PRIORITY-BASED DYNAMIC PACKET RESERVATION FOR TDMA WIRELESS NETWORKS

Tamer ElBatt
Bo Ryu

HRL Laboratories, LLC [1]
Network Analysis and Systems Dept.
Malibu, CA 90265, USA
telbatt@wins.hrl.com, ryu@wins.hrl.com

ABSTRACT

In this paper we investigate the problem of dynamic bandwidth reservation over the uplink of TDMA-based wireless networks providing prioritized services. The proposed multiple access protocol has the following salient features. First, it supports multiple levels of priorities. Second, it optimizes a unique cost function, that consists of a weighted sum of allocation "Deviation" functions, on a frame-by-frame basis. This, in turn, reduces the computational complexity significantly and hence makes it more suitable to real-time operation. Furthermore, the proposed cost function is amenable to analytical optimization techniques. Third, it supports wide variety of traffic classes, namely multimedia, short message bursts and call hand-off requests. The simulation results show the performance improvement achieved by this protocol compared to a heuristic policy based on Weighted Fair Queuing.

I. INTRODUCTION

One of the driving forces of next generation wireless networks is quality of service support for multimedia services. In this paper we focus on a TDMA cellular system that carries voice and data traffic. However, the dynamic slot reservation algorithm introduced here can be applied to several types of wireless networks, *e.g.* satellite and personal communication networks. The main objective of this paper is to investigate the problem of providing packet level QoS for prioritized services over the uplink of TDMA-based wireless networks. Various bandwidth allocation schemes have been introduced in the cellular and satellite networks literature [1], [2]. Packet reservation multiple access (*PRMA*) has been introduced to integrate voice and data applications over the shared wireless medium [3]. This protocol focuses primarily on allocating bandwidth for "periodic" and "random" traffic streams with vastly different QoS constraints. A major limitation of *PRMA* is its simple form of priority support for continuous voice traffic over random data traffic. Unlike the reservation policy proposed in this paper, priorities in *PRMA* do not have any impact on the resource allocation decision, it only affects the need for issuing reservations. The TDMA-based multiple access protocols proposed for wireless ATM networks [1], [4], are variations of Slotted Aloha and PRMA that give priority for real-time traffic over non-real time traffic. However, the different priority levels associated with various services did not have any impact on the slot allocation decision, it only affects the opportunity to issue reservation requests. Furthermore, these protocols are not optimal in any sense. In [2], the authors proposed a dynamic bandwidth allocation protocol for multimedia traffic in broadband satellite networks. The drawbacks of this protocol are twofold; first the lack of supporting multiple classes of services with different levels of priority. Second, the long-run optimization of the weighted sum of the distortion rate of video traffic, packet dropping probability of voice traffic, and buffer lengths of data traffic. The authors asserted that solving the long-run optimization problem was found to be computationally intensive and, hence, real-time implementation proved to be practically infeasible. Although, priority classes have been considered in [5], [6], it only affected the ability of mobile nodes to make reservations, while the slot allocation decision was independent of priorities. Therefore, in this paper we introduce a novel priority-based slot allocation strategy that gives equal chance for all traffic streams to attempt reservations, and then leaves it for the allocation algorithm to determine the optimal partitioning. To this end, we introduce a cost function that depends on the *system state* (given by the buffer lengths of various traffic streams) and the priorities of various services. To the best of our knowledge, there is no prior known

[1] "©2001 HRL Laboratories, LLC. All Rights Reserved"

work that adopts a cost function having the following unique features: 1) Supports multiple classes of priorities and 2) Determines the optimal allocation on a frame-by-frame basis.

The paper is organized as follows: In section II, the assumptions underlying this study are introduced. This is followed by the protocol description in section III. Afterwards, the dynamic slot reservation problem is formulated and solved in section IV. The simulation results and discussion are given in section V. Finally, the conclusions are drawn in section VI.

II. ASSUMPTIONS

In order to investigate the dynamic bandwidth reservation problem at hand, we make the following assumptions and introduce appropriate notations:

1. Assume a cellular network structure consisting of one base station (BS) that serves n mobile users.
2. Each mobile node is assumed to generate multiple traffic streams of various types corresponding to voice and data applications.
3. Each traffic stream has its own priority class.
4. Assume three classes of priority, namely *High priority* services, *Medium priority* services and *Low priority* services.
5. Time is slotted.
6. The Frame duration, denoted T, is fixed.
7. Information packets are of *fixed* length.
8. Guard bands are crucial in order to keep the nodes in the network time-synchronized. More specifically, the slot duration is assumed to be larger than packet duration by an interval equal to a guard band. These bands are essential to compensate for arbitrary delays incurred by transmitted packets due to signal propagation delays or clock drifts.
9. We assume that each mobile node has a buffer dedicated for each traffic stream. Each buffer is responsible for storing packets that arrive during a frame until it is transmitted in a future frame. In case of buffer overflow, arriving packets are dropped and treated as lost packets.
10. Nodes' mobility is not considered in this study. Hence, the effect of hand-offs is not investigated in this paper.

III. PROTOCOL DESCRIPTION

In this section we introduce a multiple access protocol that facilitates communication between the mobile terminal and the BS according to the following requirements: efficient bandwidth utilization, priority support, hand-off support, and QoS support for multimedia applications. In addition, the candidate protocol is expected to support short bursts of data that have less stringent quality of service requirements. Accordingly, we propose a TDMA-based multiple access protocol that has the following two types of information slots:

- *Contention-free Slots:* provide packet level QoS for *multimedia applications* and support *hand-off* of mobile nodes.
- *Contention-based Slots:* carry random access traffic (*e.g.* short message bursts) with less stringent quality of service constraints.

The frame structure of the proposed protocol is given in Figure 1. Each frame is divided into two sub-frames, namely *Reservation Sub-Frame* and *Information Sub-Frame*. As shown in Figure 1, the reservation sub-frame is divided into n slots where each slot is dedicated to one of the mobile nodes. Furthermore, each reservation slot is divided, in turn, into *mini-slots*. These mini-slots are dedicated to carry the reservation requests of various traffic streams to the base station. The reservation requests contain the *network state information* necessary for the reservation algorithm to allocate *contention-free* slots as described in the next section. It is worth mentioning that due to the propagation delays to and from the base station, in addition to the processing delays, reservations requested in any frame are granted in the *next* frame, *i.e.* reservations at the beginning of frame K request slots in frame $K+1$.

On the other hand, the information sub-frame is divided into information slots. Each information slot is assumed to carry one packet. As pointed out earlier, these information packets are expected to carry three possible traffic types, namely multimedia and hand-off traffic streams with strict QoS constraints, in addition to bursty traffic with less strict requirements. Therefore, N contention-free slots are dedicated to serving multimedia and hand-off traffic streams, while the remaining slots are contention-based in order to serve random access traffic. In this paper we focus on the contention-free slots, more specifically, we develop the algorithm that optimally allocates the N contention-free slots to various traffic streams depending on the network state and their respective priorities. Finally, it is worth mentioning that the transmission order of various traffic streams depends solely on the service type and priority.

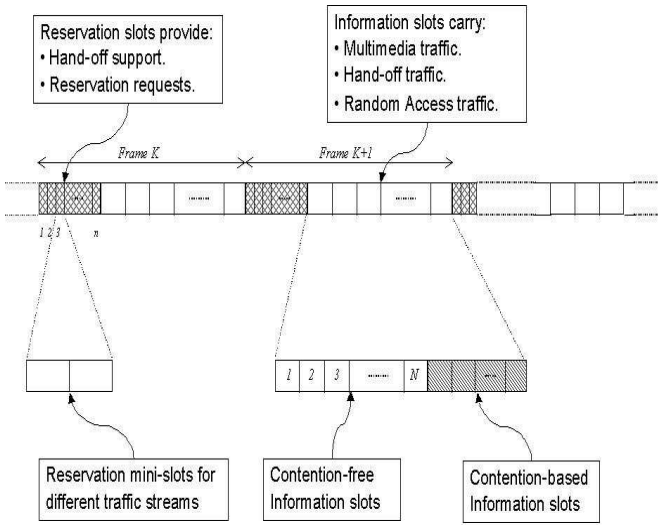

Fig. 1. TDMA Frame Structure

IV. DYNAMIC SLOT RESERVATION

As pointed out in section I, the dynamic bandwidth reservation problem has been addressed earlier in the literature. The approach we take here is fundamentally different from previous approaches due to the following features:

- Supporting multiple classes of priority.
- Solving the optimization problem on a *frame-by-frame* basis.
- Introducing a novel cost function that is related to queuing delays and packet losses.

In this section, we formulate the dynamic slot reservation problem as a *constrained optimization* problem [7]. Our objective is to construct a cost function that depends on the *system state* at the beginning of each frame, along with services' priorities, and hence can be optimized on a *frame-by-frame* basis. Furthermore, this cost function should be carefully chosen to reflect the impact of the allocation decision on the packet level quality of service parameters of interest, namely packet delays and losses. Accordingly, we attempt to solve the following constrained optimization problem at the end of the reservation period of frame K,

$$\min_{B_{ij}(K)} \quad \sum_i \sum_j [L_{ij}(K+1) - B_{ij}(K)]\ P_{ij} \qquad (1)$$

s.t.

$$\sum_i \sum_j B_{ij}(K) = N$$

where,
$B_{ij}(K)$ = Number of slots allocated to traffic stream j generated by node i in frame K and utilized in frame *K+1*.
P_{ij} = Priority level of traffic stream j generated by node i.
N = Number of contention-free slots per frame.
$L_{ij}(K+1)$ = Predicted buffer length of traffic stream j generated by node i in frame *K+1* and is given by,

$$L_{ij}(K+1) = L_{ij}(K) - B_{ij}(K-1) + A_{ij}(K) \qquad (2)$$

where $L_{ij}(K)$ = Buffer length of traffic stream j generated by node i in frame K and $A_{ij}(K)$ = Expected number of packet arrivals from stream j in node i during the interval *[K, K+1]*. In this paper, we adopt a simple traffic prediction algorithm, where $A_{ij}(K)$ is taken to be the product of the mean packet arrival rate and frame duration. Studying the impact of traffic prediction on the optimal reservation policy is out of the scope of this paper and is a subject of future research.

In order to interpret the above cost function, it is essential to define the *"Deviation"* function for each traffic stream generated in each mobile node, denoted D_{ij}, as follows:

$$D_{ij}(K+1) = L_{ij}(K+1) - B_{ij}(K) \qquad (3)$$

For each traffic stream, the deviation function represents the difference between the predicted buffer length and the number of slots to be allocated, since reservation requests in frame K are granted in frame *K+1*. Ideally, this difference should be zero if the N contention-free slots can serve all backlogged packets awaiting transmission. Unfortunately, this is not normally the case due to the upper bound on frame duration (enforced by delay constraints), lower bound on slot duration (enforced by synchronization constraints), and the inaccuracy in traffic prediction. Therefore, solving the optimization problem becomes necessary when the sum of the predicted buffer lengths $L_{ij}(K+1)$, is strictly greater than N. Thus, it is evident that minimizing this cost function is equivalent to minimizing the deviation function of each traffic stream depending on its priority level. It can be shown that packet queuing delays and packet losses are related to the deviation functions of various traffic streams. More specifically, reducing the deviation function implies reducing the expected number of waiting frames and, hence, reducing the queuing delays. Similarly, it can be shown that minimizing a weighted sum of the deviation functions reduces the probability of buffer overflow and, hence, leads to reducing the packet loss ratio.

The structure of the cost function in (1) makes it amenable to analytical optimization techniques. Therefore, we solved it using *Lagrange multipliers* [7], and the solution was found to have the following form:

$$B_{ij}(K) = \frac{L_{ij}(K+1)}{(1+\frac{\mu}{P_{ij}})} \quad (4)$$

where μ is a constant that is determined, via an iterative search algorithm, from the constraint in (1). This, in turn, reduces computational complexity dramatically compared to search-based optimization techniques [7]. This is of paramount importance since the slot allocation algorithm operates in real-time and processing delays could have significant impact on performance.

Finally, we measure the performance of other heuristic policies relative to the aforementioned optimum. Weighted Fair Queuing *(WFQ)* is a rate-based packet scheduling strategy introduced in [8], [9]. We compare the performance of the *optimal* policy to a heuristic policy based on *WFQ*, where priorities are devised as weights. This policy is called *Prioritized Proportional Allocation (PPA)*. The essence of this policy is to allocate the contention-free slots in proportion to the buffer lengths of various traffic streams weighted by their respective priorities. Thus, according to *PPA*, the number of slots allocated to stream j generated by node i would be given by,

$$B_{ij}(K) = \frac{N}{\sum_i \sum_j P_{ij} L_{ij}(K+1)} \; P_{ij} L_{ij}(K+1) \quad (5)$$

V. SIMULATION RESULTS

Number of mobile Nodes (n)	10
Number of Traffic Streams per Node	2
Data Buffer Size	100 packets
Voice Buffer Size	40 packets
Frame Duration (T)	20 msec
Number of Slots per Frame (N)	20,30,...,100
Number of Priority Classes	3
Simulation Time	800 sec

Table 1: System Parameters

The performance of the optimal reservation policy was investigated assuming the numerical parameters given in Table 1. We developed an OPNET-based model for a TDMA-based wireless network that consists of one base station serving a set of *n=10* mobile nodes. Each mobile node is assumed to generate two traffic streams that represent voice and data applications. Even though we limit our attention to this simple system, the results presented here are general and can be applied to any real system with arbitrary number of base stations and mobile nodes. In this study, we assume that voice traffic is generated according to an Interrupted Poisson Process (*IPP*), while data traffic is assumed to be Poisson.

The queuing delays and packet losses for each priority class are compared under the reservation policies introduced in section IV. The motivation behind this is to emphasize the *"Service Differentiation"* feature of the proposed optimal policy. First, the average queuing delays, for each priority class, are plotted versus the number of slots per frame as shown in Figure 2. It is evident that the optimum reservation policy outperforms the *PPA* policy for all priority classes, especially at heavy and moderate system loads. It can be easily noticed that the performance improvement achieved by the optimum is slim for low priority services, yet significant for high priority services. This is caused by the structure of the cost function in (1) that gives more weights to high priority services. This performance gain gradually diminishes as N increases (or load decreases) due to the fact that large number of resources gives more flexibility for simple heuristic policies to perform better and asymptotically approach the performance of the optimum.

The performance improvement of high priority services, compared to low and moderate priority services, can also be noticed. For voice traffic, the average queuing delays of high priority services is almost half the delays of medium priority services for a wide range of heavy to moderate loads. Also, the relative performance of medium and low priority services is given by 40% at heavy loads and slightly decreases to 35% at moderate loads. In Figure 3, the packet loss ratio for data traffic is plotted versus the number of slots per frame. It can be noticed that the packet loss ratio for high priority services outperforms medium priority services significantly at heavy and moderate load conditions. The relative performance of medium to low priority services is given by a factor of 37% at heavy loads and slightly decreases at moderate loads. It is worth mentioning that the queuing delays for data traffic and packet losses for voice traffic show trends similar to those given in Figures 2 and 3.

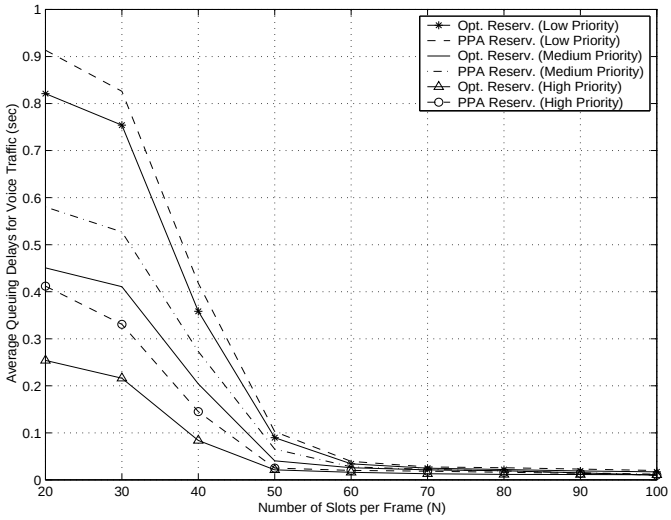

Fig. 2. Voice Queuing Delays for each Priority Class

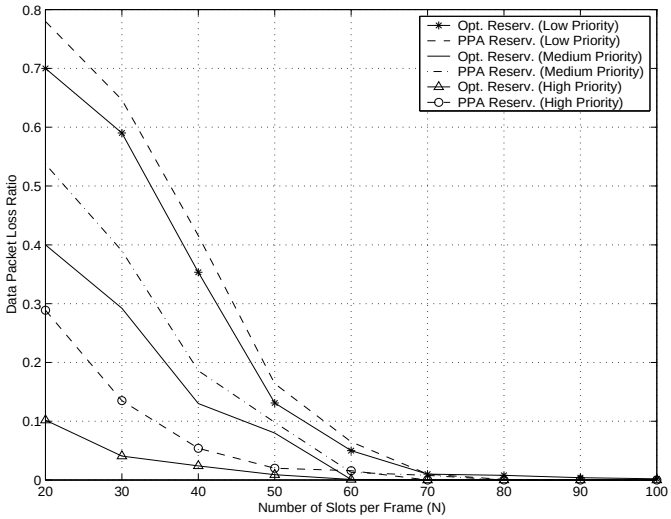

Fig. 3. Data Packet Loss Ratio for each Priority Class

VI. CONCLUSIONS

In this paper, we investigated the problem of dynamic bandwidth reservation for prioritized multimedia services over the uplink of TDMA-based wireless networks. The results of this paper are general and applicable to cellular, satellite and personal communication networks. The proposed multiple access protocol supports multiple levels of priorities. In addition, it minimizes a unique cost function, that depends on deviation functions and priorities, on a frame-by-frame basis. This, in turn, reduces the computational complexity significantly compared to long-run optimization techniques introduced earlier in the literature. Furthermore, analytical optimization techniques were applicable to this problem due to the simple structure of the cost function. Finally, this protocol supports wide variety of traffic classes, namely multimedia, short message bursts and hand-offs.

References

[1] J. Sanchez et al., "A Survey of MAC Protocols Proposed for Wireless ATM," *IEEE Network*, pp. 52-62, Nov. 1997.

[2] M. Hadjitheodosiou and E. Geraniotis, "Dynamic Bandwidth Allocation For Multimedia Traffic in TDMA Broadband Satellite Networks," In *Proceedings of the AIAA International Communications Satellite Systems Conference*, Feb. 1998.

[3] D. Goodman et al., "Packet Reservation Multiple Access for Local Wireless Communications," *IEEE Transactions on Communications*, vol. 37, No. 8, pp. 885-890, August 1989.

[4] I. Akyildiz et al., "Medium Access Control Protocols for Multimedia Traffic in Wireless Networks," *IEEE Network*, pp. 39-47, July 1999.

[5] F. Khan and D. Zeghlache, "Priority-based multiple access (PBMA) for statistical multiplexing of multiple services in wireless PCS", *Proceedings of ICUPC*, Cambridge, MA, pp. 17-21, 1996.

[6] T. Yum et al., "Analysis of a Dynamic Reservation Protocol for Interactive Data Services on TDMA-Based Wireless Networks," *IEEE Transactions on Communications*, vol. 47, No. 12, pp. 1796-1801, Dec. 1999.

[7] B. Gottfried and J. Weisman, *Introduction to Optimization Theory*. New Jersy:Prentice-Hall Inc., 1973.

[8] A. Demers, S. Keshav, and S. Shenkar, "Analysis and Simulation of a Fair Queuing Algorithm," *Internet. Res. And Exper.*, vol. 1, Dec. 1990.

[9] A. Parekh and R. Gallager, "A Generalized Processor Sharing Approach to Flow Control in Integrated Services Networks: The Single-Node Case," *IEEE/ACM Transactions on Networking*, vol. 1, No. 3, pp. 344-357, June 1993.