International Journal of Wireless Information Networks, Vol. 4, No. 3, 1997

A Bandwidth Reservation Multiple Access Protocol for Wireless ATM Local Networks

Z. Zhang,[1] I. Habib,[1] and T. Saadawi[1]

A bandwidth reservation multiple access scheme (BRMA) is proposed to resolve contention and assign bandwidth among multiple users trying to gain access to a common channel such as in mobile users contending for resources in an ATM-based cellular network or a wireless local area network (LAN) with short propagation delays. The protocol is best suited to support variable-bit-rate (VBR) traffic that exhibits high temporal fluctuations. Each mobile user is connected end-to-end to another user over virtual channels via the base station that is connected to the wired ATM B-ISDN network. The channel capacity is modeled as a time frame with a fixed duration. Each frame starts with minislots, to resolve contention and reserve bandwidth, followed by data-transmission slots. Every contending user places a request for data slots in one of the minislots. If the request is granted by the base station through a downlink broadcast channel, the user then starts transmission in the assigned slot(s). The number of assigned slots varies according to the required quality of service (QoS), such as delay and packet loss probability. A speech activity detector is utilized in order to indicate the talkspurts to avoid wasting bandwidth. Due to its asynchronous nature, BRMA is rather insensitive to the burstiness of the traffic. Since the assignment of the minislots is deterministic, the request channels are contention-free and the data channels are collision-free. Hence, in spite of the overhead (minislots) in each frame, BRMA provides higher throughput than Packet Reservation Multiple Access (PRMA) for the same QoS, especially for high-speed systems. A better delay performance is also achieved for data traffic compared to Slotted Aloha reservation-type protocol PRMA. In addition, BRMA performs better in terms of bandwidth efficiency than the conventional TDMA or the Dynamic TDMA, where speech activity detectors are very difficult to implement.

KEY WORDS: WATM LANs; wireless access; dynamic channel assignment; TDMA.

1. INTRODUCTION

Wireless communications make it possible for users to access the advanced information services that include multimedia applications from virtually anywhere and at any time. The connection-oriented Asynchronous Transfer Mode (ATM), a well-accepted concept in wired communications, has also established its position in wireless networks. The combination of these, i.e., the Wireless ATM (WATM), has been one of the hottest topics in communications in the 1990s.

In a wireless system, the performance depends largely on the media access control (MAC) protocol. A MAC protocol of WATM should support ATM traffic classes such as available bit rate (ABR), variable bit rate (VBR), constant bit rate (CBR), and unspecified bit rate (UBR) while maintaining a high wireless channel utilization. This field has been very active and many protocols have been proposed and studied [1–7]. For example, D. Petras proposed a Dynamic Slot Assignment (DSA) protocol for Mobile Broadband Systems (MBS), which

[1]Department of Electrical Engineering, City University of New York, City College, New York, New York; e-mail: {zzhang, eeiwh,

like B-ISDN. In DSA the distributed queueing information from each mobile station is included in the uplink data packets to help the base station to determine if additional slots should be assigned to the mobile station [1]. This protocol was evaluated in Ref. 2 and enhanced to DSA++ in Ref. 3. In recent years, there have already been some standardization activities to determine the MAC of wireless ATM LANs, such as Europe's Mobile Broadband Systems (MBS) [8] and ETSI HIPERLAN MAC [9].

TDMA technology has been well understood and is now widely used in many wireless networks, including the current European global system of mobile telecommunications (GSM) [10–14]. Originally designed for the circuit switching networks, most TDMA versions, including the multiservice dynamic reservation (MDR) TDMA scheme studied in Ref. 15, do not utilize the wireless bandwidth as efficiently as the wideband CDMA, especially under bursty traffic conditions such as voice communications. One reason is that in these protocols, the exploitation of the speech activity factor for voice communication is very difficult [15].

To efficiently utilize the bandwidth, PRMA, a Slotted Aloha reservation type of TDMA, was proposed by Goodman *et al.* [16] and has been widely studied by many others [17–21]. Unlike Slotted Aloha, contentions for a time slot in PRMA occur only at the beginning of each talkspurt of the conversation and unlike traditional TDMA, PRMA allows a voice user to reserve a slot only during each talkspurt rather than during the whole conversation. For voice communications where the average talkspurt is on the order of 1 s, the utilization of the channel can reach 0.68 [17], higher than that of traditional TDMA. Mitrou *et al.* [22] proposed and studied an improved version of PRMA in which a minimum portion of the available channel capacity is dedicated to the reservation channel. Slotted Aloha contentions occur in some slots that are further divided into minislots. As a result, the throughput performance under high load conditions is improved. We observe, however, that as the average length of talkspurt of the conversation decreases, the performance of this class of Slotted Aloha/TDMA reservation protocols degrades due to more frequent collisions at the beginning of each talkspurt. In the extreme case, when the user transmits one packet each spurt, the above protocols perform just like Slotted Aloha. Therefore PRMA described in Refs. 16 and 17, though good for slow time-varying and long-burst traffic such as voice traffic, is not suitable for highly time-varying traffic such

Motivated by the highly bursty nature of the multimedia traffic encountered in ATM networks and the requirements that the network must deliver different quality of services (QoS) for different types of traffic, we propose a contention-free MAC protocol, referred to as Bandwidth Reservation Multiple Access Protocol (BRMA), that meets the needs to allocate and control the wireless channel bandwidth. It is designed to achieve two main goals:

1. To ensure contention-free transmissions by avoiding collisions and retransmissions that are not suitable for high-speed real-time traffic such as video or voice.
2. To reserve bandwidth for each type of traffic in order to maintain QoS requirements. For example, video or voice traffic requires stringent delay requirements and would have priority over data traffic. Hence, the protocol must be capable of allocating different amounts of bandwidth to different types of traffic.

To achieve the first goal, we implemented a frame that is divided into time slots. Attached to each frame is an overhead made of minislots. Users contending for an amount of bandwidth send their requests to the base station via those minislots. If the bandwidth is available, then the request is honored and an average number of slots is assigned to the user every frame such that the QoS is met.

To achieve the second goal, we incorporate a bandwidth allocation algorithm into the base station. Simply stated, the algorithm first reads the requested QoS from the request packet at the time of the call establishment. It then calculates the required bandwidth and, hence, the number of slots per frame. If those slots are available, the call is then accepted; otherwise it is rejected. The advantages of incorporating this admission control policy are several: First, it harmonizes the protocol functionality with that of the ATM. Second, it avoids long-term congestion episodes due to overloading of the network's resources by avoiding admitting calls for which no sufficient resources are available. Hence, excessive packet loss due to prolonged delays or buffer overloading are minimized.

The rest of this paper is arranged as follows. Section 2 describes the system overview and BRMA protocol details. Performance analysis is given in Section 3, and simulation and numerical results are provided in Section 4. Finally, the conclusions are given in

2. SYSTEM AND PROTOCOL DESCRIPTION

In wireless TDMA schemes, time is usually divided into frames and slots. A guard time is needed between two consecutive slots or frames to accommodate the effect of the propagation delays [14, 15]. Partially for this reason, in a high-speed wireless LAN where TDMA is used, the propagation delay should be very small if high utilization is to be achieved. In other words, most high-speed wireless LANs are usually designed to cover small areas called microcells or even picocells. In fact, in some applications the confinement of the signal to a small area is a desirable feature for wireless LANs [23]. In LANs that cover such small areas, practical propagation delay is on the order of a few microseconds and has no effect on the protocol performance. BRMA is a wireless protocol for such purposes.

Figure 1 shows a possible architecture of the WATM LAN. Each end-to-end connection between two mobile stations is assigned a virtual channel. Multiple virtual channels may exist between two end stations. The base station acts as a central controller in the WATM LAN and a bridge to the B-ISDN network via a base station controller (BSC). The users of the WATM terminals request the same functionality and QoS as users of wired terminals. The QoS requirements include average and maximum cell rates, average and maximum cell delays, cell loss probability, handover and call teardown rates, etc. In this paper, for simplicity, we limit our discussion of QoS to only cell loss probability and cell delay where we evaluate the performance of BRMA within one microcell area.

The call setup is established in a separate out-of-band signaling channel. A mobile station requesting connection sends its traffic characteristic parameters and QoS requirements to the base station. The admission control scheme sitting at the base station will evaluate whether there is enough bandwidth available to support the requested connection with QoS. A virtual channel will be assigned to the connection with a minimum and an average bandwidth in slots/frame if such bandwidth is available. Otherwise the call is rejected. Examples of such admission schemes are provided in Refs. 24 and 25. Obviously, upper layer protocols (e.g., IP) must include the provisions for mobility support, since mobile stations may change their locations dynamically. Issues such as handover control and routings, among others, have been addressed by proposals such as Dynamic Host Configuration Protocol (DHCP)

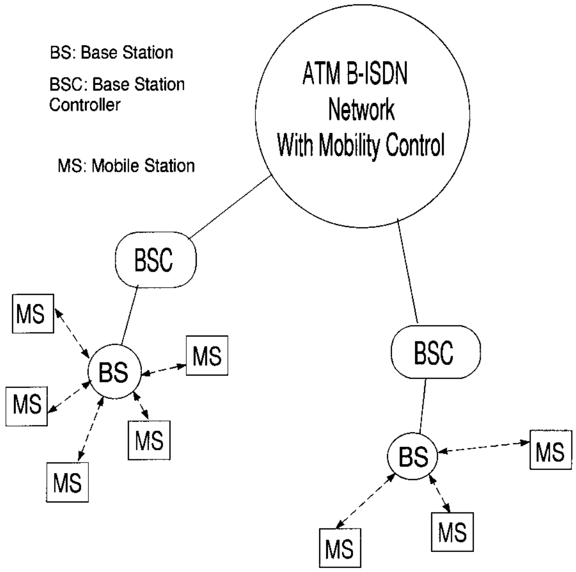

Fig. 1. Typical wired/wireless network connectivity.

Each mobile station utilizes its share of the uplink channel according to the frame-level dynamic assignment by the central base station. Users share the channel bandwidth via a time frame that is further divided into K smaller data slots. The duration of each slot equals one packet transmission time. Attached to each frame are N minislots that are deterministically assigned to the N mobile stations that have already established their connections to request data slots in each frame. This number N is dynamically updated when a virtual channel is added or deleted due to new call admissions or call completions. At the beginning of each frame, each station sends a minipacket in its assigned minislot. The minipacket includes information such as the call ID and the number of data slots this call requests in this frame. At the end of the minislots the base station knows the total number of required slots for all stations in this frame, and immediately broadcasts the data slot assignments of the current frame. Each mobile station then transmits its data packets in its assigned data slot(s); see Fig. 2.

The data slot assignments depend not only on the preassigned bandwidth, but also on the traffic conditions. If the total requested data slots exceeds the available data slots in a frame, each mobile station will get at least its minimum number of slots in this frame. Packets that are not transmitted in this frame will have to wait in the user's own buffer and be transmitted in the following frame(s). Hence, the user will request slots for both new

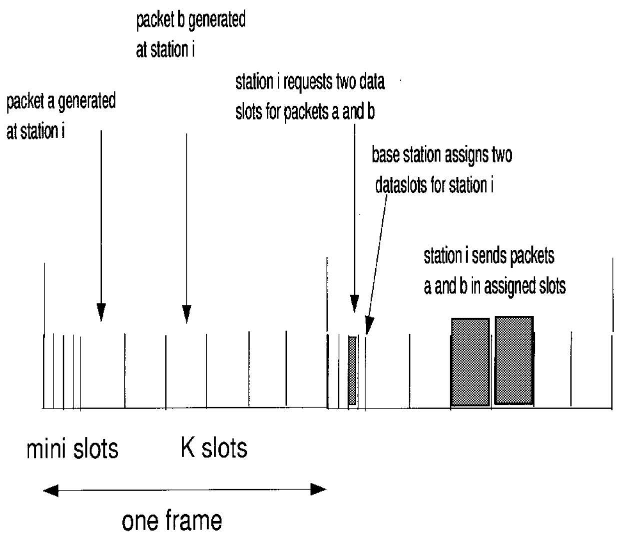

Fig. 2. Frame structure and an example of scheduling of packets in BRMA protocol.

is temporarily overloaded, this scenario may result in packet loss if the packets are delay-sensitive. However, with proper admission control and bandwidth allocation, packet loss will be bounded by a maximum value.

The above assignment procedure is repeated every frame. Hence the slot assignment is dynamic at the frame level. When voice communication is the main traffic, this protocol can accommodate the activity factor detector naturally. This is a big advantage over conventional versions of TDMA, which have to assign bandwidth to the silent portion of each conversation.

Under the control of BRMA, no two or more mobile stations in the same microcell will attempt to access the wireless channel at the same time. No packet collisions are possible; therefore the typical hidden terminal problem does not exist here.

Most functions of the ATM layer of the BISDN ATM reference model [27] are performed at the base station (e.g., (de)multiplexing, cell header addition and extraction, flow control, etc.) Therefore the base station contains an ATM header table for each admitted call. This table is updated dynamically. Packets directed from the wireless LAN through the base station to the ATM backbone network are therefore ATM cells.

BRMA is different from PRMA mainly in following aspects: In BRMA, the reservation channel is contention-free because each minislot is dedicated to each user. Also, the data channel is collision-free because tion channel by the base station. Note that in PRMA a channel is subject to possible packet collisions unless a reservation is made already. In PRMA, multiplexing for voice packets occur at talkspurt level. The duration of the frame is equal to the interarrival time of the voice packets, and at most one data slot is assigned to each voice user every frame. For BRMA, multiplexing occurs at the frame level and the frame length does not have to be the interarrival time of voice packets. One user may have multiple data slots in a frame. Of course, BRMA cannot be used to replace PRMA—they suit different systems. For example, PRMA will generate smaller packet delays for voice traffic when the system load is light.

BRMA is also different from a collision-free protocol called the bit-map method discussed in Ref. 28. Although the bit-map method also uses similar minislots and access frames, it is basically a protocol of distributed control. Therefore, it does not have the potential to perform the functions of priority, guaranteeing different QoS and bandwidth allocation. Conversely, BRMA can perform all these functions because it is basically a protocol of central control. In Bit-Map, a user declares its request in minislot, then goes ahead and uses the data slots. In BRMA, a user is allocated data slots according to the base station's assignment.

BRMA is different from the dynamic TDMA in several aspects and functionality. First, in the dynamic TDMA [15], voice activity detectors are difficult to implement; it assigns slots to voice users based on a "circuit mode." Second, in dynamic TDMA each voice user gets at most one data slot in each frame and frame length must therefore be the interarrival time of voice packets (just like in PRMA). However, BRMA, as previously indicated, assigns slots in a "statistical multiplexing" mode in the sense that every user gets a variable number of slots every frame.

3. PROTOCOL ANALYSIS

Exact modeling of the multimedia traffic is difficult since it involves the characterization of complex nonrenewal processes such as the one generated by variable-bit-rate (VBR) compressed video sources. In our study, we first consider voice traffic sources, and then the superposition of voice and data traffic sources. To compare with PRMA, we use the same non-ATM voice packet and similar channel parameters as those used in the PRMA performance study [17]. We then show that, for

We assume that a voice packet has v bits of payload, and that each packet has a header of h bits. Each minislot contains h_m bits. Each voice source generates voice packets at the peak rate of R_p kbps during the talkspurts and no packets during silent periods. Both talkspurts and silent period are exponentially distributed with means of T_A ms and T_S ms, respectively. During talkspurts, voice packets arrive every t_p ms and there are N stations connected through the base station. If the voice packets are not transmitted within D ms, they will be dropped. A satisfactory QoS here indicates a packet loss rate of under 1%. Intuitively, very long frames will lead to excessive packet loss due to prolonged delays, thus limiting the number of stations that can be supported. On the other hand, very short frames may cause the total throughput to degrade due to excessive frame overhead. The optimal number of data slots in a frame is the one that can maximize the throughput and guarantee QoS for a certain number of users.

In the following analysis, we first consider N homogeneous voice sources connected to the base station over the wireless link in a single microcell. Each source generates voice traffic following an ON/OFF model. We analyze the queue and obtain the average packet loss rate and average packet waiting time for different values of K. Second, we consider $N-1$ homogeneous voice users and one data user. We analyze the queue and obtain similar measurements.

3.1. Queueing Analysis for Voice Traffic Only

Designating d as the data slot length, L_f as the length of a frame in ms, we have

$$
\begin{aligned}
L_f &= K \times d + \frac{N \times h_m}{C} \\
d &= (h + v)/C
\end{aligned}
\tag{1}
$$

In order to simplify the analysis, we express the maximum delay limit in terms of a finite-size buffer. Considering that each buffer space is equivalent to a waiting time of one data slot d, and that each frame contains a minislot header of $(L_f - K \times d)$ ms and that there are D/L_f frames in time D, we obtain the equivalent buffer size limit B:

$$
B = \left\lfloor \left(D - \frac{D}{L_f} \times (L_f - K \times d) \right) \Big/ d \right\rfloor \tag{2}
$$

We observe the arrival process from N homogeneous voice sources at the beginning of each frame. This process can be characterized by a discrete-time Markov chain of dimension $N + 1$, where its state is defined by the number of conversations is talkspurt x. In many published papers [e.g., 29, 30] in which similar analyses were performed, the frame length was chosen to be the packet interarrival time t_p. In our study, this case is not applicable since the frame length is different from the packet interarrival time.

We use $P(x_1, x_2)$ to represent the probability that at the beginning of the $(n + 1)$th frame there will be x_1 sources in a talkspurt given that at the beginning of the nth frame there are x_2 sources in a talkspurt in the system.

Let $P_{A/S}$ be the transitional probability that a source will switch from talkspurt phase to silent phase during a single frame, and $P_{S/A}$ be the transitional probability that a source switches from silent phase to talkspurt phase during the same time period. Noticing that both talkspurt and silent periods are exponentially distributed with means T_A and T_S, respectively, we have

$$
\begin{aligned}
P_{A/S} &= 1 - e^{-L_f/T_A} \\
P_{S/A} &= 1 - e^{-L_f/T_S}
\end{aligned}
\tag{3}
$$

Now we try to obtain an expression for $P(x_1, x_2)$. Two cases arise:

1. For $x_1 \geq x_2$, for the number of sources in a talkspurt in the system to change from x_2 to x_1, there must be exactly k sources in a talkspurt changing to silent and exactly $x_1 - x_2 + k$ silent sources changing to a talkspurt in one frame, where $k = 0, 1, 2, \ldots, \min(x_2, N - x_1)$. Hence we have

$$
\begin{aligned}
P(x_1, x_2) = \sum_{k=0}^{\min(x_2, N - x_1)} & \binom{x_2}{k} P_{A/S}^{k} (1 - P_{A/S})^{x_2 - k} \\
& \times \binom{N - x_2}{x_1 - x_2 + k} \\
& \times P_{S/A}^{x_1 - x_2 + k} (1 - P_{S/A})^{N - x_1 - k}
\end{aligned}
\tag{4}
$$

DOCKET ALARM

Explore Litigation Insights

Docket Alarm provides insights to develop a more informed litigation strategy and the peace of mind of knowing you're on top of things.

Real-Time Litigation Alerts

Keep your litigation team up-to-date with **real-time alerts** and advanced team management tools built for the enterprise, all while greatly reducing PACER spend.

Our comprehensive service means we can handle Federal, State, and Administrative courts across the country.

Advanced Docket Research

With over 230 million records, Docket Alarm's cloud-native docket research platform finds what other services can't. Coverage includes Federal, State, plus PTAB, TTAB, ITC and NLRB decisions, all in one place.

Identify arguments that have been successful in the past with full text, pinpoint searching. Link to case law cited within any court document via Fastcase.

Analytics At Your Fingertips

Learn what happened the last time a particular judge, opposing counsel or company faced cases similar to yours.

Advanced out-of-the-box PTAB and TTAB analytics are always at your fingertips.

API

Docket Alarm offers a powerful API (application programming interface) to developers that want to integrate case filings into their apps.

LAW FIRMS

Build custom dashboards for your attorneys and clients with live data direct from the court.

Automate many repetitive legal tasks like conflict checks, document management, and marketing.

FINANCIAL INSTITUTIONS

Litigation and bankruptcy checks for companies and debtors.

E-DISCOVERY AND LEGAL VENDORS

Sync your system to PACER to automate legal marketing.

WHAT WILL YOU BUILD? | sales@docketalarm.com | 1-866-77-FASTCASE

fastcase® Smarter legal research.