, N, \quad Z^* = \frac{N}{\sum_{i=1}^{N} 1/a_i}.$$

And if $\gamma = \infty$, then $x = (x_1, \cdots, x_N) \in S$, $Z^* = a_\kappa$

where $\kappa = \arg\max_i \{a_i\}$, $S = \{(y_1, \cdots, y_N) \mid y_1 + \cdots + y_N = 1,\ a_1 y_1 + \cdots + a_N y_N = a_\kappa, y_i \ge 0, \forall i\}$

If $\gamma = \infty$, the constraint (5) becomes meaningless, making the problem (P) a well-known fractional knapsack problem [20]. The result of $\gamma = 1$ has been derived from the fact that the time-fraction assigned to each of users should be inversely proportional to their expected data rate so that the expected transmitted data rates of all users are equal.

3. Implementation Issues

The systematic way of updating control parameter for performance fairness and estimating the expected data rates of users for the forthcoming TFA-cycle, which need to be done every the end of a TFA-cycle, plays a key factor for successful implementation of our TF-TSM. We first consider the problem of updating the control parameter $\gamma(n)$. As the value of $\gamma(n)$ decreases, the allowable difference between the maximum and minimum amount of the expected transmitted data becomes smaller, thus strengthening the performance fairness among users. As an extreme case, when $\gamma(n)=1$, all the users must have the same expected transmitted data amount in order to meet constraint (5). And when $\gamma(n)=\infty$, constraint (5) becomes meaningless.

To keep the level of performance fairness within proper levels, the value of $\gamma(n)$ should be chosen carefully. In order to formalize the insights underlying $\gamma(n)$, we first define an unfairness measure index, $UM(n)$, for the n^{th} TFA-cycle as follows :

$$UM(n)=\sum_{i=1}^{N}\max(0,(\widetilde{W}_i(n)-\eta_i)/\eta_i) \qquad (9)$$

where η_i indicates the QoS target value on packet latency of user i.

From the above definition, as the fulfillment of the QoS requirements gets worse, the unfairness measure index increases. Thus, the unfairness measure index is useful for evaluating the level of performance fairness encountered by all users. In our TF-TSM, the time-fractions of users for the next TFA-cycle would be modified by changing the only control parameter $\gamma(n)$. Since the performance fairness constraint (5) simply regulates the ratio of maximum to minimum of expected transmitted data amount, the modification of $\gamma(n)$ does not always guarantee to change the individual time-fraction of users. It is then reasonable that the unfairness measure index $UM(n)$, based on which $\gamma(n)$ is determined, compromises the entire state of performance fairness for all users, not a part of users, by summing up all normalized differences between the target and realized packet latencies as in (9). This simplicity, though might cause the difficulty of managing directly the individual time-fraction of a specific user, makes the optimal solution of problem (P) be almost explicit form and the TF-TSM be appropriate for real-time application.

Noting two extreme cases ($UM(n)=0$ and ∞) and the fact $\gamma(n)\geq 1, \forall n \geq 1$, the control parameter for the next TFA-cycle, $\gamma(n+1)$ is updated by the equation given as

$$\gamma(n+1)=c^{1/UM(n)+1} \qquad (10)$$

where c (sufficiently large value) is a constant.

Among possible alternatives for the relation between $UM(n)$ and $\gamma(n+1)$, the equation (10) has been chosen due to its simplicity and property of satisfying two extreme cases :

$$UM(n)=0 \rightarrow \gamma(n+1)=c \text{ and}$$
$$UM(n)=\infty \rightarrow \gamma(n+1)=1$$

Still the following points are given to help understand the reasoning about the role of $\gamma(n)$ reducing the amount of the unsatisfied QoS requirements. Recall that the optimal solution (x_i, $i=1,\cdots,N$) of problem (P) is determined such that each of all users belongs to the one of two sets $m=\{i : a_i x_i=\min_i\{a_i x_i\}\}$ and $M=\{i : a_i x_i$

$> \min_i \{a_i x_i\}\}$ as stated in Theorem 1. It would be then most desirable if the unsatisfied QoS realization levels of the users in the set M are always larger than the ones of the users in the set m. Our TF-TSM, though does not always guarantee this desirable property, tends to lead the sum of the unsatisfied QoS realization levels of users to be diminished by reducing the difference of the amount of transmitted data between users in the sets m and M, as long as there exists a user with unsatisfied QoS realization, that is $UM(n) > 0$. The study [9] may be worth mentioning at present because the underlying idea on the time utility functions introduced therein to support real-time and non-real-time users at the same time is similar to our approach, though the details are quite different because the study have dealt OFDM-based PF scheduling method.

The estimation method of the expected data rates of users ($a_i(n)$, $i=1,\cdots,N$) should be simple enough so that the requirement of the almost-real time estimation speed would fulfilled. In this study, we simply assume that the expected data rates of users are equal to the average data rates observed in the latest frame without developing an elaborate estimation method, which will be left as the future study.

It is also worth mentioning about the length of TFA-cycle. As the length of TFA-cycle becomes larger, the accuracy of estimation on the expected data rates of users usually tends to get worse. On the other hand, for the convergence of the TSA algorithm, the length of TFA-cycle needs to be large enough so that all users get as many as the number of time slots corresponding with their time-fractions. Fortunately, the TSA algorithm converges relatively quickly as mentioned in [14], which has been confirmed in our computational experiments given in the next section.

4. Numerical Example

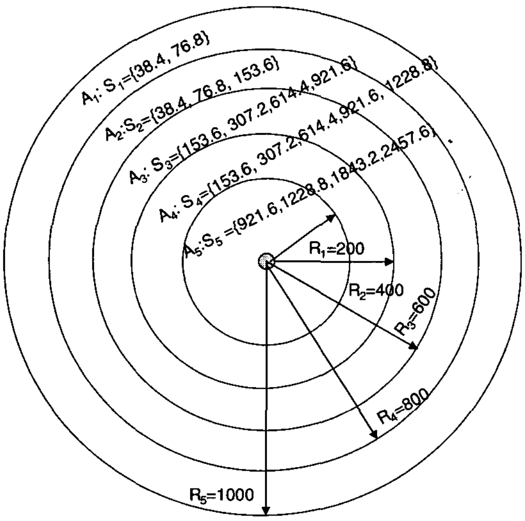
We consider a CDMA/HDR circular cell with a radius of 1000m in which 40 users ($N=40$) are randomly located. In the CDMA/HDR system, the size of time slot is 1.67ms, and a frame is composed of 1000 time slots. The data rates supported on the downlink are 38.5, 76.8, 153.6, 307.2, 614.4, 921.6, 1228.8, 1843.2, and 2457.6 kbps [1, 4]. We assume that the packet length of all users is 128 bytes and that the users are divided into four service classes according to their packet latency criteria. <Table 1> shows the packet latency QoS target values (η_k) and percentage of the number of users belonging to each class k ($k=1,2,3,4$).

〈Table 1〉 QoS criteria and percentage of users for each class

Class(k)	1	2	3	4
Criteria (sec.) (η_k)	0.15	1.0	5.0	10.0
Percentage (%)	20	40	30	10

We assume that at every initial epoch of a frame, the speed (maximum speed 40 km/hr) and moving direction of each user are randomly generated. We further assume that these remain constant during each frame. To keep the number of users within a cell constant, we assume that if a user moves out of the cell border, it reappears at a point that is symmetric to the BS.

In reality, the data rate which can be supported to each user at each time slot is directly determined by the level of SNR (signal-to-noise-ratio) received at user [4, 5, 9] which usually depends on a variety of factors giving an influence on radio propagation. To bring the purpose of this study into focus as well as to keep our simulation process simple, while sustaining the dynamics of the time-varying data rates, we make the following assumptions : The cell area is partitioned into five doughnuts by four co-centric circles with radii $R_1 = 200m$, $R_2 = 400m$, $R_3 = 600m$, $R_4 = 800m$, as shown in <Figure 2>. The five mutually exclusive regions thus obtained are referred to as regions A_i, $i = 1,\cdots,5$ respectively. We denote the set of possible data rates of the users roaming in region A_i by S_i $i = 1,\cdots,5$ and randomly assign a discrete probability distribution for each S_i, $i = 1,\cdots,5$. Prior to the initial epoch of each time slot, the data rate of each user is generated randomly using the corresponding probability distribution. As shown in <Figure 2>, the users in the area A_i have the higher possibility of getting the higher data rates than the users in the area $A_j (i > j)$, by which we have tried to represent the effect of

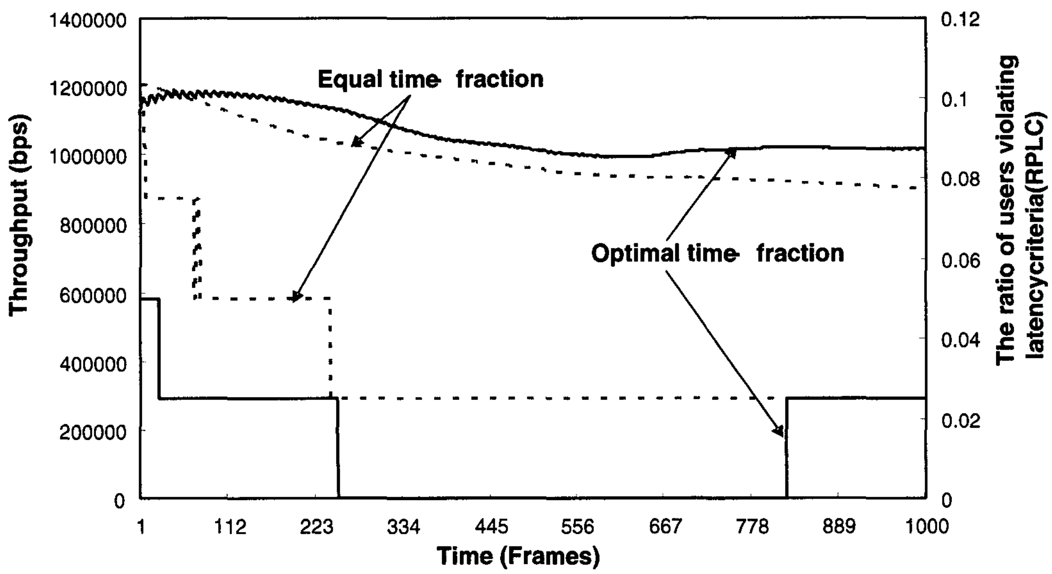


〈Figure 2〉 Achievable data rates of users in their location within cell

path loss. And the data rates of even the users in the same area, which are randomly chosen from the associated distributions of their current regions, could vary every time slot, thereby describing somewhat of the shadowing effect. We further assume that the expected data rates of the users in the next TFA-cycle are equal to the average data rates observed in the latest frame. With the length of TFA-cycle $f = 10{,}000$ (i.e., 10 frames) and $\gamma(n+1) = c^{1/UM(n)+1}$, where $c = 64 \approx \frac{\max_i\{a_i\}}{\min_i\{a_i\}} = \frac{2457.6}{38.564}$ the simulation runs for 1000 frames (1,000,000 slots).

In <Figure 3>, we show how both the performance measures, the system throughput and the ratio of users violating packet latency criteria (RPLC), change over time. Our TF-TSM (optimal time-fraction) that optimally updates the time-fractions assigned to users every TFA-cycle according to the measured packet latency of the users, is first compared with the one (equal time-fraction) that assigns a constant equal time-fraction to all users. Note that a round-robin scheme in which the users are served in a cyclic order is a special case of the TF-TSM with equal time-fraction. In [14], authors have shown that the performance of the TF-TSM with equal time-fraction is 10%~20% higher than the one of the round-robin scheme. As shown in <Figure 3>, both the performance measures in the TF-TSM with optimal time-fractions are better than those with the equal time-fractions for the almost duration of simulation, indicating that the sub-module TFA contributes significantly to the successful implementation of the TF-TSM.

In <Figure 4>~<Figure 6>, we now compare our TF-TSM (simply denoted by TF) with

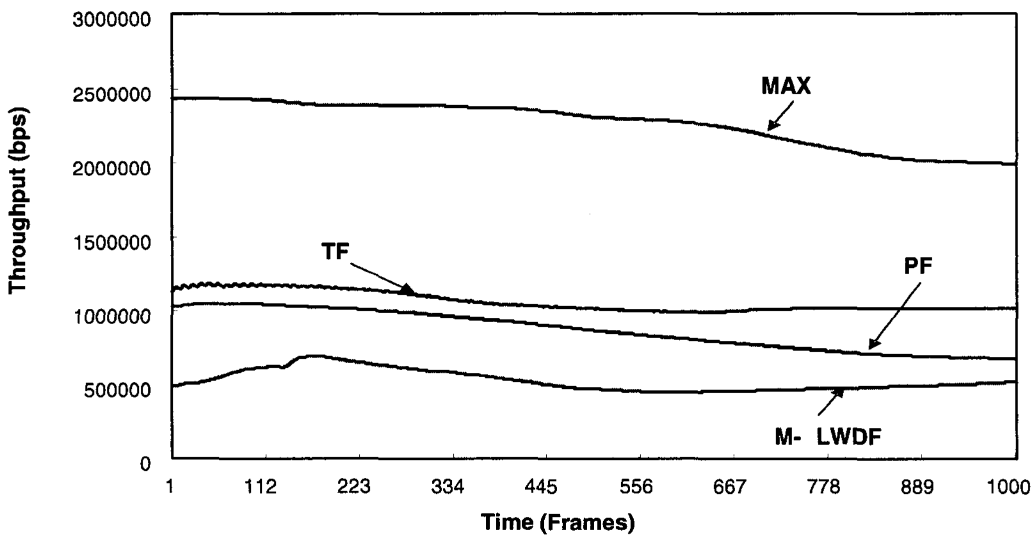


$$TH(k)\sum_{i=1}^{N} td_i(k),\ \ td_i(k) = \frac{\text{sum of tansmitted bits of user i up to the kth frame}}{\text{total number of time slots up to the kth frame}},$$

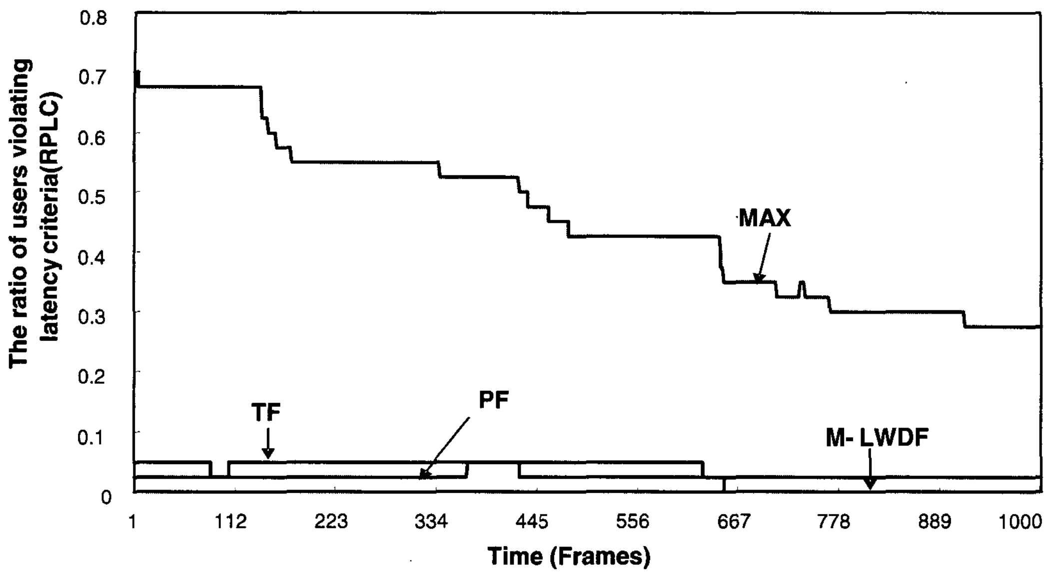
$$RPLC(k) = \frac{\text{\# of users whose packet latencis evaluate up to the kth frame are greater than their QoS criteria}}{\text{the numver of users}},$$

where TH(k) and RPLC(k) indicate the system throughput and RPLC up to the k^{th} frame, respectively.

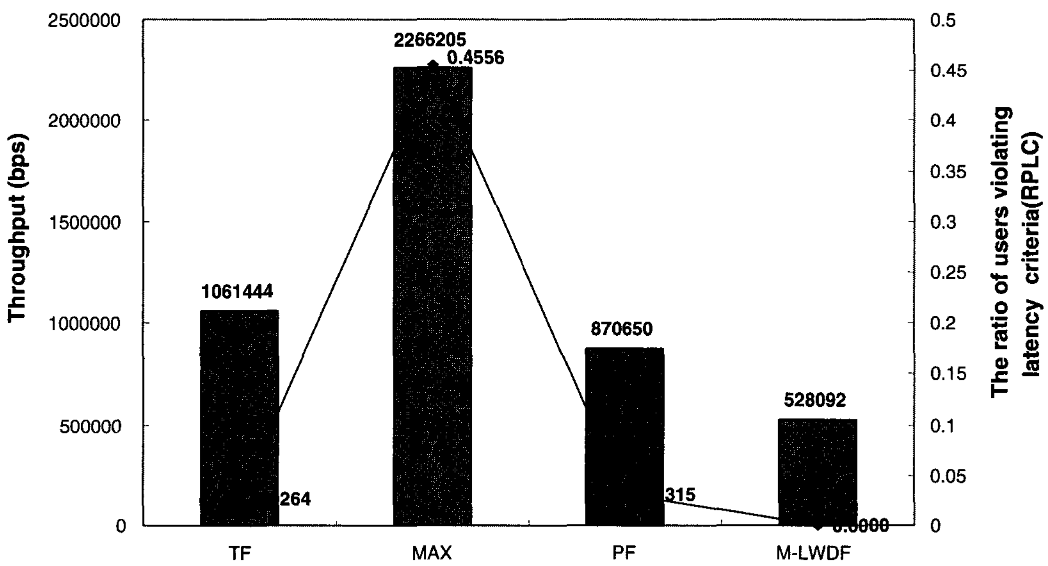
<Figure 3> Comparison of TF-TSM with optimal time-fractions and equal time-fractions : (throughput and RPLC trajectories over 1000 frames)



<Figure 4> Comparisons among different TSMs with regard to system throughput (throughput trajectories over 1000 frames)



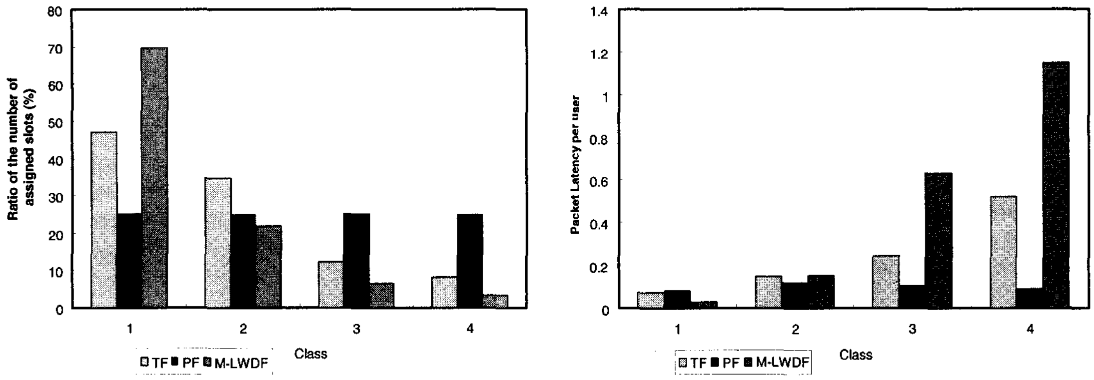
〈Figure 5〉 Comparisons among different TSMs with regard to RPLC: (RPLC trajectories over 1000 frames)



$$\text{Throughput} = \frac{\sum_{k=1}^{1000} \text{TH(k)}}{1000(\text{frames})}, \quad \text{RPLC} = \frac{\sum_{k=1}^{1000} \text{RPLC(k)}}{1000(\text{frames})},$$

where TH(k) and RPLC(k) indicate the system throughput and RPLC up to the k^{th} frame, respectively.

〈Figure 6〉 Comparisons among different TSMs with regard to system throughput and RPLC: (averages of samples over 1000 frames)



$$\text{\% of class } k = \frac{\text{total number of slots assigned to the users belonging to class } k}{\text{total number of slots assigned to the users}} \times \frac{100}{\text{the number of users in class } k}$$

(a)

$$\text{packet latency per user of class } k \ \frac{\text{sum of packet latencies of users longing to class } k}{\text{the number of users in class } k}$$

(b)

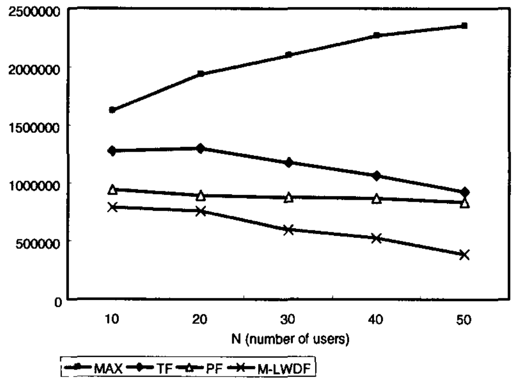
<Figure 7> Performance and slot-assignment differences between service classes (averages of samples over 1000 frames)

three other TSMs : maximum data rate (MAX), proportional fair (PF) [1, 3], and modified largest weighted delay first (M-LWDF) [12, 18]. Note that the MAX indicates the TSM which always schedules the user with the highest data rates. The parameter for PF (window size) is set at 1000 slots as indicated in [3]. For setting parameters of M-LWDF, the method given in [12] has been used except the fact that the head of line packet delay has been substituted by the average packet latency because of the different traffic model. As shown in <Figure 4>~<Figure 6>, the system throughput of our TF is better than those of PF and M-LWDF, and as far as the performance measure RPLC goes, M-LWDF reveals the best performance and TF is a little better than PF. As expected, the MAX shows the best system throughput and the worst RPLC. If we do not consider MAX further as a practical TSM that can be compared to the other three ones, we can say that the system throughput of TF (the RPLC of M-LWDF) is superior to the others. The superiority of RPLC of M-LWDF over our TF stems from its intrinsic nature of reflecting the instant packet latencies of users into every time slot assignment through changing adaptively the related parameters slot-by-slot. But our TF regulates the levels of packet latencies of users by updating the time-fractions assigned to them every TFA-cycle whose duration is much longer than that of a single time slot. This difference between TF and M-LWDF also explains why in the long-term, the system throughput of TF becomes better than that of M-LWDF.

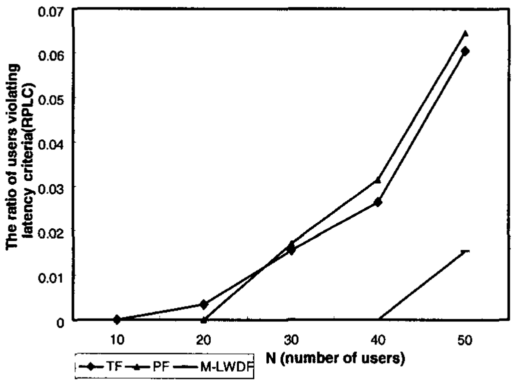
<Figure 7> depicts the percentage of time slots assigned <Figure 7(a)> and the packet latency per user <Figure 7(b)> for each of service classes. These figures confirm again that M-LWDF and PF are the most and least sensitive respectively to the QoS level of packet latency, and our TF is in between them. <Figure 8> and <Figure 9> show how the throughput and RPLC vary in the number of users. From the resultant curves, we can see that M-LWDF and our TF-TSM reveal some adaptive behav-

ior to cope with the congestion of increasing the number of users.

Now, we consider a circumstance in which there is a single class of users in the system.

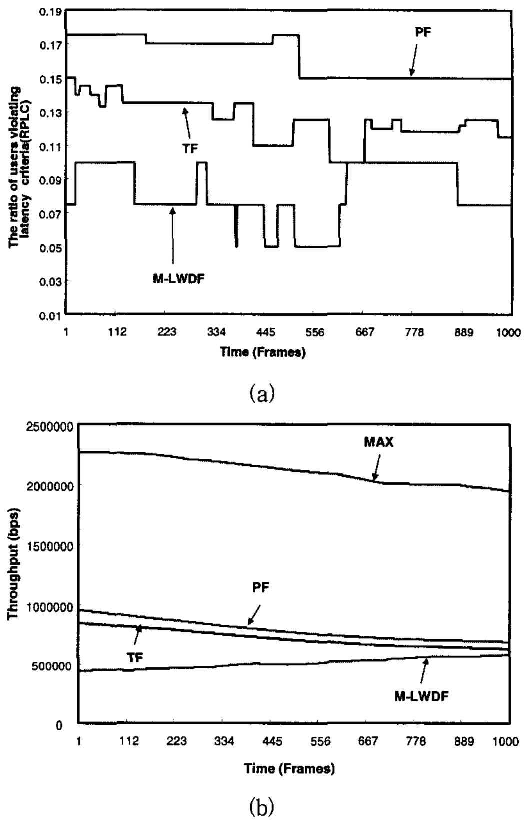


⟨Figure 8⟩ Throughput comparison among different TSMs in the number of users : (averages of samples over 1000 frames)

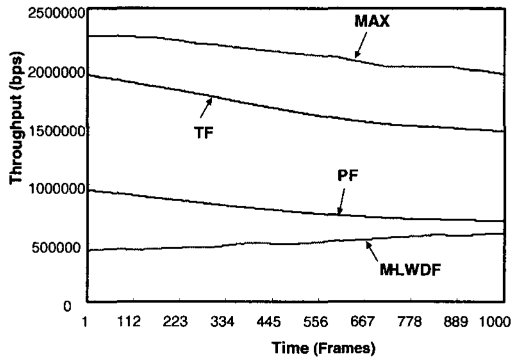


⟨Figure 9⟩ RPLC comparison among different TSMs in the number of users : (averages of samples over 1000 frames)

In <Figure 10>, we compare our TF-TSM with other TSMs when all users are in service class 1 which has the strictest constraint on packet latency delay like real-time services. As shown in this figure, M-LWDF is superior to two other TSMs with regard to RPLC, whereas PF with regard to system throughput. In the case that all users are in service class 4, which is considered



⟨Figure 10⟩ Performance comparisons among different TSMs when all users are in class 1 (trajectories over 1000 frames)



⟨Figure 11⟩ The throughput comparison among different TSMs when all users are in class 4 (trajectories over 1000 frames) to represent non-real-time users, our TF shows the most system throughput excluding MAX <Figure 11>.

To summarize, M-LWDF shows a tendency to outperform TF and PF in the performance measure RPLC at the cost of degrading the system throughput. PF performs better in the utilization fairness measure among all users, but is less effective for discriminating assignments of time slots based on diverse performance requirements as well as for the overall system throughput. Our transmission scheduling method, TF shows a balance between the fulfillment of diverse QoS requirements and the maximization of system throughput. It is also worth mentioning that the three transmission scheduling methods, TF, PF, and M-LWDF, all have their own parameters that need to be determined. The performance of these methods is fairly dependant on the values of these parameters. In case of TF, the value of the constant c in equation (10), the estimation method for the expected data rates of users for the next TFA-cycle, and the length of TFA-cycle are the factors that need to be determined prior to implementation. For our experiments, some value of these parameters (the length of cycle) are determined by an experimental test and the others are determined by intuition, instead of using a more formal systematic method which is left as a topic for future research.

5. Conclusion

This paper dealt with a mathematical problem of finding the optimal time-fractions assignment to users in TF-TSM. Given the control parameters for performance fairness, an explicit form of optimal time-fractions was presented, along with remarks on the procedures for deriving the optimal solution. A systematic procedure was also provided to incorporate the time-varying performance fairness measure into the control parameter of the TFA problem. Simulation results demonstrated that the performance of packet latency could be improved at a minimal cost of decreasing the throughput. An interesting study that arises based on our research is the application of our optimal TFA scheme to the next generation wireless systems. Another interesting issue worth investigating is a systematic method for determining the parameters for the sub-module TFA. Our TF-TSM, despite its systematic framework of deriving optimal time-fractions, has some limited applicability owing to the idealistic assumption of equal packet lengths. The real-world effectiveness of our TF-TSM should further be investigated by applying it to a real-life system with the diverse packet lengths and a more realistic traffic and radio propagation environments, which is left as a future study.

References

[1] Bender, P., P. Black, G.R. Padovani, N. Sindhushayana, and A. Viterbi, "CDMA/HDR : A badwidth-efficient high-speed wireless data service for nomadic users," *IEEE Communication. Magazine*, (2000), pp.70-77.

[2] Holtzman, J.M., "Asymptotic analysis of proportional fair algorithm," *Proc of IEEE PIMRC*, Vol.2(2001), pp.33-37.

[3] Jalali, A., R. Padovani and R. Pankaj, "Data throughput of CDMA-HDR a high efficiency-high data rate personal communication wireless system," *Proc. of IEEE Vehicular Technology Conference* 2000-Spring,

Vol.3(2000), pp.1854-1858.

[4] *1xEV : 1xEVolution IS-856 TIA/EIA Standard, Airlink Overview*, Qualcomm, Inc, 2001.

[5] *3GPP2 C.S0024-0 v4.0, cdma2000 high rate packet data air interface specification*, October 2002.

[6] Choi, Y. and Y. Han, "A channel-based scheduling algorithm for cdma2000 1xEV-DO system," *Proc. of IEEE Personal, Indoor and Mobile Radio Communications* 2002, (PIMRC02), Vol.5(2002), pp.2259-2263.

[7] Kim, K.Y., H. Kim, and Y.N. Han, "A proportionally fair scheduling algorithm with QoS and priority in 1xEV-DO," *Proc. of IEEE PIMRC*, Vol.5(2002), pp.2239-2243.

[8] Rhee, J.H., T.H. Kim, and D.K. Kim, "A wireless fair scheduling algorithm for 1xEV-DO System," *Proc. of IEEE Veh. Techno. Conference 2001 Fall*, Vol.2(2001), pp.743-746.

[9] Ryu, S. and B. Ryu, et al., Wireless packet scheduling algorithm for OFDMA system based on time-utility and channel state, ETRI Journal Vol.27, No.6(2005), pp.777-787.

[10] Ofuji, Y., S. Abeta, and M. Sawahashi, "Comparison of packet scheduling algorithms focusing on user throughput in high speed downlink packet access," *IEICE Transaction on Communications*, Vol.E86-B, No.1(2003), pp.132-139.

[11] Ohta, Y., M. Tsuru, and Y. Oie, "Framework for fair scheduling schemes in the next generation high-speed wireless links," *Proc. of 8th International Conf. on Cellular and Intelligent Communications* (CIC 2003), pp.411-415.

[12] Andrews, M., K. Kumaran, K. Ramanan, A.L. Stolyar, and P. Whiting, "Providing quality of service over a shared wireless link," *IEEE Communication Magazine*, (2001), pp.150-154.

[13] Borst S. and P. Whiting, "Dynamic-channel-sensitive scheduling algorithms for wireless data throughput optimization," *IEEE Trans. Veh. Technol.*, Vol.52, No.3(2003), pp.569-586.

[14] Liu, X., E. Chong, and N. Shroff, "Opportunistic transmission scheduling with resource-sharing constraints in wireless networks," *IEEE Journal on Selected Areas in Communications*, Vol.19, No.10(2001), pp.2053-2064.

[15] Liu, X., E. Chong, and N. Shroff, "Optimal opportunistic scheduling in wireless networks," *Proc. of IEEE Vehicular Technology Conference 2003,* Vol.3(2003), pp.1417-1421.

[16] Marbach, P. and R. Berry, "Downlink resource allocation and pricing for wireless networks," *Proc. of IEEE Infocom 2002*, Vol.3(2002), pp.1470-1479.

[17] Nguyen, H.N. and I. Sasase, "Downlink Queuing Model and Packet Scheduling for Providing Lossless Handoff and QoS in 4G Mobile Networks," *IEEE Trans. on Mobile Computing*, Vol.5, No.5(2006), pp.452-462.

[18] Shakkottai, S. and A.L. Stolyar, "Scheduling algorithms for a mixture of real-time and non-real-time data in HDR," *Proc. of the 17th International Teletraffic Congress (ITC-17)*, Salvador da Bahia, Brazil, 2001, pp.793-804.

[19] Elsayed, K.M. and A.K. Khattab, "Channel-Aware Earliest Deadline Due Fair Scheduling for Wireless Multimedia Networks," *Journal of Wireless Personal Communications*, Vol.38, No.2(2006), pp.233-252.

[20] Nemhauser, G. and L.A. Wolsey, *Integer and combinatorial optimization*, John Wiley & Sons, Inc. 1988.