A Nonpreemptive Priority-Based Access Control Scheme for Broadband *Ad Hoc* Wireless ATM Local Area Networks

Dr-Jiunn Deng and Ruay-Shiung Chang, *Member, IEEE*

***Abstract*—The DQRUMA (distributed-queuing request update multiple access) protocol has been considered as an access protocol for the BAHAMA (broadband *ad hoc* wireless ATM local area network). However, it cannot support the service discipline of integrated multimedia traffic since it does not include any priority and access control policy. In this paper, we propose a nonpreemptive priority-based access control scheme for DQRUMA protocol. Under such a scheme, modifying the CSMA/CA protocol in the contention period supports many levels of priorities such that user mobility (handoff) can be supported in BAHAMA. Besides, the proposed transmit-permission policy and adaptive bandwidth allocation scheme provide various QoS (quality-of-service) guarantees while maintaining high bandwidth utilization. Simulations show that it provides a good performance in *ad hoc* wireless ATM LAN environments.**

***Index Terms*—CSMA/CA, DQRUMA, priority, QoS, wireless ATM LAN.**

I. Introduction

ASYNCHRONOUS TRANSFER MODE (ATM) technology is anticipated as a multiplexing and switching standard for telecommunications [1], [2]. It has also been conceived as a broadband multiservice technology [3]. In the meantime, wireless communications have gained global acceptance and popularity in both voice-oriented and data-oriented markets. Consequently, many laboratories and standardization groups have focused on WATM (wireless ATM) technologies to extend ATM from the LAN/WAN infrastructure toward the wireless users [4]–[13]. In 1995, the AT&T Bell Laboratories' Karol *et al.*proposed a wireless ATM LAN prototype called BAHAMA (a broadband *ad hoc* wireless ATM local area network) [14], which is capable of supporting mobile users with multi-Mbits/s access rates. Unlike conventional star shaped WATM network architectures [15], [16], the BAHAMA employed an *ad hoc* architecture because of its low cost, plug-and-play, flexibility, and minimal human interaction requirements [17].

MAC protocols that aim to carry multimedia traffic must be able to meet the differing requirements of each of the different traffic classes. Time-bounded data, such as voice and video, are useless unless arrived in time. For example, voice traffic is delay-sensitive, with listeners noticing an irritating delay if the interpacket delay reaches 100 ms [18], even with echo cancellation. On the other hand, asynchronous data, such as e-mail or file transfer, can be delayed without causing any inconvenience. To access the bandwidth, conventional wireless MAC protocols can be classified into four categories [19]–[26]—static assignment protocol, random assignment protocol, conflict-free protocol, and scheduled access protocol.

DQRUMA protocol [26], proposed by Karol *et al.* in 1995, was a scheduled access protocol. It was considered as an access protocol for the BAHAMA [14]. In DQRUMA protocol, mobile users need to send requests to the base station only for packets that arrive to an empty buffer. For packets that arrive to a nonempty buffer, transmission requests are placed collision-free by piggybacking the requests with packet transmissions. However, it cannot support service discipline of integrated multimedia traffic since admitted inactivated traffic and new connection requests have the same priority to contend the RA (request access) slots for reservation [27]–[29]. Besides, it does not include any access control policy for various types of multimedia traffic and user mobility (handoff) was not taken into consideration as well.

A simple access control scheme of DQRUMA protocol in wireless ATM LAN has been proposed in [29]. Voice services are given higher priority and periodical assignment. Guaranteed assignment keeps data traffic from starvation. Performance for transporting integrated multimedia traffic is also examined. However, end-to-end QoS requirements cannot be satisfied in this scheme. Further, this scheme does not include any priority scheme to support user mobility, nor does it apply any bandwidth allocation strategy for handoff calls. In this paper, we propose an advanced, pragmatic, and yet more complete nonpreemptive priority based access control scheme for DQRUMA protocol. By modifying the CSMA/CA protocol in the contention period, the protocol provides many levels of priorities such that user mobility (handoff) can be supported. Besides, the proposed transmit-permission policy and adaptive bandwidth allocation scheme not only separate admitted inactivated users from newly requesting access users, but also provide various QoS (quality-of-service) guarantees while maintaining high bandwidth utilization. The proposed scheme is performed at each PBS (portable base station) in a distributed manner. Besides, this scheme can be implemented in a broad class of algorithms with relatively minor modifications.

Manuscript received September 23, 1999; revised February 23, 2000. This work was supported by ROC MOE under Contract 89-H-FA07-1-4.

D.-J. Deng is with the Department of Electrical Engineering, National Taiwan University, Taipei, Taiwan (e-mail: djdeng@cobra.ee.ntu.edu.tw).

R.-S. Chang is with the Department of Computer Science and Information Engineering, National Dong Hwa University, Hualien 974, Taiwan (e-mail: rschang@csie.ndhu.edu.tw).

Publisher Item Identifier S 0733-8716(00)07133-X.

Simulation results show that it provides a good performance in *ad hoc* wireless ATM LAN environments.

The remainder of this paper is organized as follows. Section II describes the proposed scheme in detail. Simulation results are shown in Section III. Section IV concludes this paper.

II. The Proposed Nonpreemptive Priority-Based Access Control Scheme

In this section, we describe the proposed scheme in detail. Our method can be divided into three parts: enforcing priorities for request access, the packet transmit-permission policy, and the adaptive bandwidth allocation strategy.

A. Enforcing Priorities for Request Access

Since a mobile travels while a connection is alive, the QoS might degrade because of some physical constraints. The problem will become even more challenging because recent wireless networks have been implemented using architecture based on small-size cells (i.e., micro-cells or pico-cells) to obtain higher transmission capacity and to achieve better performance. In most of the solutions, bandwidth is reserved for handoff mobiles in advance to reduce the dropping probability [30]. Some improvements have also been discussed [31]. However, when reserved and unused, the bandwidth is simply wasted. This is where priority schemes come in. In this section, we propose a novel method to modify the CSMA/CA protocol to get many levels of priorities such that handoff requests might be given higher priority over new connection requests. The method is simple, efficient, flexible, scalable, and also easy to implement. It could be used as the random access protocol for the RA channel in the DQRUMA protocol.

The collision avoidance portion of CSMA/CA is performed through a random backoff procedure. The random backoff time is an integer value that corresponds to a number of time slots. Initially, a station computes a backoff time in the range 0–7. If a station with a frame to transmit initially senses the channel to be busy, it waits until the channel becomes idle, and then the station decrements its backoff timer until the medium becomes busy again or the timer reaches zero. If the timer has not reached zero and the medium becomes busy, the station freezes its timer. When the timer finally decrements to zero, the station transmits its frame. If two or more stations decrement to zero at the same time, a collision will occur, and each station will have to generate a new backoff time in the range 0–15. For each retransmission attempt, the backoff time grows as $\lfloor ranf(\,) \cdot 2^{2+i} \rfloor \bullet$ Slot_Time, where i is the number of consecutive times a station attempts to send a frame, $ranf(\,)$ is a uniform variate in (0, 1), and $\lfloor x \rfloor$ represents the largest integer less than or equal to x. For more information about CSMA/CA protocol, see [23].

The basic idea of this method is that priority access to the wireless medium is controlled through different backoff time. The shorter backoff time a mobile waits, the higher priority this mobile will get. Therefore, we first change the backoff time generation function to $\lfloor ranf(\,) \cdot (2^{2+i}/2) \rfloor$ for high priority mobiles and $2^{2+i}/2 + \lfloor ranf(\,) \cdot (2^{2+i}/2) \rfloor$ for low priority mobiles. This technique divides the random backoff time into two parts: $0 \sim (2^{2+i}/2) - 1$ and $(2^{2+i}/2) \sim 2^{2+i} - 1$. The high priority mobiles use the former, and the low priority mobiles use the latter. For example, initially, the high priority mobiles generate a backoff time in the range 0–3, and the low priority mobiles generate a backoff time in the range 4–7. Thus, the former will have higher priority in contending the channel.

TABLE I
EXAMPLES OF BACKOFF TIME OF INDIVIDUAL TRAFFIC

Types of requests (k, m, n) / Backoff slot numbers / Consecutive times (i)	1st	2nd	3rd	4th
Real-time handoff traffic (0, 1, 1)	**0 – 3**	**0 -7**	**0 – 15**	**0 – 31**
Admitted inactivated video traffic (1, 1, 1)	**4 – 7**	**8 - 15**	**16 – 31**	**32 - 63**
Non-real-time handoff traffic New request traffic (2, 2, 1)	**8 – 15**	**16 - 31**	**32- 63**	**64 – 127**

Then, to support multiple-level priorities, the backoff time generation function is changed to $\lfloor ranf(\,) \bullet 2^{1+i} \rfloor + k \bullet 2^{1+i}$, where k is the level of priority. The lower k a mobile has, the higher priority this mobile will get.

However, for fixed backoff range, the probability of collisions in the same priority level will increase if the number of contended mobiles with the same priority increases. In other words, this scheme should have the ability to extend or reduce the backoff range arbitrarily. That is, we allow different backoff ranges for different priority levels in this scheme. To this end, we change the backoff time generating function to $\lfloor ranf(\,) \bullet 2^{m+i} \rfloor + k \bullet 2^{n+i}$, where k is the level of priority, and m and n are the parameters used to decide the number of slots in individual priority levels and the number of slots between each priority levels, respectively. In this paper, the real-time handoff traffic requests have the highest priority among all other requests, and the second priority class is admitted inactivated video traffic. The new requests and handoff data traffic will have the lowest priority level, as shown in Table I. Note that we give wider range to the lowest priority level since there is more traffic that contends in this priority level.

It is also noteworthy that when a mobile decrements its backoff timer and the medium becomes busy, the mobile freezes its timer. This means that a mobile will raise its priority automatically after several times of transmission failure. Hence, starvation will not occur in this method.

B. The Packet Transmit-Permission Policy

One of the important challenges of traffic control in ATM networks is how to decide whether a network accepts a new connection or not, i.e., the packet transmission policy, which is also used by the PBS to determine which mobile gets permission to transmit a packet. In this section, we propose a packet transmit-permission policy for DQRUMA protocol to support integrated multimedia traffic. Our scheme is an enhanced version of the transmitting policy originally proposed by Chang *et al.* in [32]. However, our scheme is simpler and more efficient. In addition, we also take handoff traffic into consideration. Under such a scheme, all voice traffic satisfies their jitter constraint, all

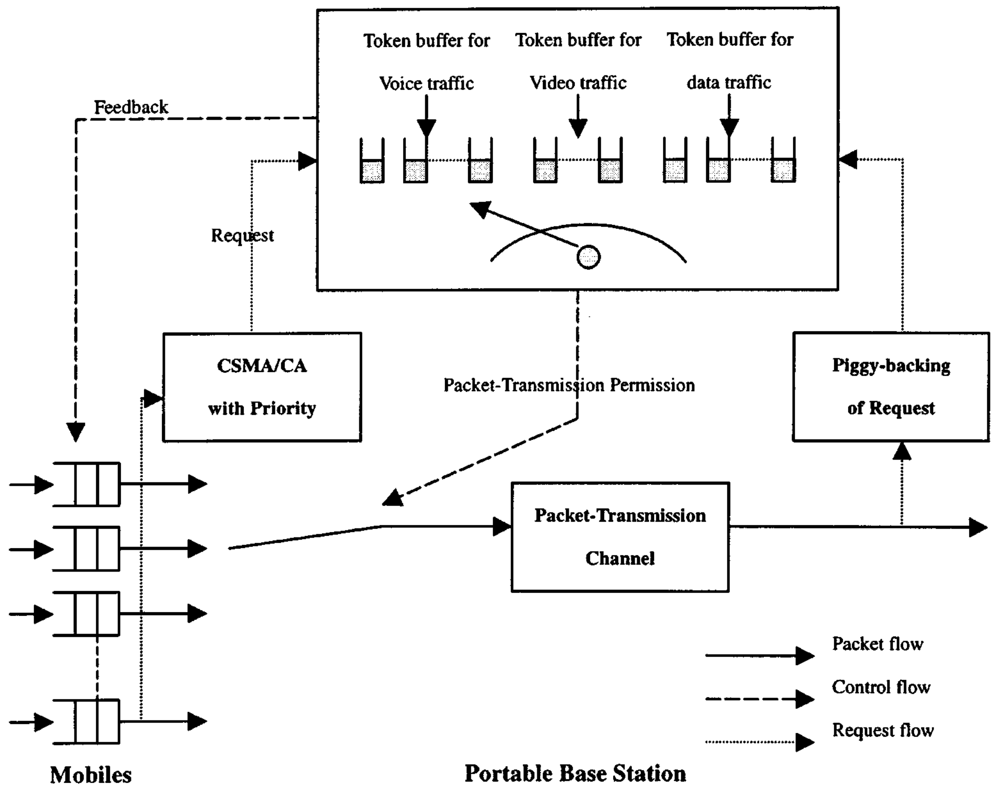


Fig. 1. Proposed packet transmit-permission policy.

video traffic satisfies their delay constraints, and the remaining bandwidth is shared by data traffic fairly and efficiently.

Three types of traffic are considered. The first is voice traffic which is characterized by two parameters (r_c, δ), where r_c is the rate of the source and δ is the maximum tolerable jitter (packet delay variation) for this stream. The second is video traffic which is characterized by three parameters (r_v, β, d), where r_v is the average rate of the source, β is the maximum burstiness of the source, and d is the maximum tolerable delay (packet transfer delay) for this stream. The third type is pure data.

The PBS implements a token buffer for each source. In token buffers for voice sources, the smaller the maximum jitter, the higher the priority. In token buffers for video sources, the priority is assigned in a similar way. That is, the one with the smallest maximum delay constraint has the highest priority among all video sources. We depict the packet transmit-permission policy in Fig. 1. At the beginning, when the PBS starts to transmit a packet, it performs the following tasks.

1) The PBS first scans the token buffers of voice sources according to the preset priority order. If a token is found, it removes one from this token buffer and transmits a packet for this voice source. Then, the PBS generates the next token for this voice source after $(1/r_c) - t_p$ second if the piggyback was set while transmitting the packet, where t_p is the time to transmit a packet.
2) If no tokens are found in the token buffers of voice sources, the PBS continues to scan the token buffers for video sources according to the preset priority order. If a token is found, it transmits a packet for this video source. It will not remove the token if the piggyback was set while transmitting this packet. If the piggyback was not set and it is not the last packet (end-of-file) either, the PBS removes the token, and then generates the next token for this video source after η seconds if there is no new token generated for this video source within η, where η will be defined later.
3) When there is no token found in the token buffers of voice and video sources, the PBS scans the token buffers for pure data sources using both FCFS (first-come first-serve) and round robin scheduling algorithms. If a token is found in the token buffers of pure data sources, it transmits a packet for this pure data source without removing the token until the piggyback was not set.
4) If there is no token found in all token buffers, the PBS will not know which, if any, of the mobiles have packets to transmit, then, the uplink Xmt channel will be idle in the next time slot. However, to avoid wasting the valuable transmission time, the PBS uses the downlink Xmt_Perm channel to announce that the next uplink Xmt channel will be converted into multiple RA channels, and the next downlink Xmt channel is similarly converted into multiple ACK channels.

In the following theorems, we provide sufficient conditions for all the voice packets to satisfy their maximum jitter constraint and for all the video packets to satisfy their maximum delay constraint. Under such conditions, the admission control is simple. If all the inequalities in Theorem 3.2.2 are satisfied, then the request of a new video source is admitted. Otherwise, it is rejected. For the request of a new voice source, it requires more computation. All the conditions in Theorem 1 and Theorem 2 need to be examined at the same time. The following proofs use mathematical induction [35] and calculus [32]–[34].

Assume there are n_c voice sources, indexed by $i = 1, \ldots, n_c$, and n_v video sources, indexed by $j = 1, \ldots, n_v$. Denote (r_{ci}, δ_i) as the traffic parameters of the ith voice source, (r_{vj}, β_j, d_j) as the traffic parameters of the jth video source, and π_i as the time needed for handoff for source i.

Theorem 1: Let $\delta_1^* = t_p$ and $\delta_i^* = t_p + \sum_{k=1}^{i-1} \lceil r_{ck}/r_{ci} \rceil \bullet t_p$, $i = 2, \ldots, n_c$ and t_p is the time to transmit a packet. If $\delta_i^* < (1/r_{ci})$ and $\delta_i^* \le \delta_i$ for all $i = 1, 2, \ldots, n_c$, then all the packets generated by new-call voice sources meet their jitter constraints. Furthermore, if $\delta_i^* + \pi_i < (1/r_{ci})$ and $\delta_i^* + \pi_i \le \delta_i$ for the ith source which is handoffed from other cells, then the packet generated by the ith source after handoff meets its jitter constraint.

Proof: We first prove the handoff part. Suppose that the first token generated from the ith voice source after handoff from other cells has a maximum waiting time $\overline{\delta}_i$. We want to prove that $\overline{\delta}_i \le \delta_i$ for $1 \le i \le n_i$. For $i = 1$, $\overline{\delta}_1 \le \pi_1 + t_p \le \pi_1 + \delta_1^* \le \delta_1$, which establishes the induction basis.

Suppose our induction hypotheses hold up to the $(i-1)$th voice sources, i.e., $\overline{\delta}_{j-1} \le \delta_{j-1}$ for $1 \le j \le i-1$. Now we consider the ith voice source. Let the instant of the beginning of handoff be at time 0. Assume that $\overline{\delta}_i > \delta_i^* + \pi_i$. Then it means that up to time $\delta_i^* + \pi_i$ the channel must be serving all the voice sources from 1 to $i-1$. Since the total amount of packets that can be served within $(0, \delta_i^* + \pi_i)$ for these $i-1$ voice sources is at most $\sum_{k=1}^{i-1} \lceil r_{ck} \bullet (\delta_i^* + \pi_i) \rceil$. Hence, the total amount of time to serve these packets is bounded above by $(\sum_{k=1}^{i-1} \lceil r_{ck} \bullet (\delta_i^* + \pi_i) \rceil + 1) \bullet t_p$, and since $\delta_i^* + \pi_i < (1/r_{ci})$, we have

$$\left(\sum_{k=1}^{i-1} \lceil r_{ck} \bullet (\delta_i^* + \pi_i) \rceil + 1 \right) \bullet t_p < \left(\sum_{k=1}^{i-1} \left\lceil \frac{r_{ck}}{r_{ci}} \right\rceil + 1 \right) \bullet t_p = \delta_i^* \le \delta_i^* + \pi_i.$$

This contradicts our assumption that $\overline{\delta}_i > \delta_i^* + \pi_i$. Hence, $\overline{\delta}_i \le \delta_i^* + \pi_i \le \delta_i$.

Consequently, by the principle of induction, our statement is true. Similarly, we can prove that all the packets generated by new-call voice sources will meet their jitter constraints in the same way. Q.E.D.

Let

$$\overline{\beta}_0 = t_p \bullet (n_c + 1), \quad \overline{r}_{v0} = t_p \bullet \sum_{i=1}^{n_c} r_{ci},$$

$$\overline{\beta}_j = t_p \bullet (\beta_j + 1),$$

$$\overline{r}_{vj} = t_p \bullet r_{vj}, \quad d_1^* = \eta_1 + \frac{\overline{\beta}_0 + \overline{\beta}_1}{1 - \overline{r}_{vo}},$$

and

$$d_j^* = \eta_j + \frac{\sum_{k=0}^{j} \overline{\beta}_k + t_p \bullet \sum_{k=1}^{j-1} (r_{vk} \bullet d_k^*)}{1 - \sum_{k=0}^{j-1} \overline{r}_{vk}},$$

where $j = 2, \ldots, n_v$.

Theorem 2: If $\sum_{k=0}^{n_v} \overline{r}_{vk} \le 1$ and $d_j^* \le d_j$ for all j, $1 \leqq j \leqq n_v$, then the delay constraints are satisfied for all the new-call video sources. Furthermore, if $d_j^* - \eta_j \le d_j - \pi_j$ for jth source which is handoffed from other cells, then the packet generated by the jth source after handoff meets its delay constraint.

Proof: Consider a nonnegative, left limited, and right continuous stochastic process $A \equiv \{a(t), t \ge 0\}$. Let $A(t_1, t_2) = \int_{t_1}^{t_2} a(t)\,dt$. We say that A is (β, r_v)-upper constrained if $A(s, t+s) \le r_v t + \beta$ for all $s, t \ge 0$. Similarly, A is (β, r_v)-lower constrained if $A(s, t+s) \ge r_v t + \beta$ for all s, $t \ge 0$. Since the number of departures in $(t_1, t_2]$ from a (β, r_v)-leaky bucket is bounded above by $\beta + \lceil r_v(t_2 - t_1) \rceil$, the departure process from a (β, r_v)-leaky bucket is $(\beta + 1, r_v)$-upper constrained.

Now consider the first video source. Let $C_1 \equiv \{c_1(t), t \ge 0\}$ be the stochastic process that denotes the available bandwidth to the first video source at time t. If the channel is available to the first video source at time t, then $c_1(t) = 1$. Otherwise, $c_1(t) = 0$.

As mentioned above, the maximum number of packets from the n_c voice sources that can be served in $(t_1, t_2]$ is at most $\sum_{i=1}^{n_c} \lceil r_{ci}(t_2 - t_1) \rceil$. Hence, the bandwidth that is available to the first video source in $(t_1, t_2]$ is at least $t_2 - t_1 - t_p \bullet \{1 + \sum_{i=1}^{n_c} [r_{ci}(t_2 - t_1) + 1]\}$.

Thus, $C_1(t_1, t_2) \ge (1 - t_p \bullet \sum_{i=1}^{n_c} r_{ci})(t_2 - t_1) - t_p(n_c + 1)$. That is, C_1 is $(\overline{\beta}_0, 1 - \overline{r}_{v0})$-lower constrained.

Let $A_1 \equiv \{a_1(t), t \ge 0\}$ be the amount of workload of video source 1 that arrives at the channel at time t. Since the number of departures in $(t_1, t_2]$ from the first video traffic is $(\beta_1 + 1, r_{v1})$-upper constrained, we have $A_1(t_1, t_2) \le t_p \bullet [r_{v1}(t_2 - t_1) + \beta_1 + 1]$. This shows A_1 is $(\overline{\beta}_1, t_p r_{v1})$-upper constrained.

Consider an instant after the last packet was sent (but not the EOF packet) by the first video source. Mark the instant as time 0. Let $q_1(t)$ be the amount of backlogged workload from the first video source in the channel at time t, we have $q_1(0) = 0$. Since the next token for the first video source will be generated at time η_1 at the latest. We have $q_1(t) = A_1(0, t) - C_1(\eta_1, t)$. Note that the delay for an arrival at time t is bounded by the amount of time needed to deplete $q_1(t)$, and the time to deplete $q_1(t)$ is bounded by $\inf\{d \ge 0\colon A_1(0, t) - C_1(\eta_1, t+d) \le 0\}$. Maximizing over t, we have the following upper bound for the maximum delay:

$$d_1^* = \sup_t \inf\{d \ge 0\colon A_1(0, t) - C_1(\eta_1, t+d) \le 0\}$$

or

$$d_1^* = \sup_t \inf\{d \ge 0\colon A_1(0, t) - C_1(\pi_1, t+d) \le 0\}$$

(for the handoff traffic).

Since $\sum_{k=0}^{n_v} \overline{r}_{vk} \le 1$, we have $\overline{r}_{v0} + \overline{t}_p r_{v1} \le 1$. Applying the upper constraint for A_1 and the lower constraint for C_1, we have

$$d_1^* = \eta_1 + \frac{\overline{\beta}_0 + \overline{\beta}_1}{1 - \overline{r}_{vo}}$$

or

$$d_1^* = \pi_1 + \frac{\overline{\beta}_0 + \overline{\beta}_1}{1 - \overline{r}_{vo}}.$$

This completes the argument for the first video source.

The argument of the jth video source is essentially the same as that for the first video source. However, the lower constraint for the channel needs to be modified since the jth video source only uses the remaining channel after all the voice sources and the first $j-1$ video sources. Since the maximum delay of the kth video source is bounded above by d_k^*, $k = 1, \ldots, j-1$, the number of packets from the kth source that can be served in $(t_1, t_2]$ is bounded above by $\beta_k + [r_{vk}(t_2 - t_1 + d_k^*)]$. Hence, the amount of workload from the kth source that can be served in $(t_1, t_2]$ is bounded above by $[r_{vk}(t_2 - t_1) + \beta_k + 1 + r_{vk} d_k^*] \bullet t_p$. Parallel to the argument for the first video source, the maximum delay of the jth video source is bounded above by

$$\eta_j + \frac{\sum_{k=0}^{j} \overline{\beta}_k + t_p \bullet \sum_{k=1}^{j-1} (r_{vk} \bullet d_k^*)}{1 - \sum_{k=0}^{j-1} \overline{r}_{vk}}$$

or

$$\pi_j + \frac{\sum_{k=0}^{j} \overline{\beta}_k + t_p \bullet \sum_{k=1}^{j-1} (r_{vk} \bullet d_k^*)}{1 - \sum_{k=0}^{j-1} \overline{r}_{vk}}$$

for the handoff traffic. Q.E.D.

Finally, we still need to engineer η_j to complete this scheme. In order to maximum the bandwidth utilization, one should have η_j as large as possible. The largest η_j can be obtained by solving $d_j^* = d_j$. However, larger η_j will lead to unsmooth video traffic. Therefore, we give a higher priority to the admitted inactivated video traffic in the contention period in order to compensate for this shortcoming.

C. The Adaptive Bandwidth Allocation Strategy

In this subsection, we propose an adaptive bandwidth allocation strategy. Our strategy tries to maximize the bandwidth utilization and reduce the handoff dropping probability and blocking probability. It also guarantees a minimum bandwidth for data traffic. In addition, this strategy is simple to implement without any extra computation.

The total bandwidth is divided into three parts: channel I, channel II, and channel III. We allocate channel I to new-call/handoff voice/video traffic and channel II to handoff voice/video traffic. Besides, we allow real-time traffic to use bandwidth exclusively with preemptive priority over data traffic to reduce the dropping and blocking probability. In other words, a new-call voice/video is blocked if there is not enough free bandwidth in channel I, and a handoff voice/video attempt is dropped if no bandwidth is available in both channel I and channel II. Channel III is only reserved for data traffic.

However, after bandwidth is allocated, network conditions may change. Therefore, the proposed strategy can also adjust the amount of allocated bandwidth based on the measured dropping probability, blocking probability, and bandwidth utilization. The algorithm to control the size of the allocated bandwidth is summarized in the following.

```
Function Adaptive_Bandwidth_Allocation:
IF monitored dropping probability >
threshold_D THEN
  IF bandwidth utilization < μ THEN
    size of allocated bandwidth II = min
    {max {size of allocated bandwidth I,
    size of allocated bandwidth II} ×
    up_γ, total bandwidth}
  ELSE
    size of allocated bandwidth II = min
    {max {size of allocated bandwidth I,
    size of allocated bandwidth II} ×
    up_γ, total bandwidth ×
    threshold_up_II}
ELSE
  IF monitored blocking probability >
  threshold_B THEN
    IF bandwidth utilization < μ THEN
      size of allocated bandwidth I = min
      {size of allocated bandwidth I ×
      up_γ, total bandwidth ×
      threshold.1_up_I}
    ELSE
      size of allocated bandwidth I = min
      {size of allocated bandwidth I ×
      up_γ, total bandwidth ×
      threshold.2_up_I}
  ELSE
    IF bandwidth utilization < μ THEN
      size of allocated bandwidth II = max
      {size of allocated bandwidth II ×
      down_γ, total bandwidth ×
      threshold_down_II}
      size of allocated bandwidth I = max
      {size of allocated bandwidth I ×
      down_γ, total bandwidth ×
      threshold_down_I}
```

As the pseudo-code illustrates, the handoff dropping probability is the first measure used to adjust the allocated bandwidth. If the dropping probability over the threshold, threshold_D, and the bandwidth utilization is not good enough (less than the threshold value μ), it implies that there is not so much data traffic. Hence, we increase the size of channel II by a factor up_γ to its maximum (total bandwidth). Otherwise, we guarantee a minimum bandwidth for data traffic by only increasing the size of channel II to the threshold (total bandwidth $\times$ threshold_up_II). Then, we use the blocking probability to adjust the allocated bandwidth of channel I in the same way. That is, to lower dropping probability will get higher priority than to lower blocking probability in adjusting bandwidth allocation. Finally, the allocated bandwidth will be stable in a good situation if the bandwidth over the threshold μ. That is, both dropping probability and blocking probability under the threshold and the bandwidth utilization is above the threshold value μ. This algorithm can be run periodically.

III. Simulations and Performance Evaluation

In this section, we evaluate the performance of the proposed scheme.

A. Simulation Model

The simulation models are built using the Simscript tool [36]. The model represents a cell in the BAHAMA network. Several assumptions have been made to reduce the complexity of the model. First, the "hidden terminal" and "exposed terminal" problems [37] are not addressed in the simulation model. Second, no mobiles operate in the "power-saving" mode. Third, no interference is considered from nearby cells. Finally, traffic error detection and retransmission methods are not considered.

Three types of traffic are considered in the simulation.

1) Pure Data: The arrival of data frames from a mobile's higher- layer to MAC sublayer is Poisson. Frame length is assumed to be exponentially distributed with mean length 1024 octets.

2) Voice Traffic: Voice stream is characterized by two parameters (γ_c, δ), where γ_c is the rate of the source and δ is the maximum tolerable jitter (packet delay variation) for this stream. Frames of voice traffic that are not successfully transmitted within its maximum jitter constraint are assumed to be lost. Each connection duration is exponentially distributed with mean time 3 min.

3) Video Traffic: Video stream is characterized by three parameters (r_v, β, d), where r_v is the average rate of the source, β is the maximum burstiness of the source, and d is the maximum tolerable delay (packet transfer delay) for this stream. We use a source model in [38]. The bit rate of a single source for the nth frame, $\lambda(n)$, is defined by the recursive relation: $\lambda(n) = a\lambda(n-1) + bw(n)$ [bit/pixel], where $a = 0.8781$, $b = 0.1108$, and $w(n)$ is a sequence of independent Gaussian random variables which have mean 0.572 and variance 1. Like voice frames, video frames that are not successfully transmitted within its maximum tolerable delay, d, is assumed to be lost.

Assume video, voice, and data are mixed in the ratio of 1 : 1 : 1. The default values used in the simulation are listed in Table II. The values for the simulation parameters are chosen carefully in order to closely represent the realistic scenarios as well as make the simulation feasible and reasonable.

TABLE II
Default Attribute Values Used in the Simulation

Attribute	Value	Meaning & Explanation
Channel rate	10 Mb/s	Data rate for the wireless channel
Mobiles	10	10 mobile hosts in a base station
RA channel	64 bits	Bandwidth needed for each request
PGBK	8 bits	Bandwidth needed for each piggybacking
r_c	32 kb/s	Voice source data rate
δ	32 ms	Tolerable jitter for voice source
β	5	Maximum burstiness
d	50 ms	Maximum packet delay for video source
threshold_D	0.1	Maximum allowable dropping probability
threshold_B	0.1	Maximum allowable blocking probability
μ	0.8	Minimum bandwidth utilization wanted
up_ r	1.1	Bandwidth allocation each time increases 10%
down_ r	0.9	Bandwidth allocation each time decreases 10%
threshold_up_II	0.9	Channel II uses at most 90% of total bandwidth
threshold.1_up_I	0.8	Channel I uses at most 80% of total bandwidth when blocking rate is too large and bandwidth utilization too low
threshold.2_up_I	0.7	Channel I uses at most 70% of total bandwidth when blocking rate is too large and bandwidth utilization is high enough
threshold_down_II	0.5	Channel II uses at least 50% of total bandwidth
threshold_down_I	0.3	Channel I uses at lease 30% of total bandwidth
handoff probability	0.25	The probability that a mobile moves out of the range of a base station

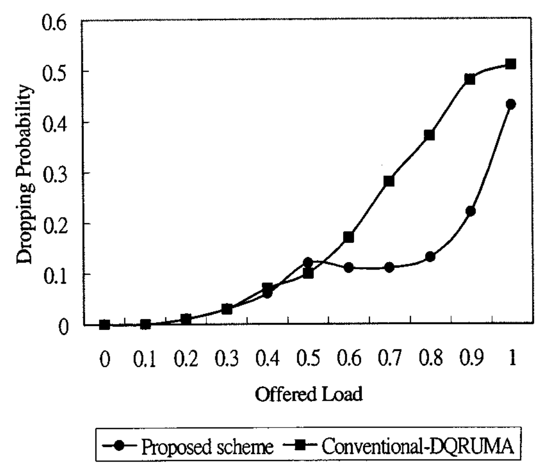


Fig. 2. Dropping probability of real-time handoff connections.

B. Simulation Results

We compare the proposed scheme to the conventional DQRUMA. In the conventional DQRUMA, slotted-ALOHA was adopted as the random access protocol for the RA channel; and a round-robin discipline was chosen as the packet transmission policy, which is used by the base station to determine which mobile gets permission to next transmit a packet. That is, all traffics have the same priority. The admission control scheme in conventional DQRUMA is very simple and intuitive. Assuming there are totally k requests in the request table, if $k \bullet t_p \leq \delta_i$ for the ith voice source or $k \bullet t_p \leq d_j$ for the jth video source, the request of a new voice or video source is admitted. Otherwise, it is rejected. Understandably, the time needed for handoff, π, will be added for the handoff mobile in admission control.

Simulation results are shown below in the form of plots. Figs. 2 and 3 show the dropping probability of real-time handoff connections and blocking probability of real-time new connections for the proposed scheme and conventional DQRUMA. These two figures show the tradeoff between the dropping probability and blocking probability in the proposed scheme. The handoff dropping probability is the first measure used to adjust the allocated bandwidth, and we also allow the handoff real-time traffic to use bandwidth exclusively with preemptive priority over other traffics in the reserved region, channel II. The dropping probability will be kept under the threshold (threshold_D) usually.

Figs. 4 and 5 show average access delay of voice and video traffic, respectively. Note that the average access delay of the

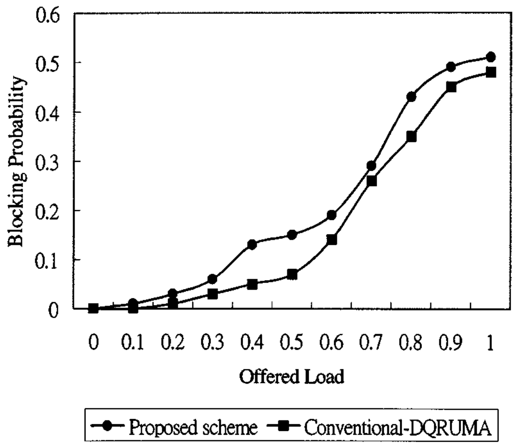


Fig. 3. Blocking probability of real-time new connections.

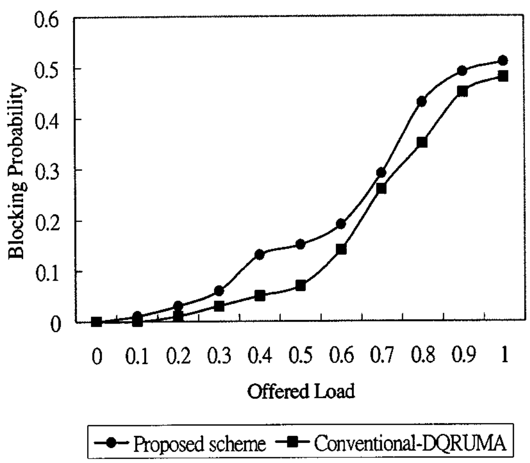


Fig. 4. Average access delay of voice traffic.

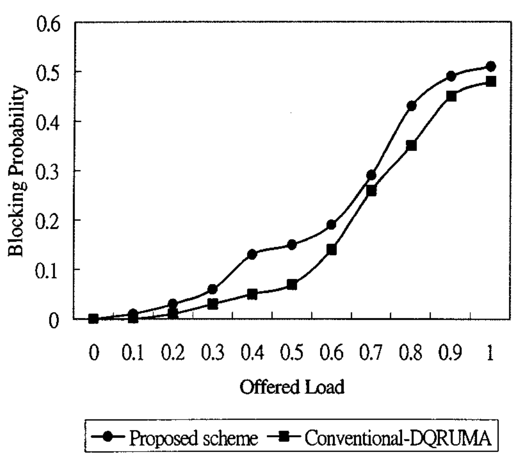


Fig. 5. Average access delay of video traffic.

proposed scheme remains low when the offered load is high, but the conventional DQRUMA shows a sharp rise as the load increases.

Fig. 6 shows the average access delay of data traffic under multimedia traffic condition. As expected, the average access delay of data traffic in the proposed scheme is worse than the conventional DQRUMA since it is of low priority.

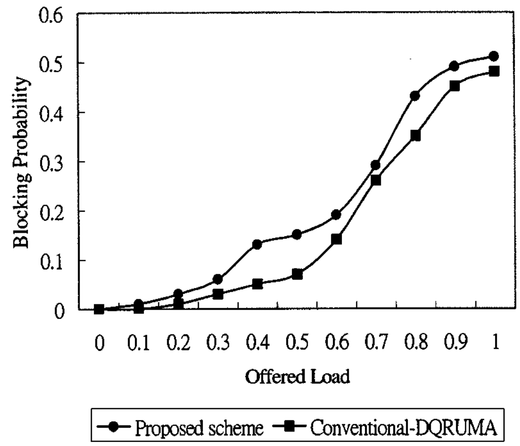


Fig. 6. Average access delay of data traffic.

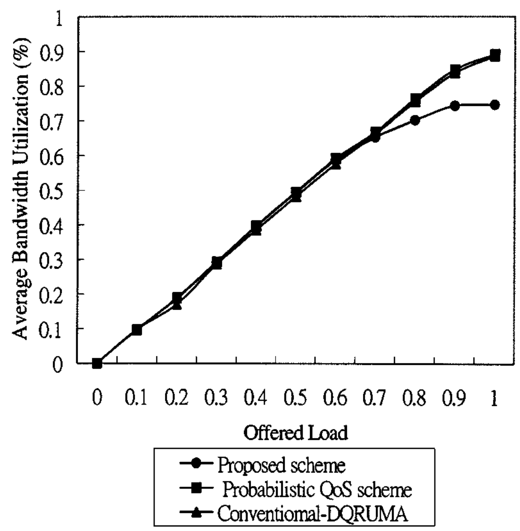


Fig. 7. Average bandwidth utilization.

Fig. 7 presents the average bandwidth utilization as a function of the offered load. Average bandwidth utilization is the percentage of the bandwidth actually being used in the total bandwidth. As illustrated in Fig. 7, the average bandwidth utilization is lower for the proposed scheme in a highly loaded system, because to maintain the QoS, it must be more conservative in admitting new connections. It reveals that there is a clear tradeoff between deterministic (hard) QoS supporting and bandwidth utilization. For comparison, we have conducted a simulation of probabilistic (soft) QoS scheme. In the probabilistic scheme, the requirements in Theorems 1 and 2 are not checked before admitting a connection. It can be seen that the bandwidth utilization is increased in the high load area. We conclude that our proposed scheme reduces the handoff dropping probability without sacrificing the bandwidth utilization too much.

IV. Conclusions

The design of priority-sensitive network protocols continues to be an important problem, and broadband wireless links constitute a subclass where prioritization is key to optimizing overall performance. In this paper, we proposed a pragmatic nonpreemptive priority based access control scheme built on well-known protocols, and offered easily implemented and yet flexible criteria for traffic prioritization in a mobile environment. We also demonstrate the performance in a quantitative way.

Wireless network is a rapidly emerging field of activity in computer network because it supports mobility. That is, it provides user connectivity without being tethered off by wired networks. In the meantime, another new technology is poised to impact business computing in an equally dramatic way. Networked multimedia computer applications will significantly affect users and network managers, and have a tremendous impact on computing and network infrastructures. Predictably, a global, ubiquitous wireless network will allow its users to communicate with anyone, anywhere, and at any time in the future. As wireless networks become common, new applications will evolve to take full advantage of this technology to affect the way we work and play.

However, the success of wireless networks depends on the availability of corresponding backbone wired infrastructure and the evolution of the software applications. The new generation wireless technologies should support universal wide-band access to a variety of services such as cordless telephony, Internet access, multimedia conference, remote audio, and flexible positioning of audio system. This means that various QoS requirements are needed in the future. Thus, multilevel priorities, bandwidth allocation, connection admission control, and traffic policing all need to be considered together to satisfy various QoS flows in future networks.

Acknowledgment

The authors would like to thank the anonymous referees for their helpful comments that greatly improved the quality of this paper.

References

[1] I. Tokizawa, H. Ueda, and K. Kiskuchi, "ATM transport system architecture and field trial," in *Conf. Rec. Globecom'93*, Nov. 29–Dec. 2, 1993, pp. 1454–1460.

[2] W. Fischer, E. Wallmeier, T. Worster, S. P. Davis, and A. Hayter, "Data communications using ATM: Architectures, protocols, and resources management," *IEEE Commun. Mag.*, vol. 32, pp. 24–33, Aug. 1994.

[3] G. Woodruff and R. Kositpaiboon, "Multimedia traffic management principles for guaranteed ATM network performance," *IEEE J. Select. Areas Commun.*, vol. 8, pp. 437–445, 1990.

[4] D. Raychaudhuri *et al.*, "WATMnet: A prototype wireless ATM system for multimedia personal communication," *IEEE J. Select. Areas Commun.*, vol. 15, pp. 83–95, Jan. 1997.

[5] G. P. Fettweis *et al.*, "A closed solution for an integrated broadband mobile system (IBMS)," in *Conf. Rec. ICUPC'96*, Cambridge, MA, Sept. 1996, pp. 707–711.

[6] J. Mikkonen, J. Aldis, and G. Awater, "The magic WAND—Function overview," *IEEE J. Select. Areas Commun.*, vol. 16, pp. 953–972, Aug. 1998.

[7] D. J. Skellern, L. H. C. Lee, T. McDermott, N. E. Waste, J. Dalton, J. Graham, T. F. Wong, A. F. Myles, T. M. P. Percival, and P. J. Ryan, "A high speed wireless LAN," *IEEE Micro*, vol. 17, pp. 40–47, Jan. 1997.

[8] P. Agrawal *et al.*, "SWAN: A mobile multimedia wireless network," *IEEE Personal Commun.*, pp. 18–33, Apr. 1996.

[9] Y. Du, C. Herrmann, and K. P. May, "Wireless ATM LAN with and without infrastructure," *IEEE Commun. Mag.*, pp. 90–95, Apr. 1998.

[10] Institut fuer Mobil- und Satellitenfunktechnik GmbH (IMST). MEDIAN: Wireless broadband CPN/LAN for Professional and Residential Multimedia Applications. [Online]http://www.imst.de/median/median.html

[11] Defense Advanced Research Projects Agency (DARPA). RDRN: Rapidly deployable radio networks. [Online]http://www.ittc.ukans.edu/Projects/RDRN/

[12] Univ. Kansas, Rome Lab., "Wireless ATM adaptive voice/data networks,", http://www.ittc.ukans.edu/Projects/Wireless_ATM/.

[13] AT&T Laboratories, "Radio ATM,", http://www.cam-orl.co.uk/radio/.

[14] K. Y. Eng, M. J. Karol, M. Veeraraghavan, E. Ayanoglu, C. B. Woodworth, P. Pancha, and R. A. Valenzuela, "BAHAMA: A broadband *ad-hoc* wireless ATM local-area network," in *Conf. Rec. ICC'95*, pp. 1216–1223.

[15] L. M. Correia and R. Prasad, "An overview of wireless broadband communications," *IEEE Commun. Mag.*, vol. 35, pp. 28–38, Jan. 1997.

[16] D. Raychaudhuri and N. D. Wilson, "ATM-based transport architecture for multiservices wireless personal communication network," *IEEE J. Select. Areas Commun.*, vol. 12, pp. 1401–1413, Oct. 1994.

[17] Y. Du, S. N. Hulyalkar, and D. Evans *et al.*, "System architecture of a home wireless ATM network," in *Conf. Rec. ICUPC'96*, Cambridge, MA, Sept. 1996, pp. 471–481.

[18] E. K. Wesel, *Wireless Multimedia Communications*. Reading, MA: Addison-Wesley, 1997, pp. 5–7.

[19] P. Agrawal, "Mobile and wireless computing," in *Tutorial 4th Mobile Computing Workshop*, Taiwan, Mar. 1998, pp. 17–25.

[20] Q. Kubbar and H. T. Mouftah, "Multiple access control protocols for wireless ATM: Problem definition and design objectives," *IEEE Commun. Mag.*, pp. 93–99, Nov. 1997.

[21] N. D. Wilson *et al.*, "Packet CDMA versus dynamic TDMA for multiple access in an integrated voice/data PCN," *IEEE J. Select. Area Commun.*, vol. 11, pp. 870–883, Aug. 1993.

[22] A. S. Tanenbaum, *Computer Networks*, 3rd ed. Englewood Cliffs, NJ: Prentice-Hall, 1996, pp. 247–250.

[23] Wireless Medium Access Control and Physical Layer WG, IEEE Draft Standard P802.11, "Wireless LAN," IEEE Stds. Dept, D3, Jan. 1996.

[24] Z. Zhang and A. S. Acampora, "Performance of a modified polling strategy for broadband wireless LAN's in a harsh fading environment," *Telecommun. Syst.*, vol. 1, no. 3, pp. 279–294, Feb. 1993.

[25] D. J. Goodman, R. A. Valenzuela, K. T. Gayliard, and B. Ramamurthi, "Packet reservation multiple access for local wireless communications," *IEEE Trans. Commun.*, vol. 37, pp. 885–890, Aug. 1989.

[26] M. J. Karol, Z. Liu, and K. Y. Eng, "Distributed-queueing request update multiple access (DQRUMA) for wireless packet (ATM) network," in *Conf. Rec. ICC'95*, pp. 1224–1231.

[27] S. Lee, Y. J. Song, D. H. Cho, Y. B. Dhong, and J. W. Yang, "Wireless ATM MAC layer protocol using WDWEDF and two-phase scheduling algorithm," *IEICE Trans. Commun.*, vol. E81-B, no. 12, pp. 2432–2443, Dec. 1998.

[28] J. Sanchez, R. Martinez, and M. W. Marcellin, "A survey of MAC protocols proposed for wireless ATM," *IEEE Network*, pp. 52–62, Nov./Dec. 1997.

[29] M. Yamamoto, S. Machida, and H. Ikeda, "Access control scheme for multimedia ATM wireless local area networks," *IEICE Trans. Commun.*, vol. E81-B, no. 11, pp. 2048–2055, Nov. 1998.

[30] B. Moon, C. Oh, A. Ahmad, and K. Kim, "A study of bandwidth allocation strategies in multimedia wireless networks," in *Conf. Rec. APCC'97*, pp. 509–513.

[31] C. Oliveira, J. Bae, and T. Suda, "An adaptive bandwidth reservation scheme for high-speed multimedia wireless networks," *IEEE J. Select. Area Commun.*, vol. 16, pp. 858–873, Aug. 1998.

[32] C. S. Chang, K. C. Chen, M. Y. You, and J. F. Chang, "Guaranteed quality-of-service wireless access to ATM networks," *IEEE J. Select. Area Commun.*, vol. 15, pp. 106–118, Jan. 1997.

[33] R. L. Cruz, "A calculus for network delay, Part I: Network elements in isolation," *IEEE Trans. Inform. Theory*, vol. 37, pp. 114–131, 1991.

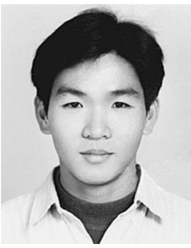
[34] ——, "A calculus for network delay, Part II: Network analysis," *IEEE Trans. Inform. Theory*, vol. 37, pp. 132–141, 1991.

[35] R. P. Grimaldi, *Discrete and Combinatorial Mathematics*, 3rd ed. Reading, MA: Addison Wesley, 1994, pp. 183–186.

[36] CACI Products Company, Simscript II.5, Calif. 92037. (1997, Sept.). [Online]http://www.caciasl.com/

[37] A. S. Tanenbaum, *Computer Networks*, 3rd ed. Englewood Cliffs, NJ: Prentice-Hall, 1996, pp. 263–264.

[38] B. Maglaris, S. Anastassiou, P. Sen, G. Karlsson, and J. D. Robbins, "Performance models of statistical multiplexing in packet video communications," *IEEE Trans. Commun.*, vol. 36, pp. 834–844, July 1998.



Dr-Jiunn Deng received the M.S. degree in information management from National Taiwan University of Science and Technology, Taipei, Taiwan, in 1999.

He is currently pursuing the Ph.D. degree in the Department of Electrical Engineering, National Taiwan University, Taipei. His main interests include quality-of-service, mobile multicasting, and wireless medium access control protocols.

Ruay-Shiung Chang (M'98) received the B.S.E.E. degree from National Taiwan University in 1980 and the Ph.D. degree in computer science from National Tsing Hua University in 1988.

From 1988 to 1992, he was a Senior Researcher at the Chung Shan Institute of Science and Technology. From 1992 to 1998, he was with the Department of Information Management, National Taiwan University of Science and Technology. In August 1998 he joined the Department of Computer Science and Information Engineering, National Dong Hwa University, where he is currently the Chairman. His research interests include internet, high-speed networks, and wireless networks.

Dr. Chang is a member of the ACM and ROC Institute of Information and Computing Machinery.