

# Utility-oriented Adaptive QoS and Bandwidth Allocation in Wireless Networks

Yaxin Cao and Victor O. K. Li

*Abstract*—In this paper we propose a general utility-oriented adaptive Quality of Service (QoS) model for wireless networks and establish a framework for formulating the bandwidth allocation problem for users with time-varying links. For slow link variations, it is inadequate to only employ low-level adaptive mechanisms at the symbol or packet level, such as error correction coding or swapping packet transmission opportunities. To improve bandwidth utilization and satisfy users' QoS requirements, high-level adaptive mechanisms working at larger time scale is needed. We propose an adaptive bandwidth allocation scheme, which is capable of providing QoS guarantees, ensuring long-term fairness, and achieving high bandwidth utilization. A finite-state Markov channel model (FSMC) is used to model wireless links.

## I. INTRODUCTION AND MOTIVATION

With the rapid growth in popularity of wireless data services and the increasing maturity of wired multimedia networks, extending multimedia services into wireless networks is inevitable. However, deploying such services is a challenging task. Wireless communications pose special problems, such as limited bandwidth and high error rate, that do not exist in wired networks. There exist two dimensions of adverse dynamics in wireless networks, namely, the dynamics of physical channels and the dynamics of users' requests. The dynamics of physical channels refer to the inherent variability of wireless channels, which is both time-dependent and location-dependent. Uncoordinated users' transmission requests make up the other dimension of the challenging dynamics. The broadcast nature of wireless communication causes conflicts or interference among users' transmissions.

Due to the above problems, deterministic service guarantees and bandwidth allocation, commonly used in wired networks, become inadequate in wireless networks. We believe that a more flexible service model which allows variable QoS is needed. Similar views of requiring new adaptive QoS service model in wireless networks can be found in [8][9][10]. However, none of the papers discusses the problems of location-dependent link variations and links with multiple physical states. The only channel variation considered is the variation of the total available bandwidth/capacity. However, different users may experience different link capacities due to different locations. To combat wireless link variability and improve bandwidth utilization, we believe bandwidth should be allocated in an adaptive and link-state-dependent way. Irrespective of the amount of bandwidth received, the ultimate measure of the effectiveness of network services is the level of users' satisfaction, which is dependent on the specific application type. To capture the heterogeneity of different applications and to have a consistent performance measure, we adopt utility functions in our proposed adaptive QoS model.

This research is supported in part by the Research Grants Council, Hong Kong, under Grant No. HKU7225/99E and by the NSFC/RGC Joint Research Scheme under Grant No. N\_HKU709/00.

Y. Cao is with the Dept. of Electrical Engineering, University of Southern California, Los Angeles, CA, U. S. A. (ycao@usc.edu).

V. O. K. Li is with the Dept. of Electrical and Electronic Engineering, the University of Hong Kong, Pokfulam, Hong Kong (vli@eee.hku.hk).

In this paper, we propose a utility-oriented wireless adaptive QoS model and a bandwidth allocation scheme which accounts for the users' QoS requirements and actively adapts to the dynamics of the physical channel. There has been much work on wireless resource (bandwidth) management, focusing on multiple access [4] and channel allocation [3]. Most of the previous work tackled one aspect of the bandwidth allocation problem, i.e. the dynamics of user requests. That is, resolving conflicts due to users' uncoordinated requests and allocating transmission slots or call channels appropriately to satisfy those requests. However, there is less research on adding explicit adaptive mechanisms to bandwidth allocation schemes to deal with the variations of wireless channels.

The rest of the paper is organized as follows. Section II discusses the wireless variations and the need of adaptive mechanisms for different time scales. In Section III the utility-oriented adaptive QoS model is discussed. Section IV describes the proposed bandwidth allocation scheme. Simulation results are presented in Section V. Section VI concludes the paper.

## II. DEALING WITH WIRELESS LINK VARIATIONS

Wireless links are usually subject to two types of variations, i.e. slow variations (shadowing) and fast variations (fading). For typical cellular communications, the duration of shadowing is in the order of seconds or tens of seconds, while fading usually lasts for milliseconds or shorter.

To handle physical link variability, low level mechanisms such as error correction coding [7] and swapping transmission opportunities in packet scheduling [5] are usually used. However, such mechanisms work for relatively small time scales, e.g. the duration of a symbol or a packet, which are comparable to the duration of fast fading. For slow link variations, such mechanisms alone are inadequate. In adaptive coding, a user with very bad link quality may waste a great amount of bandwidth on coding overhead. Swapping transmission slots in packet scheduling tries to improve effective bandwidth utilization. In most wireless scheduling schemes[5], where a two-state Markov channel model is used, a user will not receive any bandwidth when its link experiences a long-lasting shadowing degradation. However, in reality, the capacity of a wireless link will have more than two states. When slow variations are dominant, a more desirable approach is to change both the code length and the amount of bandwidth allocated to a user as its link changes state. Thus a user will still be able to receive some service when its link quality degrades and the overall bandwidth will be utilized more effectively by allocating more bandwidth to users who can better utilize it. When fading and shadowing occur simultaneously, the fast variations are superimposed on slow variations, with the latter actually determining the short-term (in the order of seconds) average link quality. To improve bandwidth utilization, in addition to the swapping mechanism at the packet level, a high-level bandwidth allocation scheme should adjust the average bandwidth share (e.g. the scheduling weight) of each user

as the average link quality changes.

The main focus of this paper is to establish a very general modeling framework of the high-level bandwidth allocation problem based on an adaptive QoS model and design an adaptive allocation scheme to deal with slow link variations.

### III. SYSTEM MODELING

#### A. Modeling the Time-varying Links

It has been proven that the finite-state Markov channel (FSMC) can accurately model both fading and shadowing channels[1][2]. Each channel state corresponds to some channel quality and/or response at the receiver. To completely describe such a Markov channel, we need the state transition probabilities, average state-holding times, and some parameter that reflects the physical characteristics of each state. The transition probabilities can be specified by a transition probability matrix  $\mathcal{P} = [p_{i,j}]$ , where  $p_{i,j}$  is the transition probability from state  $i$  to state  $j$ .

To avoid confusion, from now on we use *channel* to represent the overall wireless channel shared by all users, and *link* to represent a wireless link between two specific communicating parties (e.g. a base station and a mobile user). Each wireless link is modeled by an FSMC. Assuming that all the users move freely in the same region, all the links are independent and identical. To capture the link characteristics of each state of the Markov chain, we associate each state  $m$  with a parameter called *bandwidth degradation ratio*  $D_m$ , where  $0 \leq D_m < 1$ . The *bandwidth degradation ratio* represents the overall degree of bandwidth wastage incurred by unsuccessful transmissions, coding overhead, and other factors. More specifically, if the bandwidth allocated to the user is  $r$ , and its link is currently in state  $m$ ,  $D_m \cdot r$  of bandwidth will be wasted. We call  $(1 - D_m) \cdot r$  the *effective bandwidth* received by the user.

#### B. Utility-oriented Adaptive QoS Service Model

Utility, a concept originally used in economics, has been brought into networking research [6][9][10] in recent years. Utility represents the "level of satisfaction" of a user or the performance of an application. A utility function, which is monotonically non-decreasing, describes how the utility perceived by a user changes with the amount of effective bandwidth it receives. The key advantage of the utility function is that it inherently reflects a user's QoS requirements and can quantify the adaptability of a user or an application.

In an adaptive QoS model, the user applications are required to be adaptable to service degradations, and the bandwidth allocated to the user is not fixed, but adjusted according to the condition of the network. We propose a utility-oriented adaptive QoS service model for wireless networks. In this service model, each user <sup>1</sup>  $i$  signals its utility function  $U_i(r)$ , *minimum utility level*  $u_{i,min}$  and *maximum utility level*  $u_{i,max}$  to the network, where  $r$  is the amount of *effective bandwidth* received by the user. At any time instance, the *instant utility value* of the user is either zero or in the range of  $[u_{i,min}, u_{i,max}]$ . The network

<sup>1</sup>In the rest of the paper, the word "user" represents either a user or an application.

tries to dynamically allocate bandwidth such that each user's *instant utility* is maintained above  $u_{i,min}$  and in the long run the bandwidth is allocated fairly and utilized efficiently.

Based on the service model and the wireless link model we have discussed, the following is the complete description of the bandwidth allocation problem. There are  $n$  active users who share transmission bandwidth  $R$ . The communication link of each user follows a  $k$ -state Markov channel model. The statistical characteristics of the Markov model for all links are *i.i.d.* The average state-holding time of link state  $m$  is  $t_m$ , and the *bandwidth degradation ratio* of the state is  $D_m$ , where

$$0 \leq D_m < 1, \forall 1 \leq m \leq k \quad (1)$$

Without loss of generality, we assume that

$$D_m > D_n, \text{ if } m < n \quad (2)$$

To associate the *degradation ratio* to each user, if user  $i$ 's link is currently in the  $m$ th state,  $D_{i,m}$  will be used. The utility function of user  $i$  is  $U_i(r)$ , where

$$\begin{cases} U_i(r) = 0 & : r < r_{i,min} \\ u_{i,min} \leq U_i(r) \leq u_{i,max} & : r_{i,min} \leq r \leq r_{i,max} \\ U_i(r) = u_{i,max} & : r > r_{i,max} \end{cases} \quad (3)$$

We call  $r_{i,min}$  the *minimum effective bandwidth level* and  $r_{i,max}$  the *maximum effective bandwidth level* of user  $i$ . If at a particular time instance, user  $i$  is allocated  $r_i$  amount of bandwidth and its link is in state  $m$ , then the *instant utility* it receives is

$$u_i = U_i((1 - D_{i,m}) \cdot r_i) \quad (4)$$

To fully utilize the bandwidth, no bandwidth is reserved at any time, i.e.

$$\sum_{i=1}^n r_i = R \quad (5)$$

is always satisfied. It is assumed that each user  $i$ ,  $i = 1, 2, \dots, n$ , always generates enough traffic to fully consume the allocated bandwidth as long as the *effective bandwidth* it receives does not exceed  $u_{i,max}$ . How do we dynamically allocate the bandwidth?

To answer the above question, we first need to clarify our objectives. The key issues of all bandwidth allocation problems are: QoS requirements, fairness, and bandwidth utilization.

1. *QoS requirements*: One of the objectives of the bandwidth allocation scheme is to guarantee  $u_{i,min}$  for each user  $i$ . Define *utility outage* as the event that user  $i$ 's *instant utility* level falls below  $u_{i,min}$ . Therefore, the bandwidth allocation scheme should guarantee for each user that the probability of *utility outage* is smaller than a certain threshold  $p_{outage}$ .

2. *Fairness*: Since the ultimate service criterion is the amount of received utility, the fairness criterion should also be based on utility. Consider users  $i$  and  $j$  with *average utility*,<sup>2</sup>  $u_{i,avg}$  and  $u_{j,avg}$ , respectively,

$$G_i = \frac{u_{i,avg} - u_{i,min}}{u_{i,min}} \quad (6)$$

<sup>2</sup>In this general model, the averaging method and the measuring window are flexible. It can be simple time average or weighted average, and the measuring window may start from the beginning of the data session or may just be a fixed time interval, which should be much larger than the state-holding times. The choice can be up to the service provider.

represents the *normalized gap* of the *average utility* received by user  $i$  and its *minimum utility level*. Since the average statistics of all user's link models are the same, we want all users to have the same *normalized gap* in the long run, i.e.

$$G_i \simeq G_j, \forall i, j \quad (7)$$

3. *Bandwidth utilization*: The total effective utility delivered,  $\sum_{i=1}^n u_{i,avg}$ , is the criterion for measuring the bandwidth utilization.

#### IV. BANDWIDTH ALLOCATION SCHEME

##### A. Adaptive Bandwidth Allocation

When a wireless link changes to a state with a larger  $D_{i,m}$ , we say the link degrades. When a link changes to a state with a smaller  $D_{i,m}$ , we say the link upgrades. The basic idea of the bandwidth allocation scheme is that when a user's link degrades, it may surrender some bandwidth to another user with a smaller *normalized gap*, such that there is a net gain in the combined *instant utility*. When a link upgrades, the user may receive some bandwidth from another user with a larger *normalized gap* to achieve a gain in the combined *instant utility*.

We assume that accurate knowledge of link states is available. The adaptive bandwidth allocation mechanism is invoked whenever there is a link-state change. Assume at a particular time user  $i$ 's link state changes to state  $p$ . The following four steps are the operations performed. First, all the users' *average utility* level and *normalized gap* are updated. Second, users are sorted in increasing order of *normalized gap*. Third, if the *instant utility level* of user  $i$  is below  $u_{i,min}$ , some other users' bandwidth will be reduced and reallocated to user  $i$  to meet its  $u_{i,min}$ . Last, if there is no step three, user  $i$  may give up part of its bandwidth to another user if the link degrades, whereas it may receive some extra bandwidth if the link upgrades. We call the user who gives up part of its bandwidth to others the *benefactor*, and the user who receives bandwidth from others the *beneficiary*.

In the third step, to satisfy user  $i$ 's  $u_{i,min}$ , the bandwidth allocation scheme searches for *benefactor(s)* starting from the user with the largest *normalized gap*. Suppose the user with the largest *normalized gap* is user  $j$ , whose link is currently in state  $q$  and it is above  $u_{j,min}$ . User  $j$  will yield

$$\min\left(\frac{r_{i,min}}{1 - D_{i,p}} - r_i, r_j - \frac{r_{j,min}}{1 - D_{j,q}}\right) \quad (8)$$

amount of bandwidth to user  $i$ , where  $r_i$  and  $r_j$  are the bandwidth allocated to users  $i$  and  $j$ , respectively, before the link state transition. If  $\min(\cdot)$  takes the value of the second term in the parenthesis, it means user  $j$  can provide enough bandwidth to user  $i$  to satisfy  $u_{i,min}$  while maintaining  $u_{j,min}$ . If  $\min(\cdot)$  takes the value of the first term in the parenthesis, it means user  $j$ 's surplus bandwidth alone is insufficient for user  $i$  to reach  $u_{i,min}$ . In this case, the *instant utility* of user  $j$  is kept at the minimum level, and all surplus bandwidth is allocated to user  $i$ . Then the user with the second largest *normalized gap* will be the next candidate for *benefactor*. This procedure will be repeated until  $u_{i,min}$  can be reached or all the users have been checked.

If after the link state transition, user  $i$  is still above  $u_{i,min}$ , user  $i$ 's bandwidth share may be adjusted to improve the bandwidth utilization. When user  $i$ 's link degrades, the bandwidth

allocation scheme will search for an appropriate *beneficiary* to receive some bandwidth from user  $i$  and decide the amount to be transferred. The users are checked in increasing order of *normalized gap*. Suppose user  $j$ , whose link is in state  $q$ , is the *beneficiary* candidate being checked. We have

$$\begin{cases} u_i = U_i((1 - D_{i,p}) \cdot r_i) \\ u_j = U_j((1 - D_{j,q}) \cdot r_j) \end{cases} \quad (9)$$

The bandwidth allocation scheme tries to maximize the combined *instant utility* of the two users with some constraints. The optimization problem is:

maximize:

$$u'_i + u'_j \quad (10)$$

where

$$\begin{cases} u'_i = U_i((1 - D_{i,p}) \cdot (r_i - x)) \\ u'_j = U_j((1 - D_{j,q}) \cdot (r_j + x)) \end{cases} \quad (11)$$

subject to:

$$\begin{cases} x \geq 0 \\ u'_i \geq u_{i,min} \\ u'_j \geq u_{j,min} \end{cases} \quad (12)$$

(12) can be simplified as

$$\max(0, \frac{r_{j,min}}{1 - D_{j,q}} - r_j) \leq x \leq r_i - \frac{r_{i,min}}{1 - D_{i,p}} \quad (13)$$

Since the utility functions are bounded and monotonically non-decreasing and the constraints are linear, the above optimization problem is guaranteed to have solution(s), which can easily be solved by numerical methods. If the solution of  $x > 0$ , then reallocate the bandwidth of users  $i$  and  $j$  as:  $r_i = r_i - x$  and  $r_j = r_j + x$ . If  $x = 0$ , then the same procedure is repeated for the user with the next smallest *normalized gap*. This procedure is repeated until one *beneficiary* is found or all users with smaller *normalized gap* than user  $i$ 's have been searched.

The main difficulty in allocating bandwidth is how to combine utilization and fairness considerations and strike a balance between achieving high bandwidth utilization and fairness among users. If achieving high bandwidth utilization is the sole objective, some users may suffer starvations. If absolute fairness, such as keeping all the users at the same instant utility level in [9], is maintained at all times, bandwidth utilization is sacrificed. The operations described actually combine the considerations of both long-term fairness and short-term maximization of bandwidth utilization. First, only users who are lagging behind user  $i$  in *normalized average utility* are in the *beneficiary* candidate list. Considering long-term fairness objective, when a user gives up its bandwidth, such bandwidth is transferred to the users who have received less utility than its fair share, so that they can catch up. The smaller the user's *normalized gap*, the higher its priority in the candidate list. Second, reallocating bandwidth between the *benefactor* and the *beneficiary* is aimed at maximizing the combined *instant utility*, and hence the bandwidth utilization.

Similarly, when user  $i$ 's link upgrades, user  $i$  becomes the *beneficiary* and users with larger *normalized gap* are the candidates for *benefactor*. The scheme checks the candidates starting from the one with the largest *normalized gap*. Suppose user  $j$ , whose link is in state  $q$ , is the *benefactor* candidate being checked. The optimization problem becomes:

maximize:

$$u'_i + u'_j \quad (14)$$

where

$$\begin{cases} u'_i = U_i((1 - D_{i,p}) \cdot (r_i + x)) \\ u'_j = U_j((1 - D_{j,q}) \cdot (r_j - x)) \end{cases} \quad (15)$$

subject to:

$$\max(0, \frac{r_{i,\min}}{1 - D_{i,p}} - r_i) \leq x \leq r_j - \frac{r_{j,\min}}{1 - D_{j,q}} \quad (16)$$

If  $x > 0$ , then reallocate the bandwidth of user  $i$  and  $j$  as:  $r_i = r_i + x$  and  $r_j = r_j - x$ . If  $x = 0$ , then the same procedure is repeated for the user with the next largest *normalized gap*. This is repeated until one *benefactor* is found or all the users with larger *normalized gap* than user  $i$ 's have been searched.

Besides the link state changes, adjustments in bandwidth allocation is also needed when the following events take place. Bandwidth needs to be collected from (or distributed to) the users in the network when the overall available bandwidth decreases (or increases) or a new user arrives (or departs).

If  $r$  is the amount of deficient bandwidth which needs to be collected from the current users, either because of a decrease in overall bandwidth or a user's arrival, user  $j$  having the largest *normalized gap*  $G_j$  is to give up  $\min(\max(0, r_j - \frac{r_{j,\min}}{1 - D_{j,q}}), r)$  amount of bandwidth, where  $q$  is the current link state of user  $j$ . If it is still not enough, the user with the second largest *normalized gap* is chosen to give up bandwidth to make up the deficit. This procedure will be repeated until enough bandwidth has been collected or all the current users have been searched. If after searching all the current users, the collected bandwidth is still not enough, the bandwidth allocation scheme will start the second round of collection, again starting from the user with the largest *normalized gap*. This time each chosen user will give up all of its bandwidth or the deficient amount until the deficit becomes zero. Similarly, if there is surplus bandwidth, the users with the first  $k$  smallest *normalized gap* are chosen to receive the surplus bandwidth. Each user can increase its *effective bandwidth* up to the *maximum effective bandwidth level*.

### B. Admission Control and Utility Outage Probability

It is clear that to guarantee users' *minimum utility level*, certain admission control should be enforced to limit the number of users in the system. Given an FSMC's transition probability matrix  $\mathcal{P}$ , the steady state probability  $\Pi = [\pi_1, \pi_2, \dots, \pi_k]$  can be calculated by solving the following vector equation.

$$\Pi = \Pi \mathcal{P} \quad (17)$$

If state  $i$ 's average holding time is  $t_i$ , then at a particular time the probability of the link being in state  $i$  is

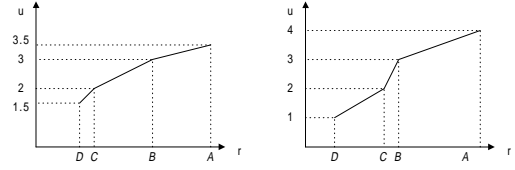
$$p_i = \frac{\pi_i \cdot t_i}{\sum_{i=1}^k \pi_i \cdot t_i} \quad (18)$$

Recall that when a user's *instant utility* falls below its *minimum utility level*, then there is a *utility outage* for the user. If the total bandwidth needed to keep all the users above their *minimum utility levels* is larger than the available bandwidth, then there will be at least one user experiencing a *utility outage*. The probability  $p_o$  of such event at any time is

$$p_o = P_r \left\{ \sum_{i=1}^n \frac{r_{i,\min}}{1 - D_{i,m_i}} > R \right\} \quad (19)$$

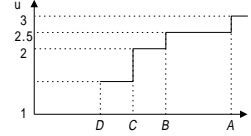
$$= \sum_A \prod_{1 \leq i \leq n} p_{m_i} \quad (20)$$

Fig. 1. Three classes of users in the simulations



(a) Class I:  $D = 7.5, C = 10, B = 20, A = 30$

(b) Class II:  $D = 15, C = 23, B = 24, A = 36$



(c) Class III:  $D = 10, C = 15, B = 20, A = 30$

where  $A = \{m_1, m_2, \dots, m_n \mid 1 \leq m_1, m_2, \dots, m_n \leq k, \sum_{l=1}^n \frac{r_{l,\min}}{1 - D_{l,m_l}} > R\}$ ,  $R$  is the total available bandwidth,  $m_i$  is user  $i$ 's link state at the time instance. It is obvious that  $p_{i,o} \leq p_o$ , where  $p_{i,o}$  is the *utility outage probability* of user  $i$ . Therefore, if we can guarantee that  $p_o \leq p_{outage}$ , where  $p_{outage}$  is the desired threshold, we will have  $p_{i,o} \leq p_{outage}$ .

Consequently, the admission control policy is as follows. When a new user arrives, calculate  $p_o$  according to (20), where  $n$  is the total number of users including the new user. If  $p_o \leq p_{outage}$ , the new user is admitted; otherwise, the user is rejected. If a user  $j$  is admitted, it is initially allocated  $r_{j,\min}/(1 - D_{j,q})$ , where  $q$  is user  $j$ 's current link state. The assigned amount of bandwidth to  $j$  is contributed by the users currently in the network following the algorithm we described previously. Given  $p_{outage}$ ,  $R$ , the channel model and all the possible classes of utility functions, (20) can also be used to find the feasible region of providing statistical minimum utility guarantees.

## V. SIMULATIONS

Without loss of generality, we use the piece-wise linear utility functions in Fig. 1 to represent three classes of users. The horizontal axis is the *effective bandwidth* allocated to a user, and the vertical axis is the utility received. In each figure,  $D$  is the *minimum effective bandwidth level* and  $A$  is the *maximum effective bandwidth level*. The total available bandwidth  $R = 100$ . Since all the values are of relative importance only, we do not specify any units for the parameters in the simulation.

$$\mathcal{P} = \begin{bmatrix} 0 & 0.8 & 0.2 \\ 0.5 & 0 & 0.5 \\ 0.2 & 0.8 & 0 \end{bmatrix} \quad (21)$$

A three-state Markov channel model is used to model wireless links. The transition probability matrix is shown in (21). The average state-holding times are  $t_1 = 3$ ,  $t_2 = 5$ , and  $t_3 = 4$ , and the degradation ratios are  $D_1 = 0.4$ ,  $D_2 = 0.2$ , and  $D_3 = 0$ .

Since there is no existing bandwidth allocation scheme for networks with multiple-state (more than two) links, the adaptive

TABLE I  
COMPARISONS WHEN  $K_1 = 2$ ,  $K_2 = 2$ , AND  $K_3 = 2$

	Normalized Gap		Utility Outage Time	
	Adaptive	Static	Adaptive	Static
user 1	0.35304	0.04177	23.2607	2012.6972
user 2	0.35325	0.03624	19.3312	2038.2643
user 3	0.35320	0.40172	41.5280	1985.2418
user 4	0.35398	0.38246	41.5824	2148.8781
user 5	0.35330	0.07510	10.0698	1913.9185
user 6	0.35298	0.11179	11.9377	1878.1480
Utilization Improvement	17.12%			

TABLE II  
COMPARISONS WHEN  $K_1 = 2$ ,  $K_2 = 1$ , AND  $K_3 = 3$

	Normalized Gap		Utility Outage Time	
	Adaptive	Static	Adaptive	Static
user 1	0.45199	0.11268	2.1284	2012.6972
user 2	0.45218	0.10734	0	2038.2643
user 3	0.45186	0.72108	0	1985.2418
user 4	0.45218	0.06785	0	2148.8781
user 5	0.45228	0.07510	0	1913.9185
user 6	0.45175	0.11179	0	1878.1480
Utilization Improvement	22.38%			

bandwidth allocation scheme is compared with a static bandwidth allocation scheme in terms of utility outage times, fairness property and bandwidth utilization. In the static bandwidth allocation scheme, each user is allocated a fixed amount of bandwidth  $r_i$ , given by

$$r_i = R \frac{r_{i,min}}{\sum_{i=1}^n r_{i,min}} \quad (22)$$

where  $r_{i,min}$  is each user's *minimum effective bandwidth level* and  $n$  is the number of users. The initial allocation of the bandwidth for the adaptive scheme at the beginning of the simulation also follows (22). For fair comparison, the overall available bandwidth  $R$  and the user set are fixed during each simulation. The duration of each simulation is 10000.

Table I shows the numerical results of the case where there are two users in each class. The table shows that for the adaptive scheme the maximum difference in *normalized gap* between any two users is 0.001 (between user 4 and user 6), whereas for the static scheme the corresponding value is 0.36548 (between user 3 and user 1). Since the difference in *normalized gaps* is a measure of the fairness, it is clear the adaptive scheme is much fairer. The result also shows that users experience much less utility outage time in the adaptive scheme than in the static scheme. Using equation (20) it can be calculated that the upper bound  $p_o$  is 0.00416, which corresponds to the maximum *utility outage time* 41.6. It is shown that none of the user's *utility outage time* is larger than the upper bound. Finally, there is a 17.12% utilization improvement in terms of the total utility delivered for the adaptive scheme.

Various user sets are tested in our simulations. Due to space limitations, not all results are presented. For illustration another example is presented in table II. From Tables I and II it is clear that the adaptive scheme outperforms the static scheme in all three aspects: minimum QoS guarantee, fairness, and utilization.

Since in the static scheme there is no dynamic adjustment of

TABLE III  
COMPARISONS WHEN  $K_2 = 2$  AND  $K_3 = 2$

	<i>max</i> $ G_i - G_j , \forall i, j$	Max. Utility Outage Time		Utilization Improvement	
	Adaptive	Static	Adaptive		Static
$K_1 = 1$	0.00068	0.75880	0	0	6.35%
$K_1 = 2$	0.00100	0.36548	41.5824	2038.2643	17.12%
$K_1 = 3$	0.00205	0.30330	674.4124	2192.5303	21.58%
$K_1 = 4$	0.00147	0.16114	2046.0076	7407.9264	219.84%

bandwidth, when the network load increases as the number of user increases the users will spend a great amount of time experiencing *utility outage*, which lowers bandwidth utilization significantly. However, in the adaptive scheme such a problem is less serious because of dynamic allocation. Therefore, it is expected that the utilization improvement of the adaptive scheme is more significant with higher network load. This is demonstrated by the data in Table III. In Table III the number of class II and III users are fixed at  $K_2 = K_3 = 2$ , while the number of class I users is increased from 1 to 4. We see that as  $K_1$  increases, the advantages of the adaptive scheme strengthens in terms of utilization improvement and *utility outage* time. At the same time, the fairness property of the adaptive scheme is maintained.

## VI. CONCLUSION

In this paper we proposed a general utility-oriented adaptive QoS service model for wireless networks and addressed the problem of bandwidth allocation with links having multiple states. Due to the generality of the service model and the bandwidth allocation problem, they can be applied to different networks, such as cellular networks and satellite networks, with heterogeneous classes of users. We also designed an adaptive bandwidth allocation scheme, which is capable of providing QoS guarantees, ensuring long-term fairness, and achieving high bandwidth utilization.

## REFERENCES

- [1] H. Wang, N. Moayeri, "Finite state markov channel - a useful model for radio communication channels," IEEE Trans. Vehicular Tech., pp. 163-171, Feb. 1995.
- [2] F. Babich, G. Lombardi, "A Markov Model for the Mobile Propagation Channel," IEEE Trans. Vehicular Tech., pp. 63-73, Jan. 2000
- [3] I. Katzela and M. Naghshineh, "Channel Assignment Schemes for Cellular Mobile Telecommunication Systems: A Comprehensive Survey," IEEE Personal Commun., pp. 10-30, June 1996.
- [4] I. F. Akyildiz, J. McNair, L. C. Martorell, R. Puigjaner and Y. Yesha, "Medium Access Control Protocols for Multimedia Traffic in Wireless Networks," IEEE Network, pp. 39-47, July-Aug. 1999.
- [5] Y. Cao and V. O. K. Li, "Scheduling algorithms for broadband wireless networks," Proc. of the IEEE, Vol. 89, No. 1, pp. 76 - 87, Jan. 2001.
- [6] S. Shenker, "Fundamental Design Issues for Future Internet," IEEE JSAC Vol.: 13 7, pp. 1176 -1188, Sep. 1995.
- [7] M. Elauod and P. Ramanathan, "Adaptive Use of Error-correction Codes for Real-time Communication in Wireless Networks," Proc. INFOCOM'98, pp. 548-555, Mar. 1998.
- [8] S. Singh, "Quality of Service Guarantees in Mobile Computing," Computer Communications, vol. 9, pp3. 59-371, Apr. 1996.
- [9] G. Bianchi, A. Campbell and R. Liao, "On Utility-Fair Adaptive Services in Wireless Networks," 1998 Sixth International Workshop on Quality of Service, pp. 256-267, 1998.
- [10] D. Reininger, R. Izmailov, B. Rajagopalan, M. Ott and D. Raychaudhuri, "Soft QoS control in the WATMnet Broadband Wireless System," IEEE Personal Communications, pp. 34-43, Feb. 1999.