[1] presented a rate-based borrowing scheme for multimedia wireless networks. In the case of insufficient bandwidth, in order not to deny service to requesting calls (new or handoff), bandwidth is borrowed on a temporary basis from existing calls. Although the scheme is adaptive, it does not include a quantitative measure of the importance of different calls. In [2], a bandwidth adaptation scheme has been introduced to provide QoS in wireless networks. The scheme used quality satisfaction curves, which capture the human perception of service quality provided by the network, to achieve global fairness and quality satisfaction among competing calls within each cell. In [3], the authors have proposed an adaptive resource management architecture and a revenue-based rate adaptation algorithm that seeks to maximize the network revenue; however the message overhead of the algorithm is high. In [4], the bandwidth adaptation was formulated as a binary linear integer programming problem and a heuristic algorithm was designed to allocate the bandwidth of a terminated call to ongoing calls maximizing the user-perceived QoS satisfaction. All the above bandwidth allocation schemes work with adaptive traffic and some apply a quantitative measure (e.g. QoS satisfaction curve which maps bandwidth to user satisfaction) to allocate bandwidth.

This paper presents a utility-based adaptive bandwidth allocation scheme for multi-class traffic including both adaptive and non-adaptive traffic in wireless networks. In our scheme, bandwidth adaptation is decomposed into two processes - bandwidth upgrades and bandwidth degrades. The network dynamically adjusts the allocated bandwidth of ongoing calls to achieve the maximum utility for each individual cell.

The paper is organized as follows. Section 2 describes three classes of traffic and their utility functions used in our study. In Section 3, we formulate the bandwidth adaptation problem. Our bandwidth adaptation algorithm is proposed in Section 4. Sections 5 presents the simulation model and numerical results. Concluding remarks are given in Section 6.

2. TRAFFIC CLASSES AND THEIR UTILITY FUNCTIONS

Before describing traffic classes, the concept of utility function needs to be introduced. Utility functions are curves mapping resources received by applications to their performance as perceived by the user. They are monotonically increasing but not necessarily strictly monotonic; in other words, more resource should not lead to degraded application performance. The shape of the utility function varies according to the adaptive characteristics of the traffic. In this paper, motivated by Shenker's observations [5] three classes of traffic with appropriate utility functions are studied.

2.1. Elastic Traffic

Elastic traffic has no minimum bandwidth requirement (i.e. the minimum bandwidth requirement is 0). Examples include some traditional data traffic such as email, file transfer, and remote terminal access which are rather tolerant of delays. For elastic traffic, there is a diminishing marginal rate of performance enhancement as bandwidth is increased, so their utility function is strictly concave everywhere [5]. The following utility function can be used to model elastic traffic:

$$u(b) = 1 - e^{-\frac{k \cdot b}{b_{max}}}$$

Variable b is the actual allocated bandwidth. Constant k is a positive parameter and b_{max} is the maximum bandwidth requirement; they can be defined together by the user to obtain utility

same traffic class are assigned to the same utility function as described in Section 2. Bandwidth adaptation is decomposed into two processes - bandwidth upgrades and bandwidth degrades; they are performed based on each individual cell which has fixed bandwidth capacity denoted by B .

3.1. Bandwidth Upgrades

Assume that in an overloaded cell when a call is terminated due to its completion or handoff, there are n ongoing calls that have not received their maximum bandwidth. The released bandwidth of the terminated call (denoted by β) can be utilized to upgrade these ongoing calls. Denote the i-th ongoing call's utility function as $u_i(b_i)$ ($1 \le i \le n$) and its current allocated bandwidth as β_i , thus the i-th ongoing call's upgradeable utility function can be written as $u_i^{\uparrow}(b_i) = u_i(\beta_i + b_i)$ $(0 \le b_i \le b_{i,max} - \beta_i)$ where $b_{i,max}$ is the maximum bandwidth requirement. The objective of bandwidth upgrades is to find the bandwidth upgrades profile $\{b_i\}$ ($1 \le i \le n$) to maximize the total cell utility subject to bandwidth constraints, i.e.

$$\text{maximize: } \sum_{i=1}^{n} u_i^{\uparrow}(b_i) \qquad \text{subject to: } \sum_{i=1}^{n} b_i \le \beta \text{ and } 0 \le b_i \le b_{i,max} - \beta_i$$

3.2. Bandwidth Degrades

Consider an overloaded cell with n ongoing calls, when a new or handoff call comes the bandwidth of ongoing calls can be degraded to smaller values to accommodate the new or handoff call, thereby reducing the call blocking and dropping probability. Bandwidth degrades need to decide how to degrade ongoing calls and how to allocate bandwidth to the new or handoff call. Denote the i-th ongoing call's utility function as $u_i(b_i)$ ($1 \le i \le n$) and its current allocated bandwidth as β_i , thus the i-th ongoing call's degradable utility function can be written as $u_i^{\downarrow}(b_i) = u_i(\beta_i - b_i)$ $(0 \le b_i \le \beta_i)$; also denote the new or handoff call's utility function as $u_{n+1}(b_{n+1})$. The objective of bandwidth degrades is to find the bandwidth degrades profile $\{b_i\}$ ($1 \le i \le n$) for ongoing calls and the allocated bandwidth b_{n+1} for the new or handoff call to maximize the total cell utility subject to bandwidth constraints, i.e.

$$\text{maximize: } \left(\sum_{i=1}^{n} u_i^{\downarrow}(b_i)\right) + u_{n+1}(b_{n+1}) \qquad \text{subject to: } \left(\sum_{i=1}^{n} (\beta_i - b_i)\right) + b_{n+1} \le B \text{ and } 0 \le b_i \le \beta_i$$

4. OUR PROPOSED ALGORITHM

From Section 3 we know that bandwidth upgrades and bandwidth degrades are to maximize the total utility of n and $n+1$ utility functions respectively subject to bandwidth constraints. In this section without loss of generality, we only propose an algorithm to maximize the sum of n utility functions. Finding optimal solutions for such a problem is NP-hard. To design the algorithm with low complexity and for the ease of implementation, we first quantize utility function into linear piece-wise segments by dividing the utility range in a fixed number of equal intervals. After the quantization each utility function becomes a linear piece-wise function represented by a set of $< bandwidth, utility >$ points:

$$u_i(b_i) = \left((b_{i,1}, u_{i,1}), \ldots, (b_{i,j_i}, u_{i,j_i})\right)$$

Table 1
Notation for the utility maximization algorithm (Continued)

u_{agg}	the aggregated utility function consisting of all utility functions' line segments sorted by their decreasing slope
$\lvert u_{agg} \rvert$	u_{agg} 's line segments number
$u_{agg}[j]$	u_{agg} 's j-th line segment
$u_{agg}[j].id$	$u_{agg}[j]$'s call id
$u_{agg}[j].bandw$	$u_{agg}[j]$'s bandwidth
$u_{agg}[j].util$	$u_{agg}[j]$'s utility
$\{u_{adapt}\}$	the set of adaptive utility functions
$\{u_{hardr}\}$	the set of hard real-time utility functions

```
$Utility\_Maximization(u_1, \cdots, u_n)$
for $i := 1$ to $n$ do
    if $u_i \in \{u_{adapt}\}$ then
        $u_i' := eliminate(u_i)$
    else
        $u_i' := u_i$
    end if
    $b[i] := 0$
    $u[i] := 0$
end for
$u_{agg} := merge\_and\_sort(u_1', \cdots, u_n')$
$b_{avail} := B$
$u := 0$
for $j := 1$ to $\lvert u_{agg} \rvert$ do
    $i := u_{agg}[j].id$
    $b_{addit} := u_{agg}[j].bandw - b[i]$
    if $(b_{addit} \le b_{avail})$ then
        $b_{avail} := b_{avail} - b_{addit}$
        $b[i] := u_{agg}[j].bandw$
        $u[i] := u_{agg}[j].util$
    else
        if $u_{agg}[j] \notin \{u_{hardr}\}$ then
            $b[i] := b[i] + b_{avail}$
```

- In order to simulate handoffs we randomly choose 50 traffic units as mobile units and give them a speed characteristic. Mobile units can travel in one of six directions with equal probability and their speed is uniformly distributed between 10 and 60 miles per hour.

The details of the multi-class traffic used in our experiments are described in Table 2. Each call belonging to the same class is assigned to the same utility function. To illustrate how the proposed scheme can provide adaptive QoS we compare our adaptive scheme with a non-adaptive one. In the adaptive scheme, a call's allocated bandwidth can be any value between b_{min} and b_{max} (for hard real-time traffic $b_{min} = b_{max}$). While in the non-adaptive scheme, a call must be allocated maximum bandwidth to be admitted and once accepted its bandwidth cannot be changed throughout the lifetime; if such bandwidth is not available, the call is either blocked or dropped depending on whether the call is a new or hand-off call.

5.2. Numerical Results

In our experiments, apart from the traditional call blocking and call dropping probability, the total achieved cell utility is also chosen as a performance metric.

Table 2
Multi-class traffic and their characteristics

Traffic Class	Traffic Type	Bandwidth Requirement	Average Connection Duration	Utility Function (b is Mbps)	Example
I	Elastic Traffic	1 - 10 Mbps (UBR)	2 minutes	$1-e^{-\frac{4.6b}{10}}$	File Transfer & Retrieval Service
II	Adaptive Traffic	1 - 6 Mbps (VBR)	10 minutes	$1-e^{-\frac{1.045b^2}{2.166+b}}$	Interact. Multimedia & Video on Demand
III	Hard Real-time Traffic	256 Kbps (CBR)	5 minutes	$\begin{cases} 1, & b \geq 0.25 \\ 0, & b < 0.25 \end{cases}$	Video-phone & Video-conference

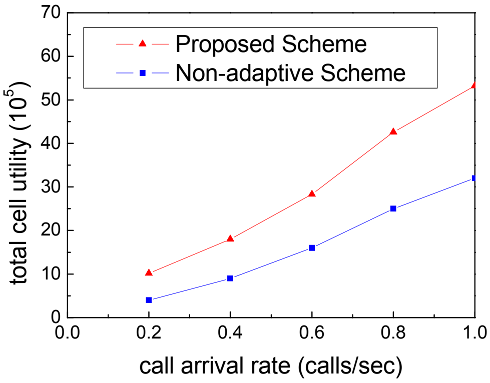

Fig. 3. The total achieved cell utility.

ACKNOWLEDGEMENT

This work is partially funded by the Office of Naval Research, Long Range Navy and Marine Corps Science & Technology Research Opportunity. Contract number: N00014-03-1-0323.

REFERENCES

[1] M. El-Kadi, S. Olariu, and H. Abdel-Wahab, "A rate-based borrowing scheme for QoS provisioning in multimedia wireless networks", IEEE Transactions on Parallel and Distributed Systems, 13(2):156–167, Feb. 2002.
[2] A. Aljadhai and T. F. Znati, "A bandwidth adaptation scheme to support QoS requirement of mobile users in wireless environment", Proceedings of the Ninth International Conference on Computer Communications and Networks, 2000.
[3] V. Bharghavan, K. W. Lee, S. W. Lu, S. W. Ha, J. R. Li, and D. Dwyer, "The timely adaptive resource management architecture", IEEE Personal Communications Magazine, 5(4): 20-31, Aug. 1998.
[4] K.-M. Ahn and S. Kim, "Optimal bandwidth allocation for bandwidth adaptation in wireless multimedia networks", Computers & Operations Research, 30: 1917-1929, 2003.
[5] S. Shenker, "Fundamental design issues for the future internet", IEEE Journal of Selected Areas in Communications, 13(7): 1176-1188, Sept. 1995.
[6] L. Breslau and S. Shenker, "Best-effort versus reservations: a simple comparative analysis", ACM Computer Communications Review, 28: 131-143, Sept. 1998.
[7] H. Kellerer, U. Pferschy and D. Pisinger, "Knapsack problems", Springer-Verlag, 317-322, 2004.
[8] A. Sinha and A. A. Zoltners, "The multiple-choice knapsack problem", Operation Research, 27: 503–515, 1979.