# A Dynamic Reservation Protocol for LEO Mobile Satellite Systems

Henry C. B. Chan, *Member, IEEE*, Jie Zhang, and Hui Chen

***Abstract*—In this paper, we propose a dynamic reservation protocol for the low earth orbit (LEO) mobile satellite system. Based on an analytical model, the dynamic reservation protocol enhances system performance by dynamically varying both the access probability and reservation bandwidth. To implement the protocol, a novel contention-pattern-analysis method is proposed to estimate the number of contending terminals at the start of a frame. The reservation protocol is employed to integrate connection-oriented and connectionless traffic over a satellite channel. Simulation and theoretical results are presented to analyze the performance of the protocol and illustrate its benefits.**

***Index Terms*—Low earth orbit (LEO) satellites, multiple-access protocols, reservation protocols.**

## I. Introduction

In RECENT years, many satellite projects have been implemented for various purposes [1]–[5]. In particular, low earth orbit (LEO) mobile satellite systems can provide worldwide personal communication services. With the advent of Internet-based multimedia applications, the next generation of satellite systems will be required to support both connection-oriented (CO) and connectionless (CL) traffic over a broadband channel, particularly using the Internet protocol (IP). Currently, multiprotocol label switching (MPLS) [6] provides an effective mechanism for forwarding IP datagrams over the well-established asynchronous transfer mode (ATM) framework. By using MPLS, many ATM-based satellite systems [2], [3] can be extended to integrate CO and CL services. Motivated by these developments, we consider an MPLS/ATM-based LEO mobile satellite system. The aim of this paper is to develop an effective and efficient access/reservation protocol for this system.

Packet reservation multiple access (PRMA) [7] was originally developed to support integrated CO and CL services for a cellular system. As it can provide bandwidth-on-demand services over the traditional time-division multiplexing framework, there has been considerable interest in extending PRMA for LEO mobile satellite systems [8]–[11]. Inspired by these reservation protocols, this paper proposes a dynamic reservation protocol with some new contributions. Essentially, reservation minislots are employed to improve the channel efficiency and the number of minislots and the access probability are dynamically varied based on the estimated number of contending terminals. To develop the dynamic reservation protocol, we need to investigate two fundamental questions. First, before obtaining the contention result, should a contending terminal that has accessed a reservation slot/minislot (tried terminal) be disabled? Note that in a satellite system, the contention result is not immediately known because of propagation delay. If the tried terminals do not contend (e.g., like PRMA [7]), many minislots may become idle and, hence, are wasted. However, if they continue to contend (e.g., like packet reservation multiple-access-hindering states (PRMA-HS) [8]), the contending terminals that have not tried to access a reservation minislot may be affected. Our new contribution is to study this issue in detail with an analytical model. Second, most packet reservation protocols use a constant access probability. However, this may not work well for the emerging broadband satellite systems because the number of contending terminals is larger and the data rate changes more quickly. To enhance the performance of the system, it would be desirable to vary the access probability dynamically. Unfortunately, the traditional Bayesian algorithm and various splitting algorithms [12] cannot be applied because they generally require immediate feedback. A frame-based Bayesian algorithm has been proposed in [13] for a delayed feedback environment. However, this method can only be employed for Poisson-based traffic. In this paper, we present an analytical model to study the above issues in detail. The model can cover both the PRMA and PRMA-HS protocols. Based on the analytical model, two methods for varying the access probabilities are proposed, namely: frame-based or minislot-based. To implement the protocol, we propose a novel contention-pattern-analysis method for predicting the number of contending terminals. Basically, this method predicts the number of contending terminals by examining the contention pattern in the previous frame. Combining the above features, this paper presents a dynamic reservation protocol for the emerging MPLS/ATM-based LEO satellite system and analyzes its performance.

The remaining sections of the paper are organized as follows. Section II provides the system model. Section III presents the analytical framework and the dynamic reservation protocol. Section IV presents the performance analysis and discusses the findings. Section V concludes the paper.

## II. System Model

Fig. 1 shows the network architecture of the LEO mobile satellite system with three segments: space, ground and user [1], [4]. Users can communicate with the LEO satellites through their fixed/mobile terminals or using appropriate devices. Like [8]–[11], the "earth fixed cell" approach is employed. The satellite system can be linked with the Internet and other

Manuscript received December 15, 2002; revised July 1, 2003 and November 10, 2003. This work was supported in part by The Hong Kong Polytechnic University under Research Accounts A-PB23 and G-T041.

The authors are with the Department of Computing, The Hong Kong Polytechnic University, Kowloon, Hong Kong (e-mail: cshchan@comp.polyu.edu.hk; csjzhang@comp.polyu.edu.hk; cshchen@comp.polyu.edu.hk).

Digital Object Identifier 10.1109/JSAC.2004.823439

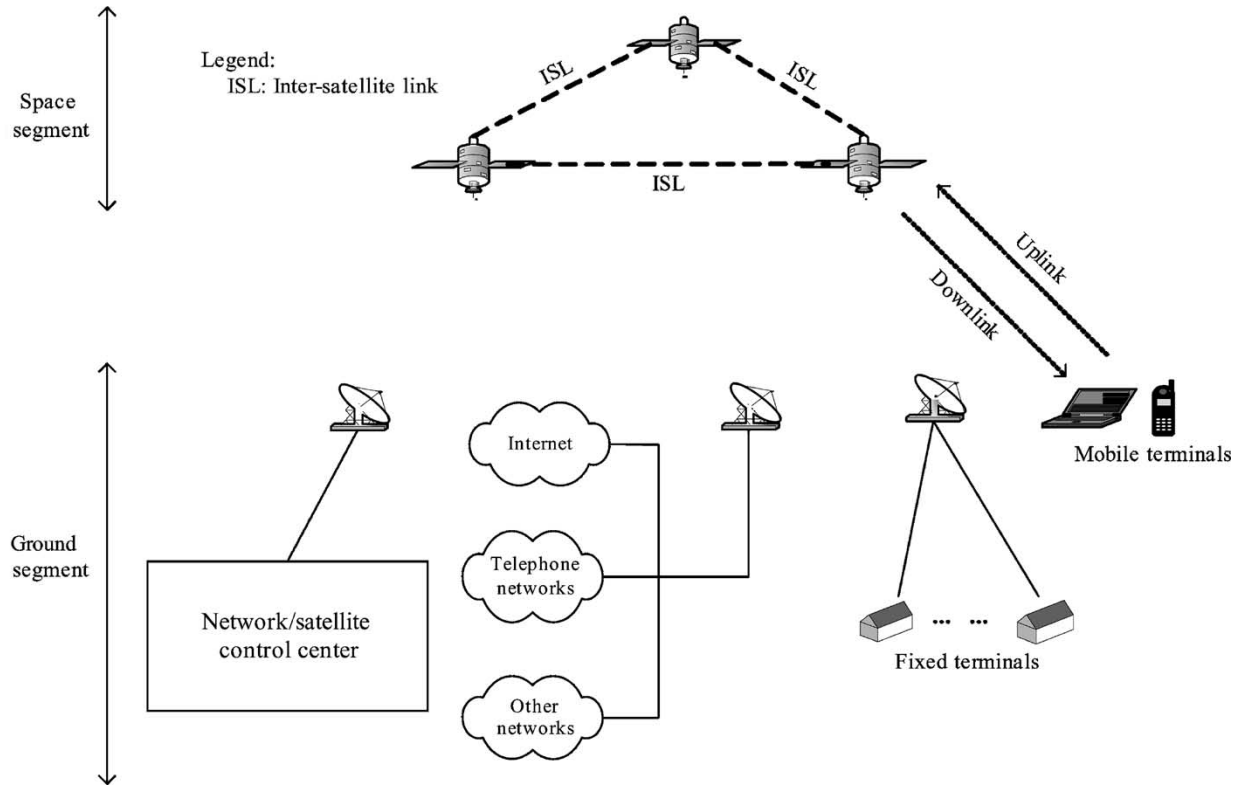


Fig. 1. General network architecture of the LEO mobile satellite system (inspired by diagrams in [1] and [4]).

networks by using suitable gateways. The network/satellite control center provides various management functions (see [1] for details). At the space segment, the LEO satellites, which have an on-board processing capability, are networked by inter-satellite links (ISLs). We consider that the satellites can support MPLS by means of ATM switching. As a further enhancement, the LEO satellites can form a hierarchical network with the medium earth orbiting (MEO) and geosynchronous earth orbit (GEO) satellites based on [14]. This paper addresses the access/reservation protocol for an uplink channel (i.e., from the user terminals/gateways to the satellite). As shown in Fig. 2, the channel is framed, with three subframes in each frame: reservation minislots for conveying reservation requests, slots for CO traffic and slots for CL traffic. There are $F$ slots per frame and the duration of the frame is $f = 24$ ms. As will be explained later, the number of slots/minislots in each subframe, as well as the access probabilities, are dynamically varied. Generally, slots are assigned in the following order subject to availability: slots for CO traffic based on the reservations, reservation minislots for CO traffic based on the estimated number of CO terminals (see Section III), slots for CL traffic based on the registered slot requirements, and reservation minislots for CL traffic from the residual capacity. As an example, we assume that each slot has a 48-B payload for carrying a user packet and a 6-B header for link management purposes. The header generally includes a modified version of the standard ATM header and additional control fields (e.g., for indicating the traffic type, mapping the payload to a cached IP address if necessary and supporting the piggybacked slot requirements to be discussed later). The size of the payload is chosen to facilitate internetworking with ATM to support cut-forwarding switching and MPLS-based services through the satellite network. Note that a standard ATM cell can be formed easily by replacing the 6-B header with a 5-B ATM header at a gateway. The 6-B header is used so that more control information can be included. A minislot is $1/t$th of a slot. We assume $t = 6$ in this paper. To develop the dynamic reservation protocol, it is assumed that the contention results for frame $k$ and the slot assignments for frame $k + 1$ can be broadcasted via the downlink channel just before the beginning of frame $k + 1$. To ensure this, the total number of minislots in each frame is limited to $t \times \lfloor (F/2) \rfloor$. Here, we assume that the round-trip delay for LEO satellites is about 10–12 ms [8]–[11], and for simplicity any processing delay is ignored. Therefore, it should be possible to satisfy the above requirements with the chosen parameters. Note that the above parameters are just an example. It is possible to apply the dynamic reservation protocol for other situations (e.g., by using a different frame/slot size, etc.). To perform the later analysis, two types of traffic are considered as follows.

### A. CO Traffic (e.g., Voice)

We consider voice an example of CO traffic. The system supports 16-kb/s voice terminals, which are modeled by the popularly used dual-state (i.e., active–idle) source (e.g., see [7]). The independent active and idle periods follow an exponential distribution, with average values of $1/\beta = 1$ s and $1/\alpha = 1.35$ s, respectively [7]–[11]. For simplicity, it is assumed that a voice terminal alters its state at the frame borders only. This means that an idle (active) voice terminal will become active (idle) in the following frame with probability $P_{ia} = 1 - \exp(-\alpha f)(P_{ai} = 1 - \exp(-\beta f))$, where $\exp(\cdot)$ denotes an exponential function or retain its current state with probability $1 - P_{ia}(1 - P_{ai})$. An active voice terminal requires one slot in each frame for transmitting the voice packets. If a voice packet cannot be transmitted

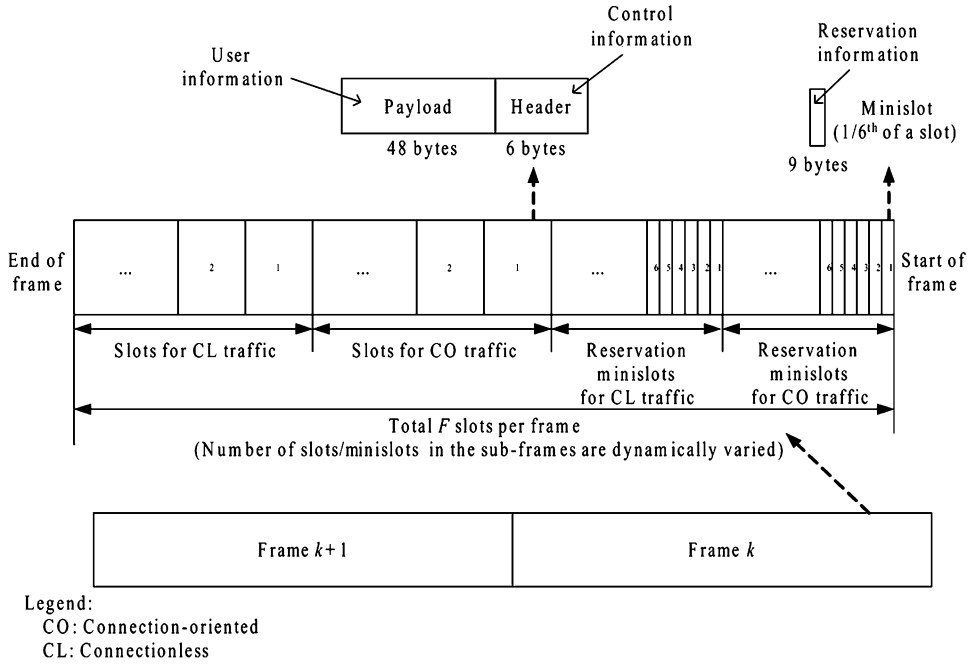


Fig. 2. Frame format for the uplink channel.

within one frame time, it will be replaced by the next voice packet (i.e., discarded). Note that the above dual-state model can also be extended to represent video traffic by aggregating the dual-state sources (or minisources) [15].

As discussed in the next section, an active voice terminal first connects to the corresponding satellite by accessing a minislot. Upon connection and subject to slot availability, the voice terminal can reserve one slot in each frame to send its voice packets. The voice packets can be forwarded through the MPLS/ATM-enabled satellites in two different ways. The simplest is by using ATM switching, in which the voice packet of each frame is switched independently. The second is by using MPLS to support IP telephony services. In this case, the voice packets of the same voice terminal collected from several frames are combined to form an IP packet. The IP packet is then transferred through the satellite network by using MPLS. To enhance the channel efficiency, the IP header information (e.g., destination and source IP addresses) can be cached or stored at the satellite during the connection phase. Furthermore, a short identifier is used for mapping to the IP header. By doing so, each voice terminal only needs to include a short identifier in the 6-B slot header to map to the corresponding IP header. Based on the mapping, the respective IP packet can be generated accordingly.

### *B. CL Traffic*

For CL traffic, the bursty data source model and the corresponding parameters given in [16] is adopted. Each data terminal is represented as an ON-OFF (i.e., active–idle) source, with the active and idle durations (e.g., $D$ with value $d$) each governed by a Weibull distribution as follows (see [16] for details):

$$\text{Prob}(D \leq d) = 1 - \exp(-(d/\psi)^{\theta}). \tag{1}$$

Based on [16], an active data terminal generates packets according to a Poisson process, with a mean rate of 20 packets/s and with each packet requiring six slots for transmission. We assume that each data terminal has a very large packet buffer.

Like the CO terminals, the active data (CL) terminals first connect to the satellite by accessing a minislot. After connection, the data segments of an IP packet can be sent through the assigned slots. Having received all of the data segments, the IP packet can then be processed accordingly (e.g., encapsulated with AAL5) and forwarded across the satellites by using MPLS.

## III. Analytical Framework and Dynamic Reservation Protocol

In this section, we first formulate an analytical model to investigate a fundamental question related to the design of the dynamic reservation protocol. Then, to implement the protocol, we present a novel contention-pattern-analysis method to estimate the number of contending terminals. Finally, we describe the protocol operation. Let us first investigate the following question: "If there are a certain number of contending terminals at the beginning of a frame, how should the access probabilities be varied in order to maximize the number of successful terminals at the end of the frame?" To investigate this interesting issue, we consider that in frame $k$, the contending terminals are classified as follows:

- nontried: contending terminals that have not tried to access a minislot in the frame;
- tried: contending terminals that have attempted to access a minislot in the frame, but that have been unsuccessful due to access conflicts (i.e., collision);
- successful: contending terminals that have successfully accessed a minislot in the frame.

The purpose is to investigate whether the tried/successful terminals should use a lower access probability. Note that due to delayed feedback, each contending terminal only knows whether it has tried or not tried to access a minislot. The successful case is for purposes of calculation only. We consider the state after the $m$th minislot as $(\sigma_{k,m}, \tau_{k,m}, \gamma_{k,m})$, where there are $\sigma_{k,m}$ nontried terminals, $\tau_{k,m}$ tried terminals, and $\gamma_{k,m}$ successful/connected terminals. The nontried terminals and the tried/successful terminals access the next minislot with probabilities $p_{k,m}$ and $q_{k,m}$, respectively. We use $m = 0$ and $m = w_k$ to denote the situations at the beginning and end of frame $k$, respectively. This means that there are $w_k$ minislots in frame $k$. It should be noted that, after the last minislot, the state remains unchanged until the end of the frame. The following are two methods of varying the access probabilities for a delayed feedback environment (i.e., the contention results are not available until the end of a frame): frame-based and minislot-based.

### A. Frame-Based Approach

Given the state $(\sigma_{k,m}, \tau_{k,m}, \gamma_{k,m})$ and the access probabilities $p_{k,m}$ and $q_{k,m}$, we first determine the state transition probability for changing from $(\sigma_{k,m}, \tau_{k,m}, \gamma_{k,m})$ to $(\sigma_{k,m+1}, \tau_{k,m+1}, \gamma_{k,m+1})$ The transition probability is denoted as $P((\sigma_{k,m+1}, \tau_{k,m+1}, \gamma_{k,m+1}) \,|\, (\sigma_{k,m}, \tau_{k,m}, \gamma_{k,m}, p_{k,m}, q_{k,m}))$, which can be found in Appendix I.

Starting from the initial state $(\sigma_{k,0}, \tau_{k,0}, \gamma_{k,0})$ and using the given access probabilities, we can calculate recursively the probability that after $w_k$ minislots, the final state becomes $(\sigma_{k,w_k}, \tau_{k,w_k}, \gamma_{k,w_k})$. We denote the respective transition probability (i.e., changing from $(\sigma_{k,0}, \tau_{k,0}, \gamma_{k,0})$ to $(\sigma_{k,w_k}, \tau_{k,w_k}, \gamma_{k,w_k})$) within the frame as $Q((\sigma_{k,w_k}, \tau_{k,w_k}, \gamma_{k,w_k}) | (\sigma_{k,0}, \tau_{k,0}, \gamma_{k,0}))$.

Having found the final state $(\sigma_{k,w_k}, \tau_{k,w_k}, \gamma_{k,w_k})$, we can compute the expected number of connected or successful terminals (i.e., the terminals that can access a minislot successfully without counting the duplicated terminals) as follows:

$$G(\gamma_{k,w_k}) = \sum_{(\sigma_{k,w_k}, \tau_{k,w_k}, \gamma_{k,w_k})} Q((\sigma_{k,w_k}, \tau_{k,w_k}, \gamma_{k,w_k}) | (\sigma_{k,0}, \tau_{k,0}, \gamma_{k,0})) \times \gamma_{k,w_k}. \quad (2)$$

For the frame-based approach, our objective is to maximize the expected number of successful terminals at the end of a frame by choosing the best values of $p_{k,m}$ and $q_{k,m}$, where $0 \le p_{k,m} \le 1$ and $0 \le q_{k,m} \le 1$. Note that for the frame-based approach, the access probabilities remain unchanged within a frame. The above model can in fact cover both the PRMA and PRMA-HS protocols by setting $\{p_{k,m} = 0.3, q_{k,m} = 0\}$ and $\{p_{k,m} = 0.6, q_{k,m} = 0.6\}$, respectively. Note that the specified probabilities are examples. However, it is not straightforward to determine the optimal access probabilities. For the dynamic reservation protocol, we consider that the access probabilities can be dynamically varied based on the initial number of contending terminals; i.e., $p_{k,m} = q_{k,m} = (1/\sigma_{k,0})$. We refer to this as the frame-based method. Here, let us assume that the initial number of contending terminals can be known. Later, we will present a novel contention-pattern analysis method to estimate $\sigma_{k,0}$. Another related problem is to determine how many minislots should be assigned for the contending terminals. As shown in Appendix II, if $p_{k,m} = q_{k,m} = (1/\sigma_{k,0})$, the average number of minislots required for all of the $\sigma_{k,0}$ contending terminals to become connected is

$$\xi(\sigma_{k,0}) = \frac{1}{\frac{1}{\sigma_{k,0}} \times \left(1 - \frac{1}{\sigma_{k,0}}\right)^{\sigma_{k,0}-1} \times \left(1 + \frac{1}{2} + \cdots + \frac{1}{\sigma_{k,0} - 1} + \frac{1}{\sigma_{k,0}}\right)}. \quad (3)$$

Furthermore, the upper and lower bounds of (3) are also worked out in Appendix II. Unless otherwise specified, we use (3) to determine the required number of minislots in a frame (i.e., $w_k$) based on the initial number of contending terminals. For a fair comparison, the formula is also applied to other schemes. Fig. 3 shows the number of minislots for a different number of contending terminals. As expected, the exact values are within the lower and upper bounds.

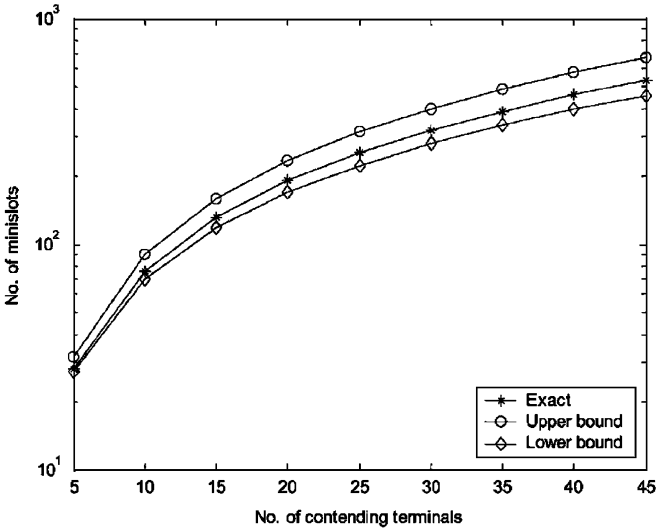


Fig. 3. Number of minislots when the number of contending terminals varies.

Theoretically, we can determine the best combination of $(p_{k,m}, q_{k,m})$ in order to maximize the expected value, as calculated by (2). To investigate this issue, Fig. 4 shows the success ratio (i.e., the fraction of the contending terminals that can be connected) for different values of $p$ and $q$ when the initial number of contending terminals is 25. For simplicity, the subscripts are removed. It can be seen that the success ratio is more sensitive to changes in $q$. In particular, when $q$ is around $1/25$, the success ratio can be kept at a high value for a wide range of $p$. Fig. 5 gives the best values of $p$ and $q$ for the frame-based (exact) method when the number of contending terminals is varied. Obviously, when there is only one contending terminal, $p$ should be set to one so that the terminal can access the minislot with certainty. As an extension to Fig. 5, Fig. 6 shows the best values of $p$ and $q$ normalized with respect to the reciprocal of the initial number of contending terminals. It can be seen that when the initial number of contending terminals is large, $p$ and $q$ should be set to about 1.2 times and 0.8 times the reciprocal of the initial number of contending

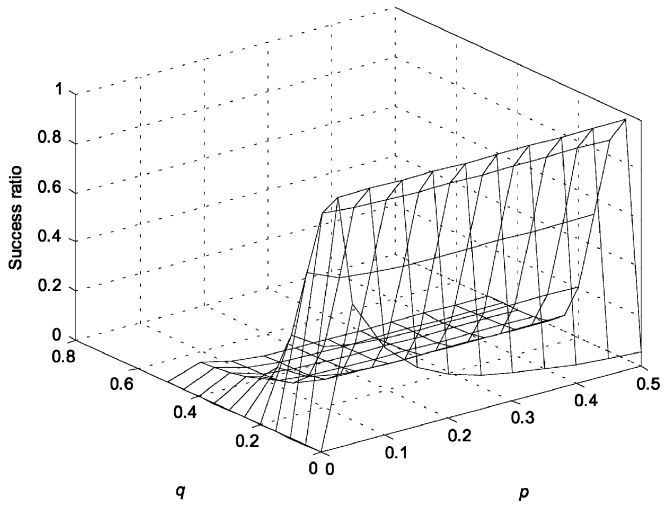


Fig. 4. Success ratio when $p$ and $q$ varies.

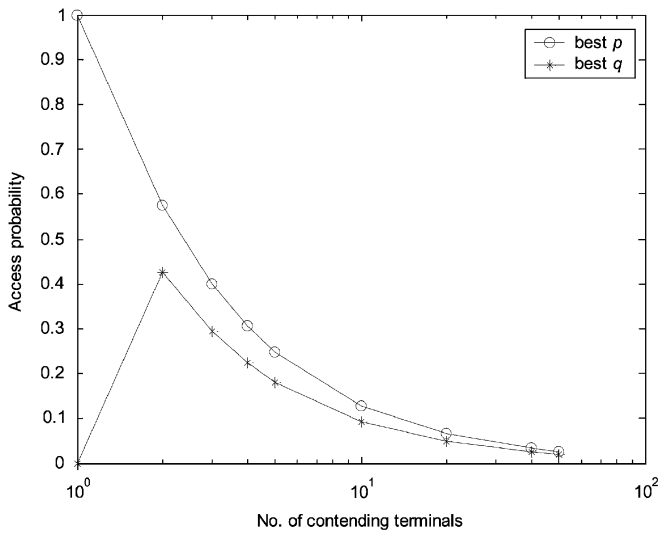


Fig. 5. Best access probability when the number of contending terminals varies.

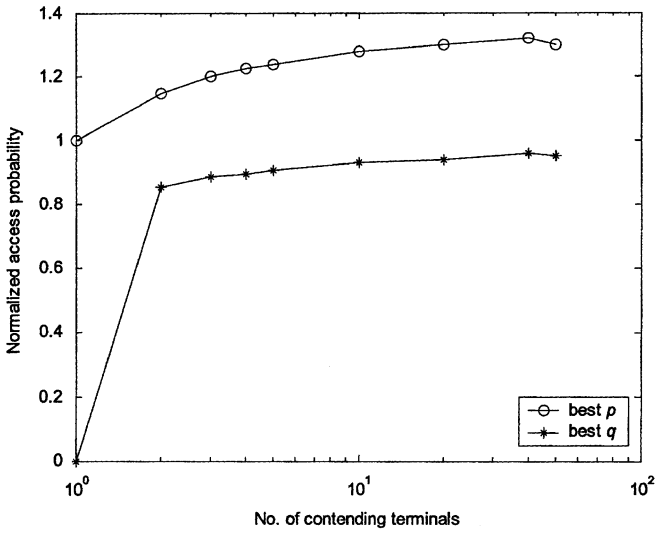


Fig. 6. Best normalized access probability when the number of contending terminals varies.

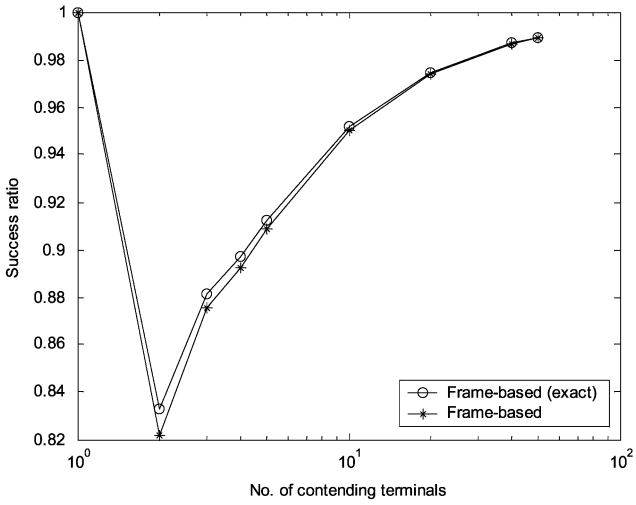


Fig. 7. Comparison of the success ratio between the frame-based (exact) scheme and frame-based scheme.

terminals, respectively. However, Fig. 7 shows that the success ratios for the frame-based and frame-based (exact) methods are in fact very close, which indicates the effectiveness of using the frame-based method. Hence, to facilitate implementation, we set $p_{k,m} = q_{k,m} = (1/\sigma_{k,0})$ for the frame-based scheme.

### *B. Minislot-Based Approach*

Next, we investigate how to vary $(p_{k,m}, q_{k,m})$ (i.e., the access probabilities for the next minislot) so that the chance of accessing the minislot by using one of the nontried or tried terminals can be maximized. Recall that due to delayed feedback, a tried terminal cannot know whether the previous access is successful until the beginning of the next frame. Mathematically, we want to determine the best combination of $(p_{k,m}, q_{k,m})$ in order to maximize the value of the following function:

$$\begin{aligned} &S(\sigma_{k,m}, \tau_{k,m}, \gamma_{k,m}) \\ &\quad = B(\sigma_{k,m}, 1, p_{k,m}) \times (1 - q_{k,m})^{\tau_{k,m} + \gamma_{k,m}} \\ &\qquad + (1 - p_{k,m})^{\sigma_{k,m}} \times B(\tau_{k,m}, 1, q_{k,m}) \times (1 - q_{k,m})^{\gamma_{k,m}} \end{aligned} \quad (4)$$

subject to the condition that $0 \le p_{k,m}, q_{k,m} \le 1$, where $B(x, y, z) = \binom{x}{y} \times z^y \times (1 - z)^{x-y}$. Note that (4) gives the probability that one of the nontried or tried terminals will successfully access a minislot. To determine the best values $(p^*_{k,m}, q^*_{k,m})$, we need to consider the turning point(s) and boundary conditions as follows:

*Case 1:* The values of $p_{k,m}$ and $q_{k,m}$ are found by considering the turning point where the derivatives of $S(\sigma_{k,m}, \tau_{k,m}, \gamma_{k,m})$ with respect to both $p_{k,m}$ and $q_{k,m}$ are zero; i.e.,

$$\begin{aligned} &\frac{\partial S(\sigma_{k,m}, \tau_{k,m}, \gamma_{k,m})}{\partial p_{k,m}} = 0 \quad \text{and} \\ &\frac{\partial S(\sigma_{k,m}, \tau_{k,m}, \gamma_{k,m})}{\partial q_{k,m}} = 0. \end{aligned} \quad (5)$$

After calculation, it is found that the optimum values of $p_{k,m}, q_{k,m}$ are

$$p_{k,m} = \frac{1 - \gamma_{k,m}}{\tau_{k,m} + \sigma_{k,m}} \quad (6)$$

$$q_{k,m} = \frac{\sigma_{k,m} \times \gamma_{k,m} + \tau_{k,m}}{\tau_{k,m}^2 + \tau_{k,m} \times \gamma_{k,m} + \sigma_{k,m} \times \gamma_{k,m} + \sigma_{k,m} \times \tau_{k,m}}. \tag{7}$$

As can be seen, this case is invalid unless $\gamma_{k,m} = 0$ or 1. If this condition is not satisfied, $p_{k,m}$ becomes negative.

*Case 2:* We need to find the maximum value of $q_{k,m}$ when $p_{k,m} = 0$. This is done by solving $(\partial S(\sigma_{k,m}, \tau_{k,m}, \gamma_{k,m})/\partial q_{k,m}) = 0$ when $p_{k,m} = 0$. The access probabilities for this case are

$$p_{k,m} = 0, \quad q_{k,m} = \frac{1}{\tau_{k,m} + \gamma_{k,m}}. \tag{8}$$

The corresponding value of $S$ is

$$S(\sigma_{k,m}, \tau_{k,m}, \gamma_{k,m}) = \frac{\tau_{k,m}}{\tau_{k,m} + \gamma_{k,m}} \times (1 - q_{k,m})^{\gamma_{k,m} + \tau_{k,m} - 1}. \tag{9}$$

*Case 3:* Similar to case 2, we determine the maximum value of $p_{k,m}$ when $q_{k,m} = 0$ (i.e., by solving $(\partial S(\sigma_{k,m}, \tau_{k,m}, \gamma_{k,m}))/(\partial p_{k,m}) = 0$ when $q_{k,m} = 0$). We get

$$p_{k,m} = \frac{1}{\sigma_{k,m}}, \quad q_{k,m} = 0. \tag{10}$$

The corresponding value of $S$ is

$$S(\sigma_{k,m}, \tau_{k,m}, \gamma_{k,m}) = (1 - p_{k,m})^{\sigma_{k,m} - 1}. \tag{11}$$

Furthermore, we should also consider the cases where $p_{k,m} = 1$ or $q_{k,m} = 1$, as well as $p_{k,m} = 0$ and $q_{k,m} = 0$. It is not difficult to see that they are not the best choices.

Comparing the above cases, it can be found that the best values of $p_{k,m}, q_{k,m}$ are shown in (12), at the bottom of the page. Thus, if the access probabilities are varied as $(p_{k,m}^*, q_{k,m}^*)$, the number of successful terminals at the end of a frame can be maximized. However, the actual value of $(\sigma_{k,m}, \tau_{k,m}, \gamma_{k,m})$ cannot be known in reality. Here, we implement the minislot-based scheme as follows. Consider that the probability for changing from $(\sigma_{k,0}, \tau_{k,0}, \gamma_{k,0})$ to $(\sigma_{k,m}, \tau_{k,m}, \gamma_{k,m})$ is $R_m((\sigma_{k,m}, \tau_{k,m}, \gamma_{k,m}) | (\sigma_{k,0}, \tau_{k,0}, \gamma_{k,0}))$. Let the expected value of $(\sigma_{k,m}, \tau_{k,m}, \gamma_{k,m})$ be $(\sigma'_{k,m}, \tau'_{k,m}, \gamma'_{k,m})$, which can be worked out as follows:

$$\sigma'_{k,m} = \sum_{(\sigma_{k,m}, \tau_{k,m}, \gamma_{k,m})} R_m((\sigma_{k,m}, \tau_{k,m}, \gamma_{k,m}) | (\sigma_{k,0}, \tau_{k,0}, \gamma_{k,0})) \times \sigma_{k,m} \tag{13}$$

$$\tau'_{k,m} = \sum_{(\sigma_{k,m}, \tau_{k,m}, \gamma_{k,m})} R_m((\sigma_{k,m}, \tau_{k,m}, \gamma_{k,m}) | (\sigma_{k,0}, \tau_{k,0}, \gamma_{k,0})) \times \tau_{k,m} \tag{14}$$

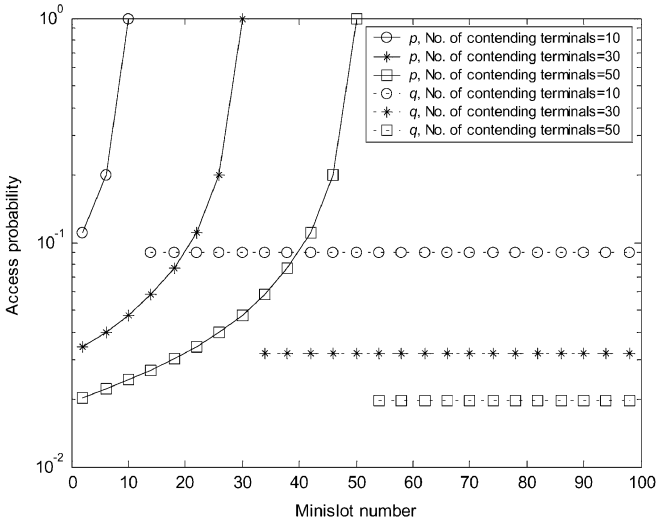


Fig. 8. Change in access probabilities for the minislot-based (exact) scheme.

$$\gamma'_{k,m} = \sum_{(\sigma_{k,m}, \tau_{k,m}, \gamma_{k,m})} R_m((\sigma_{k,m}, \tau_{k,m}, \gamma_{k,m}) | (\sigma_{k,0}, \tau_{k,0}, \gamma_{k,0})) \times \gamma_{k,m}. \tag{15}$$

With the expected values, we can calculate the expected optimum access probability $(p_{k,m}^*, q_{k,m}^*)$ by using (12), and then determine $R_{m+1}((\sigma_{k,m+1}, \tau_{k,m+1}, \gamma_{k,m+1}) | (\sigma_{k,0}, \tau_{k,0}, \gamma_{k,0}))$ accordingly based on (30) (see Appendix I). Hence, starting from the initial state $(\sigma_{k,0}, \tau_{k,0}, \gamma_{k,0})$ and $m = 0$, we can compute the expected optimum access probabilities recursively. We refer to this method and to the exact solution (i.e., where the access probabilities are calculated based on the exact value of $(\sigma_{k,m}, \tau_{k,m}, \gamma_{k,m})$) as the minislot-based and the minislot-based (exact) schemes, respectively.

Fig. 8 shows how the access probabilities should be varied for the minislot-based method. Again, the subscripts are ignored for simplicity. The figure indicates that the best strategy is to disable the tried terminals below a certain number of minislots and to increase the access probability for the nontried terminals (i.e., $p$) exponentially toward one. When all of the contending terminals have tried to access a minislot, $q$ should be set to the reciprocal of the number of contending terminals.

Finally, let us evaluate the performance of the different schemes. Fig. 9 compares the success ratio for the different schemes when the number of contending terminals is varied. When the number of contending terminals is less than four, a better performance can be achieved by using $p = 0.6$ and $q = 0.6$ (similar to PRMA-HS). However, if the number of contending terminals is above four, using $p = 0.3$ and $q = 0$ (similar to the original PRMA) is better. As expected, the

$$(p_{k,m}^*, q_{k,m}^*) = \begin{cases} \left(\frac{1}{\sigma_{k,m}}, 0\right), & \text{if } \sigma_{k,m} > 0 \quad \text{and} \quad (\gamma_{k,m} > 0 \text{ or } \tau_{k,m} > \sigma_{k,m} \text{ or } \tau_{k,m} = 0) \\ \left(0, \frac{1}{\tau_{k,m} + \gamma_{k,m}}\right), & \text{if } \sigma_{k,m} = 0 \quad \text{or} \quad (\gamma_{k,m} = 0 \text{ and } 0 < \tau_{k,m} \leq \sigma_{k,m}) \end{cases} \tag{12}$$

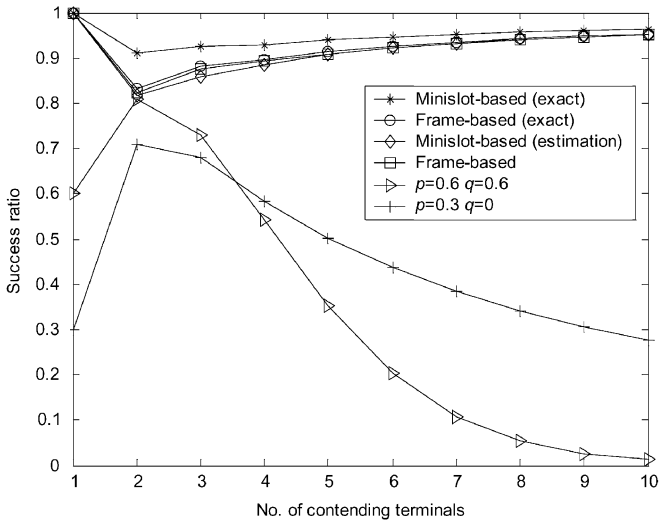


Fig. 9. Success ratio for different schemes when the number of contending terminals varies.

minislot-based (exact) method always gives the best performance. However, it is only an ideal scheme, which cannot be realized in practice. The other three schemes: frame-based (exact), frame-based, and minislot-based give almost the same performance close to the ideal result. This justifies the use of the frame-based scheme for the dynamic reservation protocol because it is simple to implement. In the rest of this paper, we will focus on using this scheme.

### C. Contention-Pattern-Analysis Algorithm

To vary the access probability dynamically using the above frame-based scheme, we need to estimate the number of contending terminals at the beginning of a frame. In this section, we introduce a novel contention-pattern-analysis algorithm for this purpose. The key question is: "Given a certain contention pattern, what number of contending terminals is most likely to produce the pattern?" The algorithm works as follows. We define $a$ as the possible number of contending terminals and $b_m$ as the number of terminals that have successfully sent their requests just after the $m$th minislot. If the number of contending terminals at the beginning of the frame is $a$, we denote $E_m$ as the probability that the pattern up to the $m$th minislot is generated. At the beginning of frame $k$, we assume that the possible numbers of contending terminals are: $0, 1, 2, 3, \ldots, u$ (i.e., $a$ can be one of these values) and that there are $w_k$ minislots. The algorithm is described as follows.

1) Set $a = 0$.
2) Set $b_0 = 0$ and $E_0 = 1$.
3) Starting from $m = 1$, for each minislot $m$, update $E_m$ based on the contention result as follows.
   - If the result is "idle" (i.e., no terminal accesses the minislot), update $E_m = E_{m-1} \times (1 - p_{k,m-1})^a$.
   - If the result is "success" (i.e., one terminal accesses the minislot and it has not previously sent a request successfully), update $E_m = E_{m-1} \times (a - b_{m-1}) \times p_{k,m-1} \times (1 - p_{k,m-1})^{a-1}$. Furthermore, set $b_m = b_{m-1} + 1$.
   - If the result is "duplication" (i.e., one terminal accesses the minislot but it has previously sent a request successfully), update $E_m = E_{m-1} \times b_{m-1} \times p_{k,m-1} \times (1 - p_{k,m-1})^{a-1}$.
   - If the result is "collision," (i.e., more than one terminal accesses the minislot) update $E_m = E_{m-1} \times (1 - (1 - p_{k,m-1})^a - a \times p_{k,m-1} \times (1 - p_{k,m-1})^{a-1})$.
4) After going through all of the $w_k$ minislots, set $Z_a = E_{w_k}$. Increase $a$ by one. If $a \leq u$, repeat steps 2)–4); otherwise, stop.

Finally, find $x$ (where $x = 0, 1, 2, 3, \ldots u$) such that $Z_x$ is the largest. Denote $\varepsilon_{k,0}$ and $\varepsilon_{k,w_k}$ as the estimated number of contending terminals at the start and end of frame $k$. Based on the above, we set $\varepsilon_{k,0} = x$. Knowing that $\gamma_{k,w_k}$ terminals have accessed a minislot successfully within the frame, we have $\varepsilon_{k,w_k} = \varepsilon_{k,0} - \gamma_{k,w_k}$. For simplicity, we can set $\varepsilon_{k+1,0} = \varepsilon_{k,w_k}$. Note that if the terminal model is known, the estimate can be further enhanced based on the terminal model. For the case of voice, we have $\varepsilon_{k+1,0} = \lceil \varepsilon_{k,w_k} + (N_v - r_k - \varepsilon_{k,w_k}) \times P_{ia} - \varepsilon_{k,w_k} \times P_{ai} \rceil$, where $N_v$ is the number of voice/bistate terminals and $r_k$ is the number of connected/reserved voice/bistate terminals at the frame border. Note that the second and third terms in the equation give the average number of terminals entering and leaving the active state, respectively. To convert the new estimate to an integer, the ceiling function is used. The access probabilities for frame $k + 1$ are updated to $p_{k+1,m} = q_{k+1,m} = (1/\mathrm{Max}(2, \varepsilon_{k+1,0}))$. Here, we assume a lower bound of two rather than one contending terminal(s) to prevent excessive collisions in the case of occasional underestimations. Furthermore, the number of minislots for the next frame is also set accordingly.

Let us use a simple example to explain the above method. Suppose that the current frame has four minislots and that at the end of the frame the contention results are *collision, idle, success*, and *duplication*. Assume that initially there can be 0, 1, 2, and 3 contending terminals and the access probability for the previous frame is $1/3$. For each possible number of contending terminals, the base terminal calculates the probability that the above contention pattern is generated (see Table I). In this example, the predicted number of contending terminals is three because it has the highest chance of producing the contention pattern: *collision, idle, success*, and *duplication*. Since one of the contending terminals has accessed a minislot successfully, the estimated number of contending terminals becomes $3 - 1 = 2$. Furthermore, the access probability should be updated to $1/2$.

We have conducted many simulations to evaluate the effectiveness of the contention-pattern-analysis method. Some representative results are presented in Fig. 10. Here, we consider that the largest possible number of contending terminals is 50 and the number of minislots is set based on (3). We repeated each case 100 000 times to obtain the results. Fig. 10 shows the distribution (frequency percent of the 100 000 runs) of the estimated number of contending terminals for a different actual number of contending terminals. It can be seen that the estimate is generally quite good. In other words, the prediction

TABLE I
SIMPLE EXAMPLE TO ILLUSTRATE THE CONTENTION-PATTERN-ANALYSIS METHOD

| Initial number | 1st minislot (collision) | 2nd minislot (idle) | 3rd minislot (success) | 4th minislot (duplication) |
|---|---|---|---|---|
| $a=0$<br>$b_0=0$ | $E_1=E_0\times\left(1-a\times p_{k,0}\times(1-p_{k,0})^{a-1}-(1-p_{k,0})^{a}\right)$<br>$=0$<br>$b_1=0$ | | | |
| $a=1$<br>$b_0=0$ | $E_1=E_0\times\left(1-a\times p_{k,0}\times(1-p_{k,0})^{a-1}-(1-p_{k,0})^{a}\right)$<br>$=0$<br>$b_1=0$ | | | |
| $a=2$<br>$b_0=0$ | $E_1=E_0\times\left(1-a\times p_{k,0}\times(1-p_{k,0})^{a-1}-(1-p_{k,0})^{a}\right)$<br>$=\frac{1}{9}$<br>$b_1=0$ | $E_2=E_1\times(1-p_{k,1})^{a}$<br>$=\frac{4}{81}$<br>$b_2=0$ | $E_3=E_2\times(a-b_2)\times p_{k,2}\times(1-p_{k,2})^{a-1}$<br>$=\frac{16}{729}$<br>$b_3=1$ | $E_4=E_3\times b_3\times p_{k,3}\times(1-p_{k,3})^{a-1}$<br>$=\frac{32}{6561}\approx 0.00487$<br>$b_4=1$ |
| $a=3$<br>$b_0=0$ | $E_1=E_0\times\left(1-a\times p_{k,0}\times(1-p_{k,0})^{a-1}-(1-p_{k,0})^{a}\right)$<br>$=\frac{7}{27}$<br>$b_1=0$ | $E_2=E_1\times(1-p_{k,1})^{a}$<br>$=\frac{56}{729}$<br>$b_2=0$ | $E_3=E_2\times(a-b_2)\times p_{k,2}\times(1-p_{k,2})^{a-1}$<br>$=\frac{224}{6561}$<br>$b_3=1$ | $E_4=E_3\times b_3\times p_{k,3}\times(1-p_{k,3})^{a-1}$<br>$=\frac{896}{177147}\approx 0.00505$<br>$b_4=1$ |

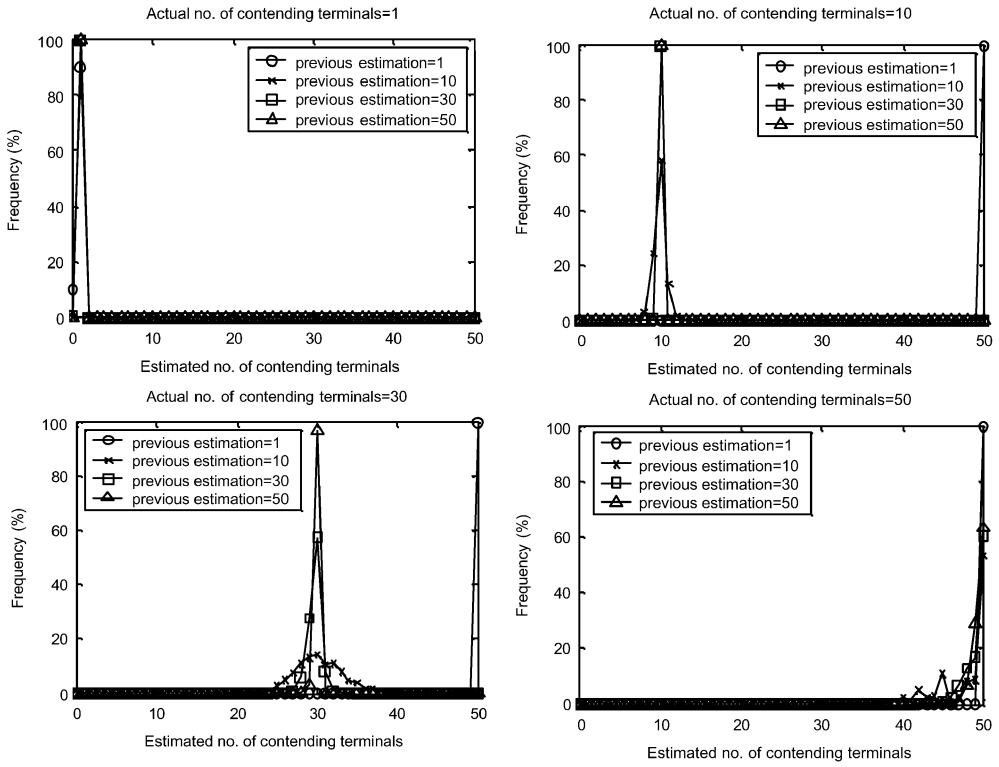


Fig. 10. Distribution of the estimated number of contending terminals.

matches closely with the actual number of contending terminals except when the previous estimate is significantly underestimated. However, in this case, the next estimate will move closer to the maximum number of contending terminals. This eventually brings the estimate close to the actual number of contending terminals. In summary, the contention-pattern-analysis method provides an effective way of estimating the number of contending terminals for the dynamic reservation protocol.

### D. Protocol Operation

Based on the above analytical model, the dynamic reservation protocol works as follows. Note that our focus is on packet-

level operations. We assume that the terminals have registered with the satellite through an appropriate mechanism. According to the contention pattern in the last frame, the satellite determines the access probability and number of reservation minislots for each type of traffic (i.e., CO and CL traffic). The minislots are then assigned based on the available (nonreserved) channel capacity. Basically, the minislots required for the voice (CO) traffic are allocated first and the unused capacity (i.e., after all of the outstanding transmission requests have been served) is assigned as minislots for the data (CL) traffic. The access probability and number of minislots for each type of traffic are broadcasted through the downlink channel, together with the slot assignments. The slot assignments (i.e., determining which terminal can use a specific slot) can be conveyed via the downlink slots. The terminals access the minislots/slots as follows.

When an idle voice terminal becomes active, it first sends a reservation request through an eligible minislot by using the respective access probability as broadcasted by the satellite. Based on the reservation requests, the satellite reserves the slots for the voice terminals accordingly. As mentioned above, each reserved voice terminal needs to keep track of the slot assignment for the subsequent frame via the downlink channel (similar to [17]). After sending the last voice packet of the talkspurt, the reservation is released. Note that the above protocol can also be extended to support variable bit rate traffic. In this case, a reserved terminal can change the bit rate (i.e., the required number of slots per frame) through the transmitted packets so that further slots (if available) can be assigned or reserved accordingly.

Similar to the voice terminals, an active data terminal first accesses an eligible minislot (i.e., a minislot for the data terminals) with the corresponding access probability, as predicted by the above contention-pattern-analysis method. Note that the access probability for the data terminals is likely to be different from that of the voice terminals. After accessing a minislot, the data terminal notifies the satellite about the number of slots required to transmit its packets. Having collected all of the requests, the available slots are assigned to the data terminals on a "first-come, first-served" basis. Once a data terminal is connected to the satellite, later slot requirements can be piggybacked through the transmitted packets, based on a similar approach used in [16]–[18] (i.e., the information is included in the header of each transmitted packet). Based on the information, the satellite can record the slot requirements and then assign the slots accordingly. By employing this method for transmitting data packets, the channel efficiency can be improved. After sending the last packet in the buffer, the data terminal is logically disconnected from the satellite (i.e., the data terminal needs to access a minislot again in order to transmit a new packet).

## IV. Performance Analysis and Discussion

In this section, we analyze the performance of the dynamic reservation protocol by using a Markov model and computer simulations. We first formulate a Markov model to evaluate the performance of the dynamic reservation protocol with the frame-based scheme for multiplexing bistate CO traffic sources (e.g., voice terminals). For purposes of comparison, we also analyze the cases $\{p = 0.6, q = 0.6\}$ and $\{p = 0.3, q = 0\}$ with the Markov model. Note that the frame-based and the frame-based (estimation) (i.e., the scheme based on the contention-pattern-analysis method) schemes use the actual and estimated number of contending terminals to compute the access probability. Hence, the frame-based scheme gives the best possible performance that can be used to evaluate the effectiveness of the frame-based (estimation) scheme. In addition, the Markov model is also employed to validate the simulation model with $F = 15$. Unfortunately, it is computationally infeasible to use the Markov model when $F$ is large because there are too many states. Consequently, we employ the validated simulation model to evaluate the performance of the system for large values of $F$. For CL traffic, the performance is analyzed using simulations.

The Markov model for CO traffic is described as follows. A Markov chain is formed with states represented by the number of contending terminals and reserved terminals at the end of a frame. Consider the situation at the end of frame $k$, where there are $r_k$ reserved terminals, $c_k$ contending terminals, and $i_k = N_v - r_k - c_k$ idle terminals. In general, the following case can happen: $\eta_k$ of the $r_k$ terminals become idle, $\phi_k$ of the $i_k$ terminals become active, and $\delta_k$ of the $c_k$ terminals finish their active periods. The corresponding probabilities are denoted as $B(r_k, \eta_k, P_{ai}), B(i_k, \phi_k, P_{ia})$, and $B(c_k, \delta_k, P_{ai})$, respectively.

It is not difficult to see that $c'_{k+1} = c_k + \phi_k - \delta_k$ and $r'_{k+1} = r_k - \eta_k$, where $c'_{k+1}$ and $r'_{k+1}$ denote the number of contending and reserved terminals at the start of frame $k+1$. Hence, we have

$$\delta_k = \phi_k + c_k - c'_{k+1} \tag{16}$$

and

$$\eta_k = r_k - r'_{k+1}. \tag{17}$$

We use $\Gamma_1((c'_{k+1}, r'_{k+1}) \,|\, (c_k, r_k))$ to denote the respective transition probabilities at the border of the frame. Note that the state at the end of a frame is the same as the state after the last minislot of the frame. The transition probability can be found by considering the general case as given above, i.e.,

$$\begin{aligned}
&\Gamma_1((c'_{k+1}, r'_{k+1}) \,|\, (c_k, r_k)) \\
&\quad = \sum_{\phi_k = \max(0, c'_{k+1} - c_k)}^{\min(c'_{k+1}, N_v - r_k - c_k)} B(i_k, \phi_k, P_{ia}) \\
&\qquad \times B(c_k, \delta_k, P_{ai}) \times B(r_k, \eta_k, P_{ai}) \\
&\quad = \sum_{\phi_k = \max(0, c'_{k+1} - c_k)}^{\min(c'_{k+1}, N_v - r_k - c_k)} B(N_v - c_k - r_k, \phi_k, P_{ia}) \\
&\qquad \times B(c_k, \phi_k + c_k - c'_{k+1}, P_{ai}) \times B(r_k, r_k - r'_{k+1}, P_{ai}).
\end{aligned} \tag{18}$$

Note that the lower and upper limits of $\phi_k$ are found by considering the following conditions:

$$0 \le \delta_k = \phi_k + c_k - c'_{k+1} \le c_k \quad \text{and} \quad 0 \le \phi_k \le N_v - r_k - c_k. \tag{19}$$

Denote $\Gamma_2((c_{k+1}, r_{k+1}) \,|\, (c'_{k+1}, r'_{k+1}))$ as the transition probability that the state changes from $(c'_{k+1}, r'_{k+1})$ to $(c_{k+1}, r_{k+1})$ within frame $k+1$, which can be found by (30)

(see Appendix I) by setting the initial values as $\sigma_{k+1,0} = c'_{k+1}, \tau_{k+1,0} = 0, \gamma_{k+1,0} = 0$. For the number of minislots, we allocate $w_{k+1} = \text{Min}(\text{Max}(t, \xi(\sigma_{k+1,0})), t \times (F - r_k)^+, t \times \lfloor (F/2) \rfloor)$. Here, we need to satisfy a number of requirements simultaneously. Essentially, the required number of minislots is calculated by (3), subject to the maximum number of available minislots (i.e., $t \times (F - r_k)^+$), and the propagation delay requirement as explained above (i.e., no more than $t \times \lfloor (F/2) \rfloor$ minislots can be assigned). Furthermore, we assume that at least $t$ minislots are provided subject to availability. Given the state $(\sigma_{k+1,w_{k+1}}, \tau_{k+1,w_{k+1}}, \gamma_{k+1,w_{k+1}})$ at the end of frame $k+1$, there are a total of $\sigma_{k+1,w_{k+1}} + \tau_{k+1,w_{k+1}}$ contending terminals at the end of the frame. Hence, we have

$$c_{k+1} = \sigma_{k+1,w_{k+1}} + \tau_{k+1,w_{k+1}}. \tag{20}$$

As we know, there are $r'_{k+1}$ reserved terminals at the beginning of frame $k+1$, and $\gamma_{k+1,w_{k+1}}$ terminals connected to the satellite within the frame. Therefore, we have

$$r_{k+1} = r'_{k+1} + \gamma_{k+1,w_{k+1}}. \tag{21}$$

Since a frame can support at most $F$ packets, $r_{k+1} - F$ terminals cannot be assigned a slot when $r_{k+1} > F$. These terminals will return to the contending state. Therefore, the number of contending and reserved terminals at the end of frame $k+1$ can be found as follows in (22) and (23) at the bottom of the page. Hence, we get

$$\begin{aligned}&\Gamma_2((c_{k+1}, r_{k+1}) \,|\, (c'_{k+1}, r'_{k+1})) \\ &\quad = \sum_{\sigma=0}^{c_{k+1}} Q((\sigma, c_{k+1} - \sigma, r_{k+1} - r'_{k+1}) \,|\, (c'_{k+1}, 0, 0)).\end{aligned} \tag{24}$$

Having found $\Gamma_1((c'_{k+1}, r'_{k+1}) \,|\, (c_k, r_k))$ and $\Gamma_2((c_{k+1}, r_{k+1}) \,|\, (c'_{k+1}, r'_{k+1}))$, we can compute the transition probability $\Gamma((c_{k+1}, r_{k+1}) \,|\, (c_k, r_k))$ for the Markov chain as follows:

$$\begin{aligned}&\Gamma((c_{k+1}, r_{k+1}) \,|\, (c_k, r_k)) \\ &\quad = \sum_{r'_{k+1}=0}^{r_k} \sum_{c'_{k+1}=0}^{N_v - r_k} \Gamma_1((c'_{k+1}, r'_{k+1}) \,|\, (c_k, r_k)) \\ &\qquad \times \Gamma_2((c_{k+1}, r_{k+1}) \,|\, (c'_{k+1}, r'_{k+1})).\end{aligned} \tag{25}$$

Denote $\Pi(c, r)$ as the limiting probability that there are $c$ contending terminals and $r$ reserved terminals at the end of a frame, which can be found by solving the following equations:

$$\Pi(c, r) = \sum_{x=0}^{N_v} \sum_{y=0}^{N_v - x} \Pi(x, y) \times \Gamma(c, r \,|\, x, y) \tag{26}$$

$$\sum_{c=0}^{N_v} \sum_{r=0}^{N_v - c} \Pi(c, r) = 1. \tag{27}$$

Having found the limiting probabilities, we can compute the packet loss ratio and utilization as explained below. For voice traffic, it is well known that the packet loss ratio should be kept within 1% (e.g., [7] and [8]). Using $\Pi(c, r)$, the packet loss ratio (PLR) can be computed as follows:

$$\text{PLR} = \frac{\sum_{c=0}^{N_v} \sum_{r=0}^{N_v - c} \Pi(c, r) \times c}{\sum_{c=0}^{N_v} \sum_{r=0}^{N_v - c} \Pi(c, r) \times (c + r)}. \tag{28}$$

It is not difficult to see that the denominator and numerator give the mean number of packets transmitted by the terminals per frame and the average number of packets lost in a frame, respectively.

For calculating the slot utilization, idle slots and minislots are excluded. If there are $r$ reserved terminals at the end of a frame, $r/F$ of the slots are utilized in the next frame. Hence, the slot utilization (SU) is computed as follows:

$$\text{SU} = \frac{\sum_{c=0}^{N_v} \sum_{r=0}^{N_v - c} \Pi(c, r) \times r}{F}. \tag{29}$$

In addition to the above Markov model, a simulation program has also been written using C++ to analyze the performance of the dynamic reservation protocol for integrating voice (CO) and data (CL) traffic, based on some of the simulation parameters in [16]. The simulation and analytical results are presented as follows. We first consider a system with only voice terminals. Fig. 11 compares the packet loss ratio for the different schemes when the number of voice terminals is varied. It can be seen that the analytical results match closely with the simulation results, thus indicating the correctness of both models. When $F = 15$, better performance can be achieved by using $p = 0.6$ and $q = 0.6$. However, when $F$ is increased to 50 and 100, using $p = 0.6$ and $q = 0.6$ gives an unacceptable performance (nearly all of the packets are discarded). In these two cases, using $p = 0.3$ and $q = 0$ is better. The figure also shows that the packet loss ratio for the frame-based (estimation) scheme is close to that of the frame-based scheme. This means that the contention-pattern-analysis method is effective in predicting the number of contending terminals and, hence, in varying the access probability dynamically. When $F = 50$, the system can accommodate about 95 voice terminals, giving a maximum multiplexing gain of 1.9 as compared with the conventional time-division multiplexing system. Fig. 12 shows the slot utilization when the number of voice terminals is varied. Again, the simulation and analytical results match closely. It can be seen that in terms of slot utilization, all of the methods give almost the same performance. When $F = 50$ and the number of voice terminals is 95, the frame-based (estimation) scheme gives a utilization of 80%. As shown later, the utilization can be further increased by adding data traffic. Next, we consider the integration of voice and data traffic. According to [16], the active and idle durations for the data terminals are computed by setting

$$c_{k+1} = \begin{cases} \sigma_{k+1,w_{k+1}} + \tau_{k+1,w_{k+1}}, & \text{if } r'_{k+1} + \gamma_{k+1,w_{k+1}} \le F \\ \sigma_{k+1,w_{k+1}} + \tau_{k+1,w_{k+1}} + r'_{k+1} + \gamma_{k+1,w_{k+1}} - F, & \text{if } r'_{k+1} + \gamma_{k+1,w_{k+1}} > F \end{cases} \tag{22}$$

$$r_{k+1} = \begin{cases} r'_{k+1} + \gamma_{k+1,w_{k+1}}, & \text{if } r'_{k+1} + \gamma_{k+1,w_{k+1}} \le F \\ F, & \text{if } r'_{k+1} + \gamma_{k+1,w_{k+1}} > F \end{cases} \tag{23}$$

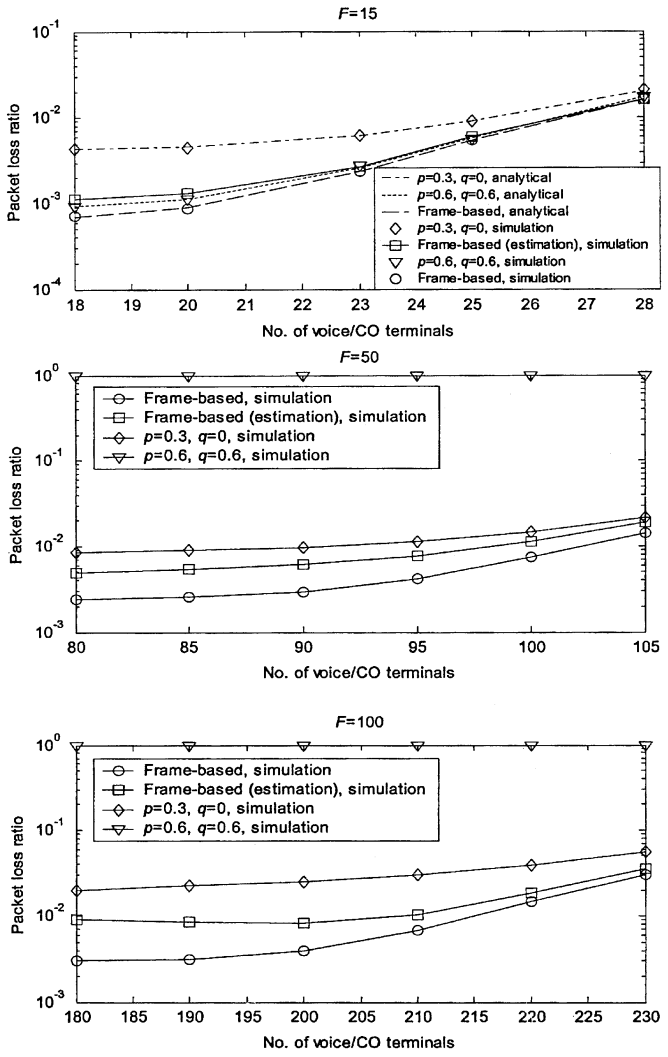


Fig. 11. Packet loss ratio for different access schemes.

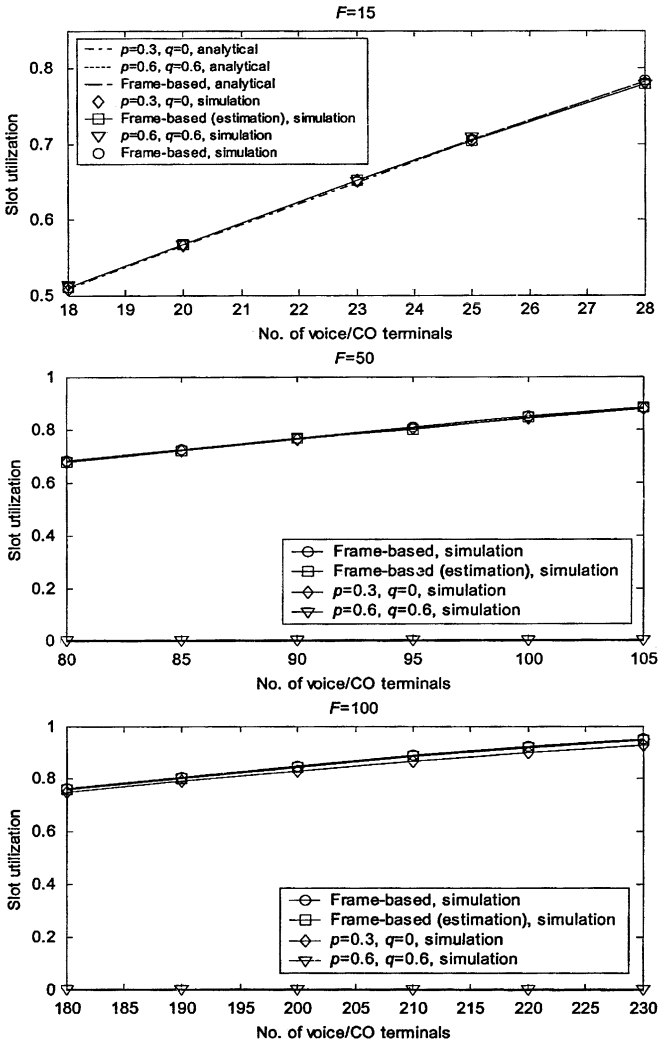


Fig. 12. Slot utilization for different access schemes.

$\{\theta = 0.88$ and $\psi = 3.10\}$ and $\{\theta = 0.88, \psi = 21.40\}$, respectively. Fig. 13 shows the mean packet delay when the number of data terminals is varied. It can be seen that the delay can be maintained at a steady low value until a certain number of data terminals is reached. For example, if there are 95 voice terminals, the system can accommodate 10 data terminals. Fig. 14 shows the slot utilization when the number of data terminals is varied. The figure indicates that the utilization increases linearly with respect to the number of data terminals. Referring back to Fig. 12, the utilization is about 80% when the number of voice terminals is 95. By introducing data traffic, the utilization can be increased to above 85%. Based on [16], we have also studied two other cases (bursty models I and III) with the same average data rate as the previous case (bursty model II) as follows:

- bursty model I: $\{\theta = 1.0$ and $\psi = 3.3\}$ for active periods and $\{\theta = 1.0, \psi = 22.8\}$ for idle periods;
- bursty model II: $\{\theta = 0.88$ and $\psi = 3.10\}$ for active periods and $\{\theta = 0.88, \psi = 21.40\}$ for idle periods;
- bursty model III: $\theta = 0.7$ and $\psi = 2.61\}$ for active periods and $\{\theta = 0.7, \psi = 18.01\}$ for idle periods.

There are 95 voice terminals in subsequent simulations. Fig. 15 shows the mean packet delay for the three cases. The result

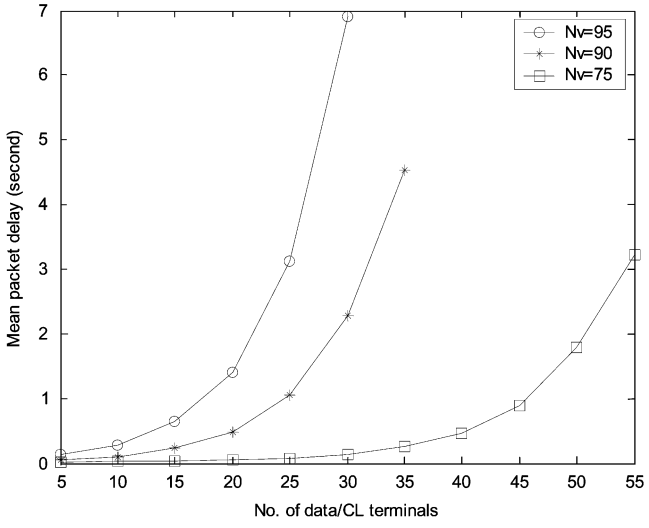


Fig. 13. Mean packet delay when the number of data terminals varies.

shows that the delay is slightly higher if $\theta$ is smaller, especially when there are more data terminals. Like [16], we have also simulated the case with exponential interarrival time (i.e., the Poisson model) using the same average data rate. It is found

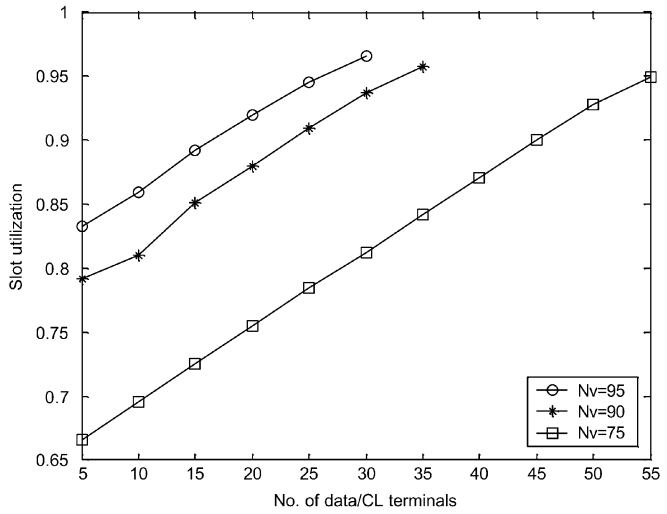


Fig. 14. Slot utilization when the number of data terminals varies.

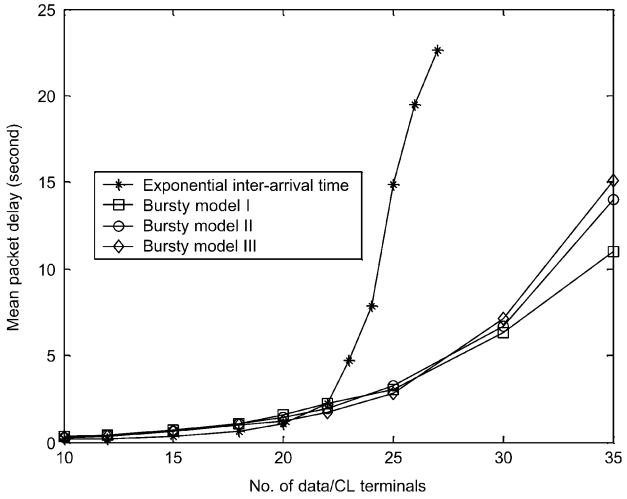


Fig. 15. Mean packet delay for different data parameters.

that this results in a slightly lower mean packet delay when the number of data terminals is small. However, the delay increases more dramatically when the number of data terminals is large. This is because data packets arrive more uniformly under the Poisson model, so in general a packet experiences a lower waiting time in the buffer. However, it also gives a larger number of contending data terminals, so the delay increases more significantly when the number of data terminals increases. Fig. 16 shows the slot utilization for different values of $\theta$ when the number of data terminals is changed. In general, the difference in slot utilization is quite small. It is found that the Poisson traffic gives a higher utilization. Fig. 17 shows the mean packet delay for different arrival rates during the active period while maintaining the same overall data rate. The figure shows that the delay is smaller when the arrival rate during the active period is larger. Note that as the overall data rate remains fixed, the packet size becomes smaller as the arrival rate is larger. Therefore, the mean packet delay becomes smaller. However, as shown in Fig. 18, the utilization is smaller when the arrival rate is larger. This is because in this case, a larger number of

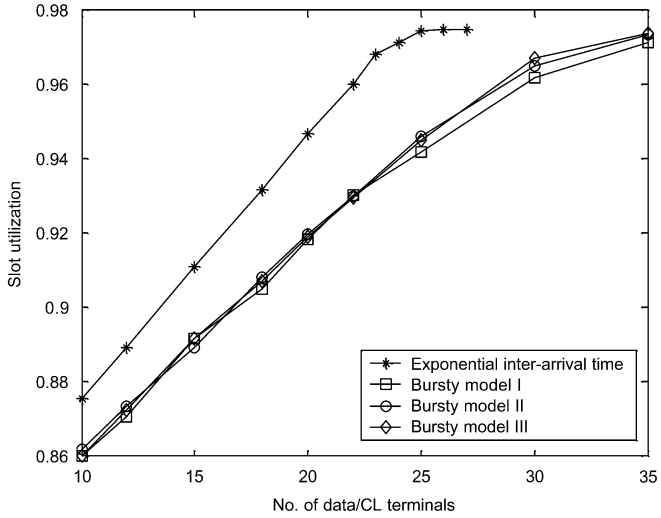


Fig. 16. Slot utilization for different data parameters.

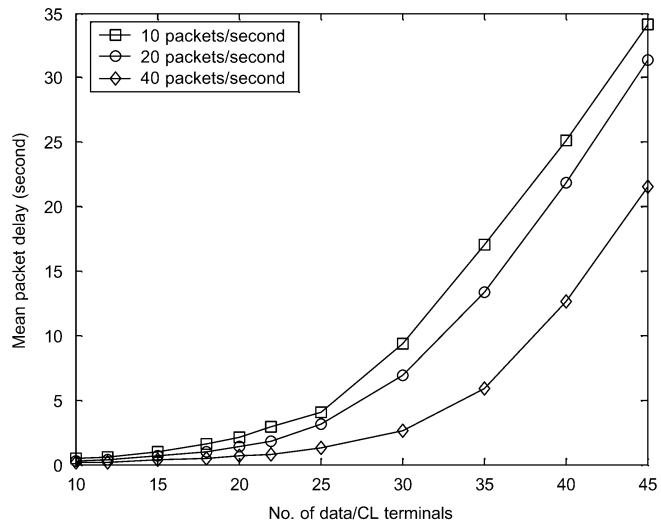


Fig. 17. Mean packet delay for different arrival rates during the active period.

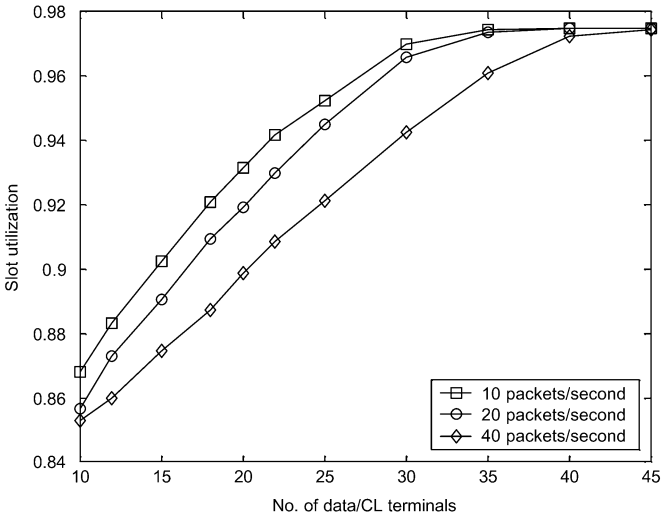


Fig. 18. Slot utilization for different arrival rates during the active period.

shorter packets arrive. The result is more contending data terminals and, hence, a lower utilization.

## V. Conclusion

In conclusion, we have presented a dynamic reservation protocol for the LEO mobile satellite system. To develop the protocol, we have formulated an analytical model to investigate how the access probability and number of reservation minislots should be varied under a delayed feedback environment in order to achieve the best performance. In particular, two approaches (frame-based and minislot-based) have been investigated and compared. Based on the analysis and taking into account the ease of implementation, it is proposed that the access probability should be varied in each frame based on the number of contending terminals. To implement the protocol, a novel contention-pattern-analysis approach has been proposed to estimate the number of contending terminals at the start of a frame. Based on the analytical model, a dynamic reservation protocol has been developed to integrate CO and CL traffic over a satellite channel. Analytical and simulation results show that the dynamic reservation protocol is effective in multiplexing CO and CL traffic.

## Appendix I
## Finding the Transition Probability

The transition probability is shown in (30). In the first and second cases, none of the $\sigma_{k,m}$ contending terminals accesses the minislot; thus, the number of contending terminals remains unchanged. In the first case, even the tried and successful terminals do not send a request, so there is no change in the state. In the second case, one of the $\tau_{k,m}$ terminals sends a request, but none of the $\gamma_{k,m}$ terminals accesses the minislot. Hence, we have $\tau_{k,m+1} = \tau_{k,m} - 1$ and $\gamma_{k,m+1} = \gamma_{k,m} + 1$. In the third and fourth cases, exactly one of the $\sigma_{k,m}$ contending terminals sends a request, so we have $\sigma_{k,m+1} = \sigma_{k,m} - 1$. In the third case, there is a successful access because none of the tried and successful terminals sends a request. As a result, we have $\tau_{k,m+1} = \tau_{k,m}$ and $\gamma_{k,m+1} = \gamma_{k,m} + 1$. In the fourth case, the access is not successful because at least one of the $\tau_{k,m} + \gamma_{k,m}$ terminals has sent a reservation request. Hence, we have $\tau_{k,m+1} = \tau_{k,m} + 1, \gamma_{k,m+1} = \gamma_{k,m}$. In the fifth case, two or more $\sigma_{k,m}$ terminals transmit a request, thus causing a collision. The probability that $i$ of the nontried terminals sends a request is $\binom{\sigma_{k,m}}{i} p_{k,m}^{i} (1 - p_{k,m})^{\sigma_{k,m} - i}$, where $2 \le i \le \sigma_{k,m}$. After sending a request, they become tried terminals. Hence, we have $\tau_{k,m+1} = \tau_{k,m} + i$ and $\gamma_{k,m+1} = \gamma_{k,m}$. See (30), at the bottom of the page.

## Appendix II
## Calculation of the Required Number of Reservation Minislots for the Frame-Based Scheme

Recall that at the start of frame $k$, there are $\sigma_{k,0}$ contending terminals and the access probabilities for the frame are $p_{k,m} = q_{k,m} = 1/\sigma_{k,0}$. Given that $n-1$ terminals have connected to the satellite, we denote $\chi_n$ as the probability that the $n$th success occurs (i.e., one of the $\sigma_{k,0} - n + 1$ nonsuccessful terminals accesses a minislot successfully). As both the successful and nonsuccessful terminals use the same access probability $1/\sigma_{k,0}$, and a new success occurs if one of the $\sigma_{k,0} - n + 1$ nonsuccessful terminals accesses a minislot successfully, we have

$$\begin{aligned}\chi_n &= \frac{\binom{\sigma_{k,0}-n+1}{1}}{\binom{\sigma_{k,0}}{1}} \times \left( \binom{\sigma_{k,0}}{1} \times \frac{1}{\sigma_{k,0}} \times \left(1 - \frac{1}{\sigma_{k,0}}\right)^{\sigma_{k,0}-1} \right) \\ &= (\sigma_{k,0} - n + 1) \times \frac{1}{\sigma_{k,0}} \times \left(1 - \frac{1}{\sigma_{k,0}}\right)^{\sigma_{k,0}-1}. \end{aligned} \tag{31}$$

$$P((\sigma_{k,m+1}, \tau_{k,m+1}, \gamma_{k,m+1}) \mid (\sigma_{k,m}, \tau_{k,m}, \gamma_{k,m}, p_{k,m}, q_{k,m}))$$

$$= \begin{cases} (1-p_{k,m})^{\sigma_{k,m}} \times (1 - B(\tau_{k,m}, 1, q_{k,m}) \times (1-q_{k,m})^{\gamma_{k,m}}), & \text{if } \begin{array}{l}\sigma_{k,m+1} = \sigma_{k,m} \\ \tau_{k,m+1} = \tau_{k,m} > 0 \\ \gamma_{k,m+1} = \gamma_{k,m}\end{array} \\ (1-p_{k,m})^{\sigma_{k,m}} \times B(\tau_{k,m}, 1, q_{k,m}) \times (1-q_{k,m})^{\gamma_{k,m}}, & \text{if } \begin{array}{l}\sigma_{k,m+1} = \sigma_{k,m} \\ \tau_{k,m+1} = \tau_{k,m} - 1 \\ \gamma_{k,m+1} = \gamma_{k,m} + 1 \\ \tau_{k,m} > 0\end{array} \\ B(\sigma_{k,m}, 1, p_{k,m}) \times (1 - (1-q_{k,m})^{\tau_{k,m}+\gamma_{k,m}}), & \text{if } \begin{array}{l}\sigma_{k,m+1} = \sigma_{k,m} - 1 \\ \tau_{k,m+1} = \tau_{k,m} + 1 \\ \gamma_{k,m+1} = \gamma_{k,m} \\ \sigma_{k,m} > 0\end{array} \\ B(\sigma_{k,m}, 1, p_{k,m}) \times (1-q_{k,m})^{\tau_{k,m}+\gamma_{k,m}}, & \text{if } \begin{array}{l}\sigma_{k,m+1} = \sigma_{k,m} - 1 \\ \tau_{k,m+1} = \tau_{k,m} \\ \gamma_{k,m+1} = \gamma_{k,m} + 1 \\ \sigma_{k,m} > 0\end{array} \\ \binom{\sigma_{k,m}}{i} \times p_{k,m}^{i} \times (1-p_{k,m})^{\sigma_{k,m}-i}, & \text{if } \begin{array}{l}\sigma_{k,m+1} = \sigma_{k,m} - i \\ \tau_{k,m+1} = \tau_{k,m} + i \\ \gamma_{k,m+1} = \gamma_{k,m} \\ 2 \le i \le \sigma_{k,m}\end{array} \\ 0, & \text{otherwise} \end{cases} \tag{30}$$

With $\chi_n$, we can compute the expected number of additional minislots to get the $n$th success. Its value is denoted as $\omega_n$. It is not difficult to see that

$$\omega_n = \frac{1}{\chi_n} = \frac{1}{(\sigma_{k,0} - n + 1) \times \frac{1}{\sigma_{k,0}} \times \left(1 - \frac{1}{\sigma_{k,0}}\right)^{\sigma_{k,0}-1}}. \tag{32}$$

Hence, the expected number of minislots for all of the $\sigma_{k,0}$ contending terminals to connect to the satellite is

$$\begin{aligned}\xi(\sigma_{k,0}) &= \sum_{n=1}^{\sigma_{k,0}} \omega_n \\ &= \frac{1}{\sigma_{k,0} \times \frac{1}{\sigma_{k,0}} \times \left(1 - \frac{1}{\sigma_{k,0}}\right)^{\sigma_{k,0}-1}} \\ &\quad + \frac{1}{(\sigma_{k,0} - 1) \times \frac{1}{\sigma_{k,0}} \times \left(1 - \frac{1}{\sigma_{k,0}}\right)^{\sigma_{k,0}-1}} \\ &\quad + \cdots + \frac{1}{1 \times \frac{1}{\sigma_{k,0}} \times \left(1 - \frac{1}{\sigma_{k,0}}\right)^{\sigma_{k,0}-1}} \\ &= \frac{1}{\frac{1}{\sigma_{k,0}} \times \left(1 - \frac{1}{\sigma_{k,0}}\right)^{\sigma_{k,0}-1}} \\ &\quad \times \left(1 + \frac{1}{2} + \cdots + \frac{1}{\sigma_{k,0} - 1} + \frac{1}{\sigma_{k,0}}\right). \end{aligned} \tag{33}$$

It can be found that (e.g., see [19])

$$\begin{aligned}1 + \frac{\log_2(\sigma_{k,0} - 1)}{2} &\le 1 + \frac{1}{2} + \cdots + \frac{1}{\sigma_{k,0} - 1} + \frac{1}{\sigma_{k,0}} \\ &\le \log_2 \sigma_{k,0}. \end{aligned} \tag{34}$$

Based on (34), the lower and upper bounds of $\xi(\sigma_{k,0})$ can be found as follows:

$$\frac{\sigma_{k,0} \times \left(1 + \frac{\log_2(\sigma_{k,0}-1)}{2}\right)}{\left(1 - \frac{1}{\sigma_{k,0}}\right)^{\sigma_{k,0}-1}} \le \xi(\sigma_{k,0}) \le \frac{\sigma_{k,0} \times \log_2 \sigma_{k,0}}{\left(1 - \frac{1}{\sigma_{k,0}}\right)^{\sigma_{k,0}-1}}. \tag{35}$$

If $\sigma_{k,0}$ is very large, $(1 - (1/\sigma_{k,0}))^{\sigma_{k,0}-1}$ converges to $e$. Hence, $\xi(\sigma_{k,0})$ is bounded as follows:

$$\begin{aligned}e \times \sigma_{k,0} \times \left(1 + \frac{\log_2(\sigma_{k,0} - 1)}{2}\right) &\le \xi(\sigma_{k,0}) \\ &\le e \times \sigma_{k,0} \times \log_2 \sigma_{k,0}. \end{aligned} \tag{36}$$

## REFERENCES


[1] R. E. Sheriff and Y. F. Hu, *Mobile Satellite Communication Networks*. New York: Wiley, 2001.

[2] I. F. Akyildiz and S.-H. Jeong, "Satellite ATM networks: A survey," *IEEE Commun. Mag.*, vol. 35, pp. 30–43, July 1997.

[3] C.-K. Toh and V. O. K. Li, "Satellite ATM network architectures: An overview," *IEEE Network*, vol. 12, pp. 61–71, Sept./Oct. 1998.

[4] E. D. Re and L. Pierucci, "Next-generation mobile satellite networks," *IEEE Commun. Mag.*, vol. 40, pp. 150–159, Sept. 2002.

[5] A. Iera and A. Molinaro, "Designing the interworking of terrestrial and satellite IP-based networks," *IEEE Commun. Mag.*, vol. 40, pp. 136–144, Feb. 2002.

[6] U. Black, *MPLS and Label Switching Networks*. Upper Saddle River, NJ: Prentice-Hall, 2002.

[7] 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.

[8] G. Benelli, R. Fantacci, G. Giambene, and C. Ortolani, "Performance analysis of a PRMA protocol suitable for voice and data transmissions in low earth orbit mobile satellite systems," *IEEE Trans. Wireless Commun.*, vol. 1, pp. 156–168, Jan. 2002.

[9] R. Fantacci, G. Giambene, and R. Angioloni, "A modified PRMA protocol for voice and data transmissions in low earth orbit mobile satellite systems," *IEEE Trans. Veh. Technol.*, vol. 49, pp. 1856–1876, Sept. 2000.

[10] E. D. Re, R. Fantacci, G. Giambene, and W. Sergio, "Performance analysis of an improved PRMA protocol for low earth orbit-mobile satellite systems," *IEEE Trans. Veh. Technol.*, vol. 48, pp. 985–1001, May 1999.

[11] G. Benelli, R. Fantacci, G. Giambene, and C. Ortolani, "Voice and data transmissions with a PRMA-like protocol in high propagation delay cellular systems," *IEEE Trans. Veh. Technol.*, vol. 49, pp. 2126–2147, Nov. 2000.

[12] D. Bertsekas and R. Gallager, *Data Networks*, 2nd ed. Upper Saddle River, NJ: Prentice-Hall, 1992.

[13] J.-F. Frigon, V. C. M. Leung, and H. C. B. Chan, "Dynamic reservation TDMA protocol for wireless ATM networks," *IEEE J. Select. Areas Commun.*, vol. 19, pp. 370–383, Feb. 2001.

[14] I. F. Akyildiz, E. Ekici, and M. D. Bender, "MLSR: A novel routing algorithm for multilayered satellite IP networks," *IEEE/ACM Trans. Networking*, vol. 10, pp. 411–424, June 2002.

[15] B. Maglaris, D. 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 1988.

[16] L. Lenzini, M. Luise, and R. Reggiannini, "CRDA: A collision resolution and dynamic allocation MAC protocol to integrate data and voice in wireless networks," *IEEE J. Select. Areas Commun.*, vol. 19, pp. 1153–1163, June 2001.

[17] X. Qiu, V. O. K. Li, and J.-H. Ju, "A multiple access scheme for multimedia traffic in wireless ATM," *Mobile Networks Applic.*, vol. 1, pp. 259–272, 1996.

[18] M. J. Karol, Z. Liu, and K. Y. Eng, "Distributed queueing request update multiple access (DQRUMA) for wireless packet (ATM) networks," in *Proc. ICC'95*, vol. 2, Seattle, WA, June 1995, pp. 1224–1231.

[19] T. H. Cormen, C. E. Leiserson, and R. L. Rivest, *Introduction to Algorithms*, 2nd ed. New York: McGraw-Hill, 2001.



**Henry C. B. Chan** (S'95–A'97–M'98) received the B.A. and M.A. degrees from the University of Cambridge, Cambridge, U.K., and the Ph.D. degree from the University of British Columbia (UBC), Vancouver, BC, Canada.

From October 1988 to October 1993, he worked with Hong Kong Telecommunications Limited, primarily on the development of networking services in Hong Kong. Between October 1997 and August 1998, he worked with BC TEL Advanced Communications, Canada, on the development of high-speed networking technologies and ATM-based services. Currently, he is an Assistant Professor in the Department of Computing, The Hong Kong Polytechnic University, Kowloon. He has authored/coauthored a textbook entitled *E-Commerce Fundamentals and Applications* (Chichester: Wiley, 2001), a chapter in the Internet Encyclopedia (Wiley), and over 40 journal/conference papers. His research interests include networking/communications, wireless networks, Internet technologies, and electronic commerce.

Dr. Chan is a Member of the Association for Computing Machinery (ACM). He is currently serving as an Executive Committee Member of the IEEE Hong Kong Section Computer Chapter. He has been listed in *Marquis Who's Who in the World*, 2002. During his graduate studies, he received a number of academic awards.

**Jie Zhang** was born in China, in 1981. She received the B.S. degree in computer science and technology from Nanjing University, Nanjing, China, in 2001. She is currently working toward the M.Phil. degree in computing from The Hong Kong Polytechnic University, Kowloon.

Her research interests include electronic commerce and wireless communications networks.

**Hui Chen** was born in China, in 1977. He received the B.S. degree in computer science and technology and the M.S. degree in engineering from Nanjing University, Nanjing, China, in 1999 and 2002, respectively.

He has worked as a Research Assistant in the Department of Computing, The Hong Kong Polytechnic University, Kowloon, for over two years. His research interests include wireless networks and electronic/mobile commerce.