State University of New York at Stony Brook
College of Engineering and Applied Sciences

Technical Report No. 759

Call Admission, Medium Access and Guaranteed Quality-of-Service Provisioning for ATM-Based Wireless Personal Communication Networks

by

Guodong Zhang and Stephen S. Rappaport

Department of Electrical and Computer Engineering
State University of New York
Stony Brook, New York 11794-2350

e-mail: gzhang@sbee.sunysb.edu, rappaport@sunysb.edu

Date: May 29, 1998

Call Admission, Medium Access and Guaranteed Quality-of-Service Provisioning for ATM-Based Wireless Personal Communication Networks

Guodong Zhang and Stephen S. Rappaport
Department of Electrical Engineering
State University of New York at Stony Brook
e-mail: gzhang@sbee.sunysb.edu, rappaport@sunysb.edu

Contents

Call Admission, Medium Access and Guaranteed Quality-of-Service Provisioning for ATM-Based Wireless Personal Communication Networks

Guodong Zhang and Stephen S. Rappaport

Department of Electrical Engineering

State University of New York at Stony Brook

e-mail: gzhang@sbee.sunysb.edu, rappaport@sunysb.edu

Abstract: An important issue in the design of next generation personal communication networks is the need to accommodate a variety of services each having predefined Quality-of-Service (QoS) requirements. A call admission control algorithm and a medium access protocol are proposed for guaranteed QoS provisioning in ATM-based wireless personal communication networks. A performance analysis model of the combined call admission control algorithm and the medium access protocol is developed. Example performance characteristics are presented.

I. INTRODUCTION

Recently there has been a considerable amount of research activity relating to ATM-based wireless personal communication networks (PCN). Rapid advances in VLSI and DSP technologies and the proliferation of personal mobile computing and communication devices are giving impetus to the evolution of next generation PCNs as wireless extensions of backbone B-ISDN/ATM communication networks. A variety of services: voice, video, data, multimedia applications, voice mail, database access, caller/calling ID and other fee-for-service options will be provided. If PCN is used to support broadband services for mobile users, the issue of meeting each service class' requirements for Quality-of-Service (QoS) emerges as a challenging problem.

The research reported in this paper was supported in part by the U.S. National Science Foundation under Grant No. NCR 94-15530 and in part by BMDO/IST under Grant No. N00014-9511217 administered by the U.S. Naval Research. General research support from Hughes Network Systems is gratefully acknowledged.

A solution lies in call admission and medium access control. It has been shown that QoS performance using call admission control is better than that of an uncontrolled system [9]. Also some research efforts in the field of medium access protocols for multimedia integration in wireless communication systems have been reported [5], [7].

We propose a call admission control algorithm and a medium access scheme for guaranteed QoS provisioning in next generation PCN's. Call admission policy is based on traffic class and is different for new call arrivals and hand-off arrivals. The medium access scheme that is proposed here is called request and report-dynamic slot allocation (R&R-DSA). It employs a time division format and efficiently utilizes the channel resource by taking the activity characteristics of each traffic class into consideration. Using the framework that has been developed in recent years, we analyze the performance of the system considering the combined effects of medium access and call admission by [1], [2], [4], [8].

The paper is organized as follows: In section II, the system description is proposed and traffic models are discussed. Section III puts forth a call admission control algorithm. In section IV, a multiple access scheme, which considers the activity characteristics of different traffic components, is proposed. The performance of this medium access is analyzed in section V. The Quality-of-Service of the combined system is analyzed in section VI. Finally, in section VII, conclusions are summarized.

II. SYSTEM DESCRIPTION AND TRAFFIC MODELS

A. System description

We consider a personal communication network with TDMA FDD structure and a spatially cellular layout based on a fixed channel assignment (FCA). We assume the communication system is homogeneous. Specifically, in each cell there are C frequency carriers. Each carrier is divided into TDMA frames with duration of F sec per frame.

The system will accommodate a variety of services. For convenient illustration, we consider three typical classes of service: voice, video, and delay-insensitive data. The problem is to accommodate these three traffic classes while assuring that their respective QoS requirements are met during session "lifetimes." We present a solution that uses call admission control and medium access control.

Call admission control regulates different traffic classes as shown in Fig. 1. In order to achieve guaranteed QoS, the characteristics of activity of each admitted traffic class should be thoroughly examined and taken into consideration. Admitted voice, video and data source traffic components are characterized in the following paragraphs.

B. Model of voice traffic

The activity of a voice call follows a pattern of alternating talkspurts and silence gaps [6]. Voice packets are generated during talkspurt periods. During silence gaps, a voice user has nothing to transmit. We assume the duration of a talkspurt is a negative exponentially distributed (ned) random variable with a mean of T_t and the duration of a silence gap is a ned random variable with a mean of T_s [6]. These mean durations are long compared with the duration of a frame. For convenience we assume that the duration of each frame, is F = 10 ms. Typically, values of T_t and T_s in [6], are T_t=1.0 sec and T_s=1.35 sec. In terms of current parameters, the mean duration of a talkspurt is 100 frames and the mean duration of a silence period is 135 frames. Given these considerations, it is clear that if an admitted voice call keeps a dedicated slot in each frame for the lifetime of the session, many payload slots will be wasted during the silence periods. So in the present scheme, we require an admitted voice call to contend for payload slot by making a reservation at the beginning of each talkspurt. If it succeeds, it gains the right to use the slot for the entire duration of that talkspurt. When a voice call enters a silence state, after one frame the base station will detect that the payload slot allocated to this call is empty. Then the base station will release this payload slot and it will become available for data packets or other voice users. When a voice call leaves a silence state and reenters a talkspurt state, it will again have to contend by making a reservation.

The activity of a single admitted voice call can be modeled as a three-state discrete time Markov chain. As shown in Fig. 2. there are three states: contention (CON), reserved talkspurt (RES) and silence gap (SIL). The *contention state* represents the condition that the voice call is at the beginning of a talkspurt period and is contending with other active calls for a reserved payload slot in its talkspurt period. The *reserved talkspurt state* represents the condition that the voice call has made a successful reservation and is using the reserved slot to transmit its packets. The transition probability from reserved talkspurt state to silence state, is denoted by α . This is the probability that a reserved talkspurt will end in a frame. Similarly, β denotes the transition probability from silence state to reserved contention state, this is, the probability that a silence gap will end in a frame. Consider a voice session which has been admitted. At the beginning of a talkspurt, a request is made using a random access scheme (which will be described subsequently). The probability that a given request results in a successful reservation is denoted by r_s. For given F, T_t and T_s, the values of α and β are given. The value of r_s will be calculated subsequently.

C. Model of video traffic

The video traffic is of variable bit-rate (VBR) type. The length of the video packet generated by a video user in a TDMA frame generally varies from frame to frame. We normalize the length of video packet to the duration of a payload slot. We let L be the maximum length of video packet (measured in payload slots) that a single video user can transmit in one TDMA frame. This is a system design parameter, which is chosen so that the frame will still have room to accommodate all admitted voice calls at their required QoS, even if all admitted video calls transmit packets of this length in a frame. If a call generates a video packet whose normalized length exceeds L, that packet must be fragmented into several parts, some of which must be transmitted in subsequent frames. Whether the video packet is transmitted in one frame as a whole packet or a fragment, is not important in the performance characterization that is developed in this paper. For simplicity, we assume that every video packet has a normalized length less than or equal to L. We also assume that an admitted video call never have a empty packet, that is for an admitted video call the normalized length of its packet at any frame is at least 1.

For our purpose of characterizing media access, it is sufficient to use a simple model to represent video traffic. A more representative but complex model of video characteristic is given in [7] (without attention to medium access). In our model, the actual length of a video packet can be any value from $\{1, 2, 3, \ldots, L\}$. Let L_{video} denote the normalized length of a video packet. We assume L_{video} is a discrete random variable distributed between 1 and L. Let $P_{L\_video}(i)$ denote the probability that L_{video} is i. The analysis in later sections is general in terms of $P_{L\_video}(i)$. It is assumed that this is the same for all admitted video calls, and that the event $L_{video} = i$ is independent from frame to frame.

Because a video call always has one packet to transmit in every frame (the actual length varies from frame to frame), every admitted video call is granted permission to transmit during the "lifetime" of its session. The number of payload slots it uses is dynamically allocated by the base station on a frame-by-frame basis according to the length of the video packet that is generated. Unused slots are made available for data traffic.

D. Model of data traffic

There are many different kinds of data, such as file transfer, e-mail, TCP/IP, fax, paging, etc. They have different characteristics that are not easily summarized in one general model. However, under pessimistic assumptions of heavy data traffic some simplification is reasonable. We assume that in each spatial cell at any time there is always a sufficient amount of offered data traffic so that, after call admission control, the admitted data traffic can be modeled as $U_d \times C$ data sources as shown in Fig. 1. That is, each of C carriers has U_d data sources, we do not consider the call admission and blocking probability and hand-off failure probability for data traffic in this paper. We assume the data traffic is of available bit-rate (ABR) type. All the admitted data traffic share the resources leftover by voice and video traffic fairly. The activity of an admitted data source is described as follows:

1. Using a round-robin approach, the base station schedules transmissions of all admitted data sources that have packets to send. If a data source has a packet to transmit, it has to report to the base station in the frame. If it is this data source's turn to transmit, the report is acknowledged by the base station before the end of the frame, and the packet is transmitted in the next frame.
2. Otherwise, the data source will report to the base station again in the next frame. The process will be repeated until it is the user's turn to transmit. During the reporting period, the data packet is stored in the user's buffer.
3. A data source will generate a new packet with a certain probability (denoted by h) in the current frame if its buffer will be clear of any previously generated packet at the end of current frame - that is, if it had no packet to transmit in the previous frame or it has a packet which

will be transmitted in this frame. After the new packet is generated, the data source will report to the base station as before.

Although this is a simple model, it reasonably characterizes some data traffic and provides insight into how data traffic utilizes system resources in the medium access scheme which will be discussed in section IV.

III. CALL ADMISSION CONTROL

Call admission control is based on the call's traffic class and is different for new calls and hand-off calls. Let class 1, 2, 3 denote the voice, video and delay-insensitive traffic respectively. The call admission control has the following functions:

1. Keep the number of voice calls sharing one carrier less than or equal to a given number specified by system design. Let U_s denote this given number. Cut-off priority is used for admission of voice calls as shown in Fig. 3. The number of carriers in a cell is C, and the number of payload slots that are reserved for hand-off voice calls in one cell is S_{h1}. A hand-off voice call will be admitted if there are less than $U_s \cdot C$ admitted voice calls, while a new voice call will be admitted only if the number of admitted voice calls is $U_s \cdot C - S_{h1}$ or less.
2. Keep the number of video calls sharing one carrier less than or equal to a given number specified by the system design. Let U_v denote this given number. Cut-off priority is used for admission of video calls as shown in Fig. 3. The number of payload slots reserved for hand-off video calls in one cell is $S_{h2} \cdot L$. A hand-off video call will be admitted if there are less than $U_v \cdot C$ admitted video calls, while a new video call will be admitted only if the number of admitted video calls is $U_v \cdot C - S_{h2}$ or less.
3. According to our assumption, there is always a sufficient amount of offered data traffic so that, U_d data sources are kept on each carrier at any time. We do not consider call admission for data traffic here.
4. A new call admission or a hand-off admission of a specific class is assigned to the carrier which has the least traffic load of that class of traffic. Ties are broken by random selection. This approach distributes the traffic load more or less uniformly among the carriers. The result is that all users of the same class will enjoy the same Quality-of-Service.

The flowchart of call admission function for new call and hand-off call is shown in Fig. 3.

IV. MEDIUM ACCESS

A medium access scheme inspired by multiservices dynamic reservation (MDR) TDMA [3] is proposed. In our scheme, there is also a request field used by voice calls. As discussed in Section II, every admitted video call is granted permission to transmit during its "lifetime" of the session, so there is no request slots for video calls. Each admitted video call is assigned a report slot to inform the base station of the number of slots that it will use in the next frame. In order to use the resources that are unused by voice and video traffic, each admitted data source is also assigned a report slot in which informs the base station if it has packet to transmit or not. We call this scheme request and report dynamic slot allocation (R&R-DSA).

Suppose in one TDMA frame there are N_s request slots in the request field used by voice calls, N_d report slots used by delay-insensitive data packets and N_v report slots used by video calls. The duration of a request slot is □. The duration of a report slot is also □. We assume that the duration of a payload slot is much larger than □. All these request slots and report slots are in the head of the frame. We also assume there are U_s voice calls, U_v $(\leq N_v)$ video calls and U_d $(\leq N_d)$ delay-insensitive data sources assigned to use this TDMA frequency carrier. Let S_p denote the total number of payload slots in each TDMA frame. For given S_p, the values of U_s and U_v are chosen so that, even if all the admitted voice and video calls have packets to transmit, S_p is large enough to accommodate them. The reason is that we assume that there is heavy offered data traffic in the system. The scenario without the heavy data traffic will be studied in future research. So there is no competition between voice packets and video packets. Both voice and video packets have higher priorities than data packets in allocation of payload slots. Only the slots which are not used by either voice or video packets can be used by data packets. The medium access scheme is described as follows:

1. When a mobile voice user enters the talkspurt period, it will send a request packet in a randomly selected request slot in the frame. Successful requests will be acknowledged by the base station by the end of the frame. Request packets which collide with other requests, will not be acknowledged by the base station and will time-out by the end of current TDMA frame.

An unsuccessful voice talkspurt request will be retransmitted in the next frame (in a randomly selected slot as before) until it either succeeds or exceeds the maximum delay constraint. If the maximum delay is exceeded for a voice talkspurt request, subsequent voice packets which were generated by the call and stored in the user's local buffer during the request period will be discarded. The next *new voice packet* will act as the beginning of the talkspurt period and will initiate contention for reservation as before.

2. Although video calls are granted permission to transmit upon admission, the first packet of a video session is not transmitted immediately. The video call sends a report packet in its report slot to inform the base station of the number of slots it will use in the next frame. Since each admitted video call is assigned a report slot to use, all report packets are successfully transmitted. In the next frame, the video packet will be transmitted and the length of the next video packet will be reported to the base station. The same operation is repeated for subsequent packets in the session.
3. A voice call which has made a successful reservation, is allocated one payload slot for its talkspurt in the next frame. Each video call is allocated a number of payload slots to use in the next frame. The number depends on the actual length of its packet. Any remaining capacity (slots) is available to data packets.
4. A data source which has a data packet to transmit, will use its own report slot in the frame to report to the base station that it has a packet to transmit. The permission to transmit is shifted among all active admitted data sources on a fair, round-robin basis. If it is this data source's turn to transmit, it will be acknowledged by the base station before the end of the frame, and the packet will be transmitted in the next frame. Otherwise, the data source will report to the base station again in the next frame, this is repeated until it is the given data source's turn to transmit. We assume that each data packet will need two payload slots to transmit. During the reporting period, the data packet is stored in the buffer.
5. The timing of the scheme is shown in Fig. 4, in which τ is the one way propagation delay. The time required to transmit the acknowledgment from the base station to mobiles is denoted by T_{ACK}. After receiving all the request and report packets, the base station processes them. The

processing time at the base station is denoted T_{p1} and similarly, T_{p2} denotes the processing

time

of an acknowledgment at a mobile. The system uses a TDMA FDD structure, so a mobile user can transmit and receive at the same time. To assure that even the last request/report packet

can

be processed before the end of the frame, we must have

$$(N_s + N_d + N_v)\cdot\sigma + 2\tau + T_{ACK} + T_{p1} + T_{p2} \le F \tag{1}$$

The structure of the R&R-DSA TDMA frame is shown in Fig. 5. The diagram of the medium access scheme is shown in Fig. 6.

V. PERFORMANCE ANALYSIS OF MEDIUM ACCESS

A. Probability of successful reservation

In order to study the performance of this medium access scheme, we must first determine the distribution of the number of voice calls that get permission to transmit. Suppose that n is the number of voice request slots in one frame. We denote the conditional probability that k calls succeed in making reservations given that there are u active calls as $P_s(k \mid u;\ n)$. A formula for this probability is developed in appendix A.

For each voice request, the probability of success given that there are u active calls and n request slots is given as

$$P_s(u;\ n) = \sum_{k=0}^{u} P\{a\ chosen\ request\ is\ among\ the\ k\ success\} \cdot P_s(k \mid u; n)$$

$$= \sum_{k=0}^{u} \frac{k}{u} \cdot P_s(k \mid u;\ n) \tag{2}$$

B. Performance analysis for voice and video traffic

The QoS metrics for voice and video traffic are average packet loss rate, average delay, blocking probability and forced termination probability. Blocking probability and forced termination probability are developed in Section VI. Here, we consider average packet loss rate and average delay.

Because data traffic uses only payload slots that are not used by voice and video traffic, in the performance analysis of voice and video traffic we do not need to consider the data traffic. We assume that the uplink and downlink channel transmissions are error free. Thus, packet loss comes only from failure to gain permission to transmit. Recall that the system is designed so that any admitted video call is granted permission to transmit during its "lifetime" and even if all the video calls generate packets of maximum length, they can still be accommodated. Thus, for video packets there is no packet loss and the delay is one TDMA frame. So we only consider the packet loss rate and delay for voice packets.

We want to determine the probability that a voice request packet makes a successful reservation. From (2), we can see that the probability depends on how many calls are making requests (including new requests and retransmitted requests) at the same time. Conclude that at the beginning of a TDMA frame, there are i ($0 \le i \le U_s$) retransmitted voice talkspurt request packets and j ($0 \le j \le (U_s - i)$) new talkspurt request packets generated from (U_s-i) voice calls which were not in contention state, i.e. in reserved talkspurt state or silence state, by the end of last frame. Let $W(j;u)$ denote the probability that j new talkspurt requests are generated from u voice calls which are in reserved talkspurt state or silence state. Let γ denote the probability that a single voice call generates a new talkspurt request packet when the voice call is in reserved talkspurt state or silence state. We find that

$$P(SIL \mid SIL \cup RES) = \frac{\alpha}{\alpha + \beta} \tag{3}$$

A voice call can generate a talkspurt request packet only when the voice call is in silence state. From Fig. 2. we can see γ is given by

$$\gamma = P(SIL \mid SIL \cup RES) \cdot \beta = \frac{\alpha \cdot \beta}{\alpha + \beta} \tag{4}$$

Then we have

$$W(j;u) = \binom{u}{j} (\gamma^{j} \cdot (1-\gamma)^{(u-j)}) \tag{5}$$

Using a discrete time Markov chain model in appendix B, we can determine the probability that there are i retransmitted voice talkspurt request packets, $R(i)$.

Then the probability that a voice talkspurt request makes a successful reservation is

$$r_s = \sum_{i=0}^{U_s} R(i) \cdot \sum_{j=0,\, i+j>0}^{U_s - i} W(j; (U_s - i)) \cdot P_s((i+j); N_s) \qquad (6)$$

Assume that the duration of each TDMA frame is F. Also assume that the maximum delay voice can tolerate is T. Let $\lfloor x \rfloor$ represent the largest integer that does not exceed x. Let M_s denote the maximum times that voice packet can be retransmitted respectively. We have

$$M_s = \lfloor T / F \rfloor - 1 \qquad (7)$$

For voice packets, the packet losses and delay only occur at the beginning of talkspurt periods. If the first packet has been retransmitted M_s times, this packet and all subsequent voice packets that were generated by the call and buffered in local buffer during the request period will be discarded and the next voice packet will act as the beginning of talkspurt to make reservation until the reservation is successfully made. Let s_{re} denote the probability that a request succeeds before being lost. Thus

$$s_{re} = \sum_{i=0}^{M_s} r_s \cdot (1 - r_s)^i$$

$$= 1 - (1 - r_s)^{M_s + 1} \qquad (8)$$

Let t denote the actual duration of talkspurt period. According to our assumption, t is a ned R.V. with a mean T_t. The distribution is

$$P(t) = e^{-\frac{1}{T_t} t} \quad , t > 0 \qquad (9)$$

Let N_p denote the number of voice packets during a talkspurt of duration t. N_p is given by

$$N_p = \lfloor t / F \rfloor + 1 \qquad (10)$$

Letting $m = \lfloor N_p / M_s \rfloor$, for a talkspurt of t sec, we have the packet loss rate $PL_s(t)$ given by

$$PL_s(t) = \frac{1}{N_p} \sum_{i=0}^{m} S_{re} \cdot (1 - S_{re})^i \cdot i \cdot M_s \qquad (11)$$

Then $$PL_s = \int_0^{\infty} P(t) \cdot PL_s(t) dt$$

$$= \int_0^{\infty} P(t) \left(\frac{1}{N_p} \sum_{i=0}^{m} s_{re} \cdot (1 - s_{re})^i \cdot i \cdot M_s \right) dt \qquad (12)$$

For $nF < t \le (n+1)F$, N_p is a constant and m is a constant. Thus, $\frac{1}{N_p}\sum_{i=0}^{m} s_{re}\cdot(1-s_{re})^i\cdot i\cdot M_s$ is a constant. Then

$$\int_{nF}^{(n+1)F} P(t)\,dt = \int_{nF}^{(n+1)F} e^{-\frac{1}{T_t}t}\,dt = T_t\cdot(e^{-\frac{nF}{T_s}} - e^{-\frac{(n+1)F}{T_t}}) \qquad (13)$$

Letting $m' = \lfloor (n+1)/M_s \rfloor$, we have

$$PL_s = \sum_{n=0}^{\infty} T_t\cdot(e^{-\frac{nF}{T_s}} - e^{-\frac{(n+1)F}{T_t}})\cdot\frac{1}{(n+1)}\sum_{i=0}^{m'} s_{re}\cdot(1-s_{re})^i\cdot i\cdot M_s \qquad (14)$$

Even if a talkspurt request succeeds without retransmission, there is one frame delay. If the request succeeds in the i-th$(i \le M_s)$ retransmission, the delay is 1+i frames. According to the medium access scheme, previous failed requests, if any, have no influence on the delay. The average delay $\overline{D}_s$ is given by

$$\overline{D}_s = 1+\sum_{i=1}^{M_s} r_s\cdot(1-r_s)^i\cdot i \ \text{(Frames)} \qquad (15)$$

We consider the scenario in which T_t=1.0 sec, T_s=1.35 sec, α =0.00995, β =0.00738, $N_s = 5$ and there are enough payload slots to accommodate the admitted voice calls within the ranges of figures 7 and 8. In the figures 7 and 8, average delay and packet loss rate are respectively shown as a function of U_s. The average delay for data packets is only a little longer than one TDMA frame, and packet loss rate are below 1×10^{-11}. The one frame delay is caused by the request at the beginning of each talkspurt. The rest of the delay is caused by the retransmission of the request until reservation is made, which is very small. We can see that in this medium access scheme, a few request slots (whose duration σ is much smaller than that of one payload slot τ) can be used to accommodate many voice calls while maintaining a reasonably small average delay and a very small packet loss rate. This also means, the overheads of this scheme, bandwidth for request slots and delay, is reasonably small.

C. Performance analysis for data packet

All the admitted data sources share the resources unused by voice and video traffic fairly. A data source that has a packet to transmit will report to the base station and wait until its turn to transmit. All these buffers together form a distributed queue. Data packets get payload slots to use in an order that is determined according to a fair, round-robin policy. The average delay is the primary QoS metric for data traffic.

According to the data traffic model and the medium access scheme, at any given time there are at most U_d reports. This queuing problem corresponds to a G/G/S/U_d queuing system with server vacations. S is the number of servers, i.e. the number of pairs of payload slots available to data packets, which is a random variable. Because only the payload slots that are unused by voice and video traffic are available for data traffic, so there is server vacation. Here, we model this problem by a discrete time Markov chain shown in Fig. 9. The state is the number of data reports which are not acknowledged by the end of the current frame. We want to determine the state probability p_i. A formula for this probability is developed in Appendix C.

With the set of p_i's, we can derive all the performance metrics now. The average length of the data request queue is given by

$$\overline{L}=\sum_{i=1}^{U_d} i\cdot p_i \tag{16}$$

Let $u(i)$ denote the throughput of the state i. $u(i)$ is given by

$$u(i)=\sum_{m=1}^{\max(U_d,\lfloor S_p/2\rfloor)} s_m\cdot m\cdot \sum_{n=\max(0,\ (m-i))}^{U_d-i} q_d(n;\ (U-i)) \tag{17}$$

The average throughput of the queue is calculated as

$$\overline{Throughput}=\sum_{i=0}^{U_d} u(i)\cdot p_i \tag{18}$$

Using Little's law [10], the average delay in the queue is given by

$$\overline{D}_{queue}=\frac{\overline{L}}{\overline{Throughput}} \tag{19}$$

The average total delay of data packet is given by

$$\overline{D}_{d\_total}=\overline{D}_{queue}+1 \tag{20}$$

The utilization of the channel resources by data traffic is given by

$$Util_d=2\cdot\overline{Throughput}\ /\ S_p \tag{21}$$

For the simplicity of illustration, we consider the scenario of U_s=4, U_v=3, N_s =6, N_v=3, L=4, N_d=10 and S_p=16. We assume that $P_{L\_video}(1) = 1/8$, $P_{L\_video}(2) = 5/24$, $P_{L\_video}(3) = 7/24$ and $P_{L\_video}(4) = 9/24$. The parameter choice for this scenario is taken only to demonstrate the performance of the medium access scheme. When there is no voice and video traffic in the system, we assume there are 10 data sources with a packet generation probability h on each carrier. The average delay and the utilization of channel are plotted as a function of h in Figures 10 and 11 respectively. We can see that the average delay is small and for heavy data sources (h is near 1) an utilization near 1 can be achieved.

Now, suppose there are 4 voice calls, 3 video calls and 10 data sources assigned to one carrier. The average total delay for data packet is plotted as a function of h in Fig. 12. Compared with Fig. 10, the delay increases 3~4 times because with the presence of voice and video traffic, data traffic can only use the payload slots that are unused by voice and video traffic. The average total delay for data packet is still reasonably small, say, less than 4 frames. We can see that using this medium access the system can accommodate plenty of voice and video traffic while at the same time accommodating heavy data traffic with their QoS requirement met.

VI. QUALITY of SERVICE ANALYSIS of THE SYSTEM

Using the proposed medium access protocol and call admission control we seek to admit a call only if the system can guarantee the required Quality-of-Service for the call during the lifetime of its session.

A. Example problem statement

We assume that there are G type mobile platforms, labeled by g=1, 2, …,G, and there are C frequency carriers in each cell. Assume that U_i is the maximum number of each class (voice/video) users that can share a carrier while at the same time the requirement of QoS, such as transmission delay and packet loss rate, is maintained. U_i is determined in medium access performance analysis developed in Section VI. No more than one call can be supported by a platform at any given time. There are I classes of calls, labeled by i=1, 2, …, I. The new i-class call origination rate from a noncommunicating g-type platform is $\Lambda(g,i)$. The number of noncommunicating g-type platforms in any cell is denoted $v(g,0)$. Thus, the total *i*-class call generation rate for g-type platforms in any cell can be denoted $\Lambda_n(g,i)=\Lambda(g,i)\times v(g,0)$. A cut-off priority scheme is used as specified in call admission control in Section III. New *i*-class calls will be blocked if the number of admitted *i*-class calls is $U_i \cdot C - S_{hi}$ or greater. Hand-off *i*-class calls will fail if the number of admitted *i*-class calls is $U_i \cdot C$.

For any call of class i, we assume the unencumbered call (session) duration is a ned random variable, $T(i)$, having a mean $\overline{T}(i)=1/\mu(i)$. The dwell time in a cell for a g-type platform is a ned random variable, $T_D(g)$, has a mean $\overline{T}(g)=1/\mu_D(g)$.

B. State description

For convenience, we assume there are 2 types of platform, i.e. *G*=2, and only one type of resource, channel. We also assume there are C TDMA carriers in a cell. Denote voice service by class 1, video service by class 2, i.e. *I*=2.

Consider a single cell. We define the state (of a cell) by a sequence of nonnegative integers. This can be conveniently written as *G* *n*-tuples [2], [4]

$$\begin{matrix} v_{11}, & v_{12} \\ v_{21}, & v_{22} \end{matrix} \tag{22}$$

where v_{gi} {*g*=1, 2, …, *G*; *i*=1, 2, …, *I*} is the number of platforms of type *g* that have an *i*-class call in progress. Here, *G*=2, *I* =2. It is convenient to order the states using an index *s*=0, 1, 2, …, $s_{\max}$. Then the state variables v_{gi} can be shown explicitly dependent on the state. That is, v_{gi}=*v*(s,g,i).

When the cell is in state s, the following characteristics can be determined. The number of *i*-class calls is

$$v(s,i)=\sum_{g=1}^{G} v(s,g,i) \ , \qquad i=1, 2. \tag{23}$$

Permissible states correspond to those sequences for which all constraint are met.

C. Driving processes and state transition flow:

1. Generation of *i*-class new calls

A transition into state *s*, due to a new *i*-class call arrival on a *g*-type platform when the cell is in the state x_{ni}, will cause the state variable $v(x_{ni}, g, i)$ to be increased by 1. A permissible state x_{ni} is a predecessor state of *s* for an *i*-class new call arrival on a *g*-type platform, if $v(s, i)$ $\leq U_i \cdot C - S_{hi}$, S_{h1} is the number of payload slots reserved for hand-off voice calls, $S_{h2} \cdot L$ is the number of payload slots reserved for hand-off video calls, and the state variables are related by

$$
\begin{aligned}
&v(x_{ni}, g, i) = v(s, g, i) - 1 \\
&v(x_{ni}, z_1, z_2) = v(s, z_1, z_2), \quad z_1 \neq g \\
&v(x_{ni}, z_1, z_2) = v(s, z_1, z_2), \quad z_2 \neq i
\end{aligned}
\tag{24}
$$

Otherwise it will be blocked.

Let $\Lambda_n(g, i)$ denote the average arrival rate per cell of *i*-class new calls from *g*-type platforms. Then, the corresponding transition flow is given by

$$\gamma_c(s, x_{ci}) = \Lambda_n(g, i). \tag{25}$$

2. Call completions of *i*-class

A transition into state *s*, due to a *i*-class call completion on a *g*-type platform when the cell is in the state x_{ci}, will cause the state variable $v(x_{ci}, g, i)$ to be decreased by 1. Thus a permissible state x_{ci} is a predecessor state of s for an *i*-class call completion on a *g*-type platform, if the state variables are related by

$$
\begin{aligned}
&v(x_{ci}, g, i) = v(s, g, i) + 1 \\
&v(x_{ci}, z_1, z_2) = v(s, z_1, z_2), \quad z_1 \neq g \\
&v(x_{ci}, z_1, z_2) = v(s, z_1, z_2), \quad z_2 \neq i
\end{aligned}
\tag{26}
$$

The corresponding transition flow is given by

$$\gamma_c(s, x_{ci}) = \mu(i) \times v(x_{ci}, g) \tag{27}$$

3. Hand-off arrivals of *i*-class

Let $\Lambda_h(i)$ be the average rate at which hand-off arrivals of i-class service impinge on the cell, and F_g denote the fraction of arrivals that are g-type platforms. A transition into state *s*, due to an i-class hand-off arrival on a *g*-type platform when the cell is in the state x_{hi}, will cause the state variable $v(x_{hi},g,i)$ to be incremented by 1. Thus a permissible state x_{hi} is a predecessor state of s for an i-class hand-off arrival on a g-type platform, if $v(s,i) \leq U_i \cdot C$ and the state variables are related by

$$v(x_{hi},g,i)=v(s,g,i)-1$$
$$v(x_{hi},z_1,z_2)=v(s,z_1,z_2), \qquad z_1 \neq g$$
$$v(x_{hi},z_1,z_2)=v(s,z_1,z_2), \qquad z_2 \neq i \tag{28}$$

The corresponding transition flow is given by

$$\gamma_h(s,x_{hi})=\Lambda_h(i)\times F_g \tag{29}$$

4. Hand-off departures of i-class

A transition into state s, due to a hand-off departure of i-class on a g-type platform when the cell is in the state x_{di}, will cause the state variable $v(x_{di},g,i)$ to be decreased by 1. Thus a permissible state x_{di} is a predecessor state of s for an i-class hand-off departure on a g-type platform, if the state variables are related by

$$v(x_{di},g,i)=v(s,g,i)+1$$
$$v(x_{di},z_1,z_2)=v(s,z_1,z_2), \qquad z_1 \neq g$$
$$v(x_{di},z_1,z_2)=v(s,z_1,z_2), \qquad z_2 \neq i \tag{30}$$

The corresponding transition flow is given by

$$\gamma_d(s,x_{di})=\mu_D(g)\times v(x_{di},g,i) \tag{31}$$

D. Flow balance equations

From the equations given above, the total transition flow into state s from any permissible state x can be expressed by

$$q(s,x)=\gamma_n(s,x)+\gamma_c(s,x)+\gamma_h(s,x)+\gamma_d(s,x) \tag{32}$$

in which $s \neq x$, and flow into a state s has been taken as a positive quantity. The total flow out of state s is denoted $q(s,s)$, and is given by

$$q(s,s)=-\sum_{k=0,\,k\neq s}^{s_{max}} q(k,s) \tag{33}$$

To find the statistical equilibrium state probabilities for a cell, we write the flow balance equations for the states. These are a set of $s_{\max}+1$ simultaneous equations for the unknown state probabilities $p(s)$. They are of the form

$$\sum_{j=0}^{s_{\max}} q(i,j)p(j)=0, \qquad i=0,\ 1,\ 2,\ \ldots,\ s_{\max}-1$$

$$\sum_{j=0}^{s_{\max}} p(j)=1 \tag{34}$$

in which, for $i \neq j$, $q(i,j)$ represents the net transition flow into state i from state j, and $q(i,i)$ is the total transition flow out of state i. These equations express that in statistical equilibrium, the net probability flow into any state is zero and the sum of the probabilities is unity. The index j in equation (34) can run up to $s_{\max}$ to provide a redundant set that may be helpful in numerical computation.

E. Quality-of-Service metrics

When the statistical equilibrium state probabilities and transition flows are found, the following QoS metrics can be calculated.

1. Blocking probability

The blocking probability for a call from a g-type platform is the average fraction of new g-type calls that are denied access to a channel. Blocking of new i-class call occurs if there are no channels to serve the call. We define the following set of states

$$B_i=\{s: U_i \cdot C - S_{hi} \leq v(s,i) \leq U_i \cdot C\} \tag{35}$$

Then the blocking probability for i-class calls is

$$P_B(i)=\sum_{s\in B_i} p(s) \tag{36}$$

2. Hand-off failure probability

The hand-off failure probability for i-class calls is the average fraction of i-class hand-off attempts that are denied access to a channel. Hand-off attempts have potential access to all channels of a cell without regard to S_{hi}. We define the following set of states

$$H_i = \{s : v(s,i) = U_i \cdot C\} \tag{37}$$

Then the hand-off probability for a *i*-th class hand-off is

$$P_H(i) = \sum_{s \in H_i} p(s) \tag{38}$$

3. Forced termination probability

There is a more important QoS metric than hand-off failure probability. It is forced termination probability, $P_{FT}(g,i)$. This is defined as the probability that an *i*-class that is not blocked is interrupted due to hand-off failure in its lifetime. For convenience, we limit our discussion here to the case where all dwell-time phases of a given platform type are statistically identical.

If we let $a(g,i)$ denote the probability that an *i*-class call on a *g*-type platform will make a hand-off attempt and will fail on that attempt. Similarly, let $b(g,i)$ denote the probability that an *i*-class call on a g-type platform will make a hand-off attempt and succeed. Using the Markovian properties of the model we have

$$a(g,i) = \mu_D(g) \cdot P_H(g,i)/(\mu(i) + \mu_D(g)) \tag{39}$$

and

$$b(g,i) = \mu_D(g) \cdot (1 - P_H(g,i))/(\mu(i) + \mu_D(g)) \tag{40}$$

Assuming hand-offs independence we get

$$P_{FT}(g,i) = \sum_{i=0}^{\infty} a(g,i) \cdot (b(g,i))^i \tag{41}$$

Summing (41) and using (39), (40) we get

$$P_{FT}(g,i) = \mu_D(g) \cdot P_H(g,i)/(\mu(i) + \mu_D(g) \cdot P_H(g,i)) \tag{42}$$

F. Results

This analysis approach is used to consider a system with number of payload slots S_p=16, number of voice request slots N_s=6, N_v=3, N_d=10. In the example, new call arrival rates for all platform types are assumed to be the same. The QoS requirements of voice, video and data traffic are shown in Table 1.

	QoS metrics				
traffic class	average delay	maximum delay	average packet loss rate	blocking probability	forced termination probability
voice traffic	N/A	4 TDMA frames	$\leq 10^{-5}$	$\leq 10^{-2}$	$\leq 10^{-3}$
video traffic	N/A	4 TDMA frames	$\leq 10^{-5}$	$\leq 2\times 10^{-2}$	$\leq 2\times 10^{-3}$
data traffic	$\leq$4 TDMA frames	infinity	$\leq 10^{-10}$	N/A	N/A

Table. 1. The QoS requirements of voice, video and data traffic.

By using the equations developed in Section VI, we find in order to meet the QoS requirements of traffic, U_s is selected as 4, U_v is selected as 3 and U_d is selected as 10. The parameter choices for the figures are taken only to demonstrate the performance of the combined medium access and call admission.

Fig. 13 shows the blocking probabilities for voice calls which is plotted as a function of call origination rate. Fig. 14 shows the forced termination probabilities for voice calls which is plotted as a function of call origination rate on a similar plot. There is a tradeoff between blocking probability and forced termination probability. The increase in blocking probability as S_{h1} is increased to favor the hand-offs can be observed. If the voice call origination rate is less than 5×10^{-4} calls/sec and S_{h1} is chosen as 2, the voice call's QoS requirement on both blocking and forced termination probabilities can be met. Its QoS requirements on average delay and packet loss rate are met by using the R&R-DSA medium access scheme.

Fig. 15 shows the blocking probabilities for video calls. Fig 16 shows the forced termination probabilities for video calls. Similarly, the tradeoff between blocking probability and forced termination probability can be seen. If the video call origination rate is less than 3.75×10^{-4} calls/s and S_{h2} is chosen as 1, the video call's QoS requirement on both blocking and forced termination probabilities can be met. Its QoS requirements on average delay and packet loss rate are met by using the medium access scheme.

VII. CONCLUSION

We propose and consider a medium access scheme called R&R-DSA and a call admission control algorithm whose combined use allow guaranteed QoS provisioning for ATM-based wireless personal communication networks. The performance of R&R-DSA is analyzed in terms of average delay and packet loss rate. We developed a performance analysis model to characterize call blocking probability, hand-off failure and forced termination probabilities of a system using the combined call admission and medium access. As we have shown in performance analysis, the R&R-DSA can accommodate a variety of traffic classes and provide efficient use of bandwidth. Combined with call admission control, it can guarantee QoS for different services in ATM-based wireless personal communication networks. The performance of an example system with combined call admission and medium access was determined under the pessimistic assumption of heavy data traffic.

REFERENCES

[1] D. Hong and S. S. Rappaport, "Traffic Model and Performance Analysis for Cellular Mobile Radio Telephone Systems with Prioritized and Non-Prioritized Hand-off Procedures," IEEE Trans. Veh. Technol., Aug. 1986, Vol. 35, No. 3, pp. 77-92.

[2] S. S. Rappaport, "Blocking Hand-off and Traffic Performance Analysis for Cellular Communication System with Mixed Platforms," IEEProceedings, Part I, Communications, Speech and Vision, October 1993, Vol. 40, No. 5, pp. 389-401.

[3] D. Raychaudhuri and N. D. Wilson, " ATM-based transport architecture for multiservices wireless personal commuincation network," *IEEE JSAC*, Vol. 12, No. 8, Oct. 1994.

[4] S. S. Rappaport and C. Purzynski, "Prioritized resource assignment for mobile cellular systems with mixed services and platform type," IEEE Trans. Veh. Technol., Vol. 45, No. 3, Aug. 1996.

[5] Yibin Yang, Junfeng He, Ming T. Liu, "A medium access protocol for vocie and data integration in received-oriented DS-CDMA PCNs," *ICPUC* 1997.

[6] S. Nanda and D. J. Goodman, " Performance of PRMA: A Packet Voice Protocol for Cellular Systems," IEEE Trans. Veh. Technol., Vol. 40, No. 3. Aug. 1991.

[7] G. Anastasi, D. Grillo, and L. Lenzini, "An Access Protocol for Speech/Data/Video Integration in TDMA-Based Advanced Mobile System," IEEE JSAC., Vol. 15., No. 8., Oct.

1997.

[8] Y. Park and S. S. Rappaport, "Cellular Communication Systems with Voice and Background Data," in D. J. Goodman and D. Raychaudhuri (Eds.) 'Mobile Multimedia Communications', (Plenum Press, New York ,1997), pp. 33-42.

[9] A. S. Acampora and M. Naghshineh, "Control and Quality-of-Service Provisioning in High-Speed Microcellular Network," IEEE Personal Communications Mag., Vol. 1, No. 2, Second Quarter, 1994.

[10] R. Wolff, Stochastic Modeling and the Theory of Queues, Englewood Cliffs, NJ: Prentice-Hall, 1989.

Appendix A: Probability of Successful Reservation

Let n denote the number of request slots for a specific class in one TDMA frame. Let u denote the number of voice calls who send requests in this TDMA frame. Each call will randomly choose a request slot from n slots and send a request packet in that slot. Let S denote the sample space. Let E be any event on S. Let $N[E]$ denote the size of E. The size of S, $N[S]$, is n^u.

Exactly k requests succeed if there are k request slots each occupied by only one request packet and any other slot is either empty (i.e. not chosen by any call) or occupied by more than one request packet (i.e. there is collision of two or more request packets). Let $P_s(k \mid u; n)$ denote the probability that k calls succeed in making reservation given that there are u active calls and n request slots. For any given n and u, there are two possibilities: $n \geq u$ or $n<u$.

For $n \geq u$, the probability can be solved as follows:

The event $k= u$ corresponds to a event E in which each active user chooses a distinct slot from n slots. This results in one request packet placed in each selected slot. In order to implement this, first we select u slots from n slots, then place one packet in each selected slot. So the corresponding probability is given by

$$P_s(k \mid u; n) = \frac{N[E]}{N[S]} = \frac{u! \cdot \binom{n}{u}}{n^u} \quad , \qquad k=u \tag{43}$$

The event $k= u$-1 is impossible because if a collision occurs at least two requests fail. So

$$P_s(k \mid u; n) = 0 \quad , \qquad k=0 \tag{44}$$

The event of any k with $1 \leq k \leq u\text{-}2$ can be implemented by the following two steps: step 1: select m slots from n slots, select k packets from u packets and k slots from these m slots, then place one packet in each of these k selected slots; step 2: arrange the remaining (u-k) packets in such a way that each of the remaining n-m slots is occupied by at least two packets.

There is some requirement on m. First, (n-m) should be greater than or equal to 1. Secondly, (u-k) should be greater than or equal to 2(n-m). Finally, m should be larger than k. Then we have

$$\begin{cases} n-m \geq 1 \\ u-k \geq 2(n-m) \\ m \geq k \end{cases} \tag{45}$$

Note that $n\text{-}\left\lfloor \frac{u-k}{2} \right\rfloor \geq k$, then the range of m is

$$n - \left\lfloor \frac{u-k}{2} \right\rfloor \leq m \leq n-1 \tag{46}$$

Let $N[step1]$ and $N[step2]$ denote the number of possible combinations in step1 and step 2 respectively. We find that

$$N[step1] = \binom{n}{m} \cdot \binom{u}{k} \cdot \binom{m}{k} \cdot k! \tag{47}$$

In order to implement step 2, first we place 2 packets in each of the (n-m) slots. The number of possible combinations in this sub-step is $\frac{[2(m-n)]!}{(2!)^{(n-m)}}$. This sub-step makes sure that each of the (n-m) slots has at least 2 packets. Then, we place the remaining (u-k)-2(n-m) packets on (n-m) slots randomly. The number of possible combinations in this sub-step is $(n-m)^{((u-k)-2(n-m))}$. So we have

$$N[step2] = \frac{[2(m-n)]!}{(2!)^{(n-m)}} (n-m)^{((u-k)-2(n-m))} \tag{48}$$

Because step 1 and step 2 are independent, the size of E is given by

$$N[E]=\sum_{m=n-\left\lceil\frac{u-k}{2}\right\rceil}^{n-1} N[step1]\cdot N[step2]$$

$$=\sum_{m=n-\left\lceil\frac{u-k}{2}\right\rceil}^{n-1}\binom{n}{m}\binom{u}{k}\binom{m}{k}k!\frac{[2(m-n)]!}{(2!)^{(n-m)}}(n-m)^{((u-k)-2(n-m))} \tag{49}$$

The corresponding probability is

$$P_s(k\mid u;n)=\frac{N[E]}{N[S]}$$

$$=\frac{1}{n^u}\sum_{m=n-\left\lceil\frac{u-k}{2}\right\rceil}^{n-1}\binom{n}{m}\binom{u}{k}\binom{m}{k}k!\frac{[2(m-n)]!}{(2!)^{(n-m)}}(n-m)^{((u-k)-2(n-m))},\quad 1\le k\le u-2 \tag{50}$$

The outcome $k=0$ corresponds to the event E in which all u calls fail. We select m slots out of the n slots and put u packets in these m slots and let the remaining $(n-m)$ slots empty. The u packets are arranged in such a way: first use 2 packets on each slot, then place the remaining $(u-2m)$ packets randomly in these m slots. Similar to previous case, the range of m is

$$1\le m\le\left\lfloor\frac{u}{2}\right\rfloor \tag{51}$$

Using the same method as above, we have the size of E

$$N[E]=\sum_{m=1}^{\lfloor u/2\rfloor}\frac{(2m)!}{(2!)^m}m^{(u-2m)} \tag{52}$$

The corresponding probability is

$$P_s(k\mid u;n)=\frac{N[E]}{N[S]}=\frac{1}{n^u}\sum_{m=1}^{\lfloor u/2\rfloor}\frac{(2m)!}{(2!)^m}m^{(n-m)}\quad,\qquad k=0 \tag{53}$$

The distribution turns out to be:

$$P_s(k \mid u;n) = \begin{cases} \dfrac{1}{n^u} \displaystyle\sum_{m=n-\left\lceil \frac{u-k}{2} \right\rceil}^{n-1} \binom{n}{m}\binom{u}{k}\binom{m}{k} k! \frac{[2(m-n)]!}{(2!)^{(n-m)}} (n-m)^{((u-k)-2(n-m))} & for\ 1 \le k \le u-2 \\ 0 & for\ k = u-1 \\ \dfrac{u!\binom{n}{u}}{n^u} & for\ k = u \\ \dfrac{1}{n^u} \displaystyle\sum_{m=1}^{\left\lceil \frac{u}{2} \right\rceil} \frac{(2m)!}{(2!)^m} m^{(u-2m)} & for\ k = 0 \end{cases} \tag{54}$$

If $u>n$, then the event, $k \ge n$ is impossible. For outcome $0<k<n$ and $k=1$, we use the same reasoning as above, but the ranges of m are different from that given above. The distribution is

$$P_s(k \mid u;n) = \begin{cases} \dfrac{1}{n^u} \displaystyle\sum_{m=1}^{\min(n,\lfloor u/2 \rfloor)} \frac{(2m)!}{(2!)^{(n-m)}} \cdot m^{(u-2m)} & for\ k = 0 \\ 0 & for\ n \le k \le u \\ \dfrac{1}{n^u} \displaystyle\sum_{m=(k,n-\left\lfloor \frac{u-k}{2} \right\rfloor)}^{n-1} \binom{n}{m} \cdot \binom{u}{k} \cdot \binom{m}{k} \cdot k! \cdot \frac{[2(m-n)]!}{(2!)^{(n-m)}} \cdot (n-m)^{((u-k)-2(n-m))} & for\ 1 \le k \le n-1 \end{cases} \tag{55}$$

Appendix B: The state probability $R(i)$

This can be modeled as a discrete time Markov chain. The state is the number of retransmitted talkspurt requests in the beginning of the current TDMA frame, i.e. the number of unsuccessful requests at the end of last TDMA frame. The transition probabilities are given as for $i \le j$ and $i, j \ne 1$,

$$p_{i,j} = \sum_{k=j-i}^{U_s-i} P_s(j \mid (j+k); N_s) \cdot W(k;(U_s - i)) \tag{56}$$

for $i>j$ and $i, j \ne 1$,

$$p_{i,j} = \sum_{k=0}^{U_s-i} P_s(j \mid (i+k); N_s) \cdot W(k;(U_s - i)) \tag{57}$$

$$p_{i,j} = 1, \qquad i=j=1 \tag{58}$$

$$p_{i,j} = 0, \qquad i=1 \text{ or } j=1 \text{ and } i \ne j \tag{59}$$

$P_s(j \mid (i+k); N_s)$ is determined in Appendix A and $W(k;(U_s - i))$ is obtained by using (5). For steady state probabilities, we have the following equations

$$\begin{bmatrix} R(0) \\ R(1) \\ R(2) \\ \vdots \\ R(U_s) \end{bmatrix} = \begin{bmatrix} p_{0,0} & p_{1,0} & \cdots & p_{U_s,0} & p_{U_s,0} \\ p_{0,1} & p_{1,1} & \cdots & p_{U_s-1,1} & p_{U_s,1} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ p_{0,U_s-1} & p_{1,U_s-1} & \cdots & p_{U_s-1,U_s-1} & p_{U_s,U_s-1} \\ p_{0,U_s} & p_{1,U_s} & \cdots & p_{U_s-1,U_s} & p_{U_s,U_s} \end{bmatrix} \begin{bmatrix} R(0) \\ R(1) \\ R(2) \\ \vdots \\ R(U_s) \end{bmatrix} \tag{60}$$

$$\sum_{i=0}^{U_s} R(i) = 1 \tag{61}$$

We solve these equations using Gauss-Seidel iteration algorithm to find $R(i)$.

Appendix C: The State Probability, p_i

In order to consider the influence of voice traffic on data traffic, we determine *the state probability of a reserved talkspurt* (shown in Fig. 2).

From Fig. 2., we have the following equations:

$$\alpha \cdot P(RES) = r_s \cdot P(CON) \tag{62}$$

$$\alpha \cdot P(RES) = \beta \cdot P(SIL) \tag{63}$$

$$P(RES) + P(SIL) + P(CON) = 1 \tag{64}$$

We solve the equations to find

$$P(RES) = \frac{r_s \cdot \beta}{\alpha \cdot \beta + r_s(\alpha + \beta)} \tag{65}$$

Let $q_s(j;\ U_s,\ N_s)$ denote the probability that j voice calls have reserved their slots among the U_s admitted voice calls. This can be expressed as

$$q_s(j;\ U_s,\ N_s) = \binom{U_s}{j} \cdot [P(RES)]^j \cdot [1 - P(RES)]^{U_s - j} \tag{66}$$

Let $q_v(j;\ U_v)$ denote the probability that the U_v admitted video calls occupy j payload slots. Then it is given as

$$q_v(j;\ U_v) = \sum_{\sum_{i=1}^{U_v} n_i = j,\ 1 \le n_i \le j} \prod_{i=1}^{U_v} P_{L\_video}(n_i) \tag{67}$$

Let s_i denote the probability that 2i or 2i+1 slots (which can accommodate i data packets) available. For i=0, 1, 2, ..., $\lfloor S_p/2 \rfloor$, we define a set of integer pair of (j, k). Let X_i denote the set. Then X_i can expressed as

$$X_i = \{(j,k) \mid 0 \le j \le U_s, 0 \le k \le U_v, j+k+2i = S_p \ or \ (S_p - 1)\} \tag{68}$$

Then s_i is given as

$$s_i = \sum_{(j,k)\in X_i} q_s(j;U_s,N_s)\cdot q_v(k;U_v) \tag{69}$$

At state *i*, there are *i* data reports which are not acknowledged by the end of the current frame. These reports will be retransmitted in the next frame. According the data traffic model, every data source of the remaining $U_d - i$ data sources will generate a new packet which will trigger a new report in the next frame with a probability of g. Let $q_d(j;\ (U_d - i))$ denote the probability that *j* new reports will be generated among these $U_d - i$ data sources, that is the probability that there will be *j* new reports in the next frame. Then, $q_d(j;\ (U_d - i))$ is given by

$$q_d(j;\ (U_d - i)) = \binom{U_d - i}{j}\cdot(g^j \cdot (1-g)^{U_d-i-j}) \tag{70}$$

The transition probabilities at state *i* of the discrete time Markov chain are as follows:

$$p_{i,j} = \begin{cases} \displaystyle\sum_{m=(j-i)}^{U_d-i} q_d(m;(U_d-i))\cdot s_{m-(j-i)}\ , & 0 \le (j-i) \le (U_d-i),\ j \ne 0 \\ \displaystyle\sum_{m=0}^{\lfloor S_p/2 \rfloor + (j-i)} q_d(m;\ (U_d-i))\cdot s_{m-(j-i)}\ , & 0 < (i-j) < \min(\lfloor S_p/2 \rfloor i),\ j \ne 0 \\ \displaystyle\sum_{m=0}^{\max((U_d-i),\lfloor S_p/2 \rfloor)} q_d(m;(U_d-i))\cdot \sum_{l=i+m}^{\lfloor S_p/2 \rfloor} s_l\ , & 0 \le (j-i) \le (U_d-i),\ j = 0 \\ 0\ , & \text{otherwise} \end{cases} \tag{71}$$

For steady state probabilities, we get equations as follows:

$$\begin{bmatrix} p_0 \\ p_1 \\ p_2 \\ \vdots \\ p_{U_d} \end{bmatrix} = \begin{bmatrix} p_{0,0} & p_{1,0} & \cdots & p_{U_d-1,0} & p_{U_d,0} \\ p_{0,1} & p_{1,1} & \cdots & p_{U_d-1,1} & p_{U_d,1} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ p_{0,U_d-1} & p_{1,U_d-1} & \cdots & p_{U_d-1,U_d-1} & p_{U_d,U_d-1} \\ p_{0,U_d} & p_{1,U_d} & \cdots & p_{U_d-1,U_d} & p_{U_d,U_d} \end{bmatrix} \begin{bmatrix} p_0 \\ p_1 \\ p_2 \\ \vdots \\ p_{U_d} \end{bmatrix} \tag{72}$$

$$\sum_{i=0}^{U_d} p_i = 1 \tag{73}$$

We solve these equations using the Gauss-Seidel iteration algorithm.

$$\begin{cases} 0 & for\ n \le k \le u \\ \dfrac{1}{n^u} \displaystyle\sum_{m=\max(k,n-\left\lceil \frac{u-k}{2} \right\rceil)}^{n-1} \binom{n}{m}\binom{u}{k}\binom{m}{k} k! \frac{[2(m-n)]!}{(2!)^{(n-m)}} (n-m)^{((u-k)-2(n-m))} & for\ 1 \le k \le n-1 \\ \dfrac{1}{n^u} \displaystyle\sum_{m=1}^{\min(n,\left\lceil \frac{u}{2} \right\rceil} \frac{(2m)!}{(2!)^{(n-m)}} m^{(u-2m)} & for\ k = 0 \end{cases}$$

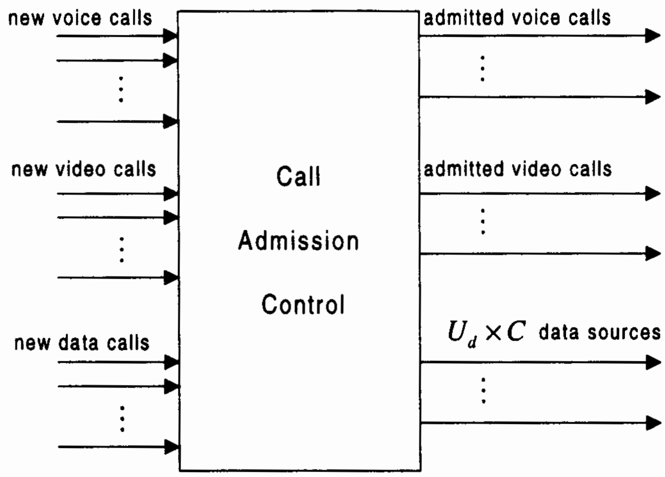

Fig. 1. Call admission control regulates each traffic class in each cell.

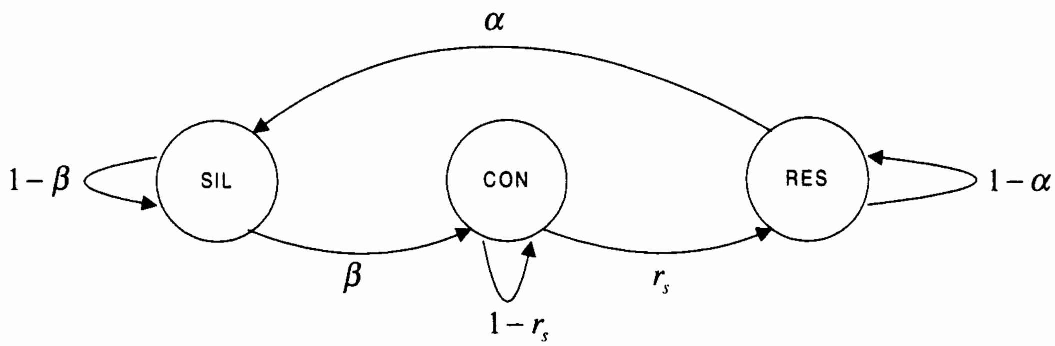

Fig. 2. Activity of a single admitted voice call

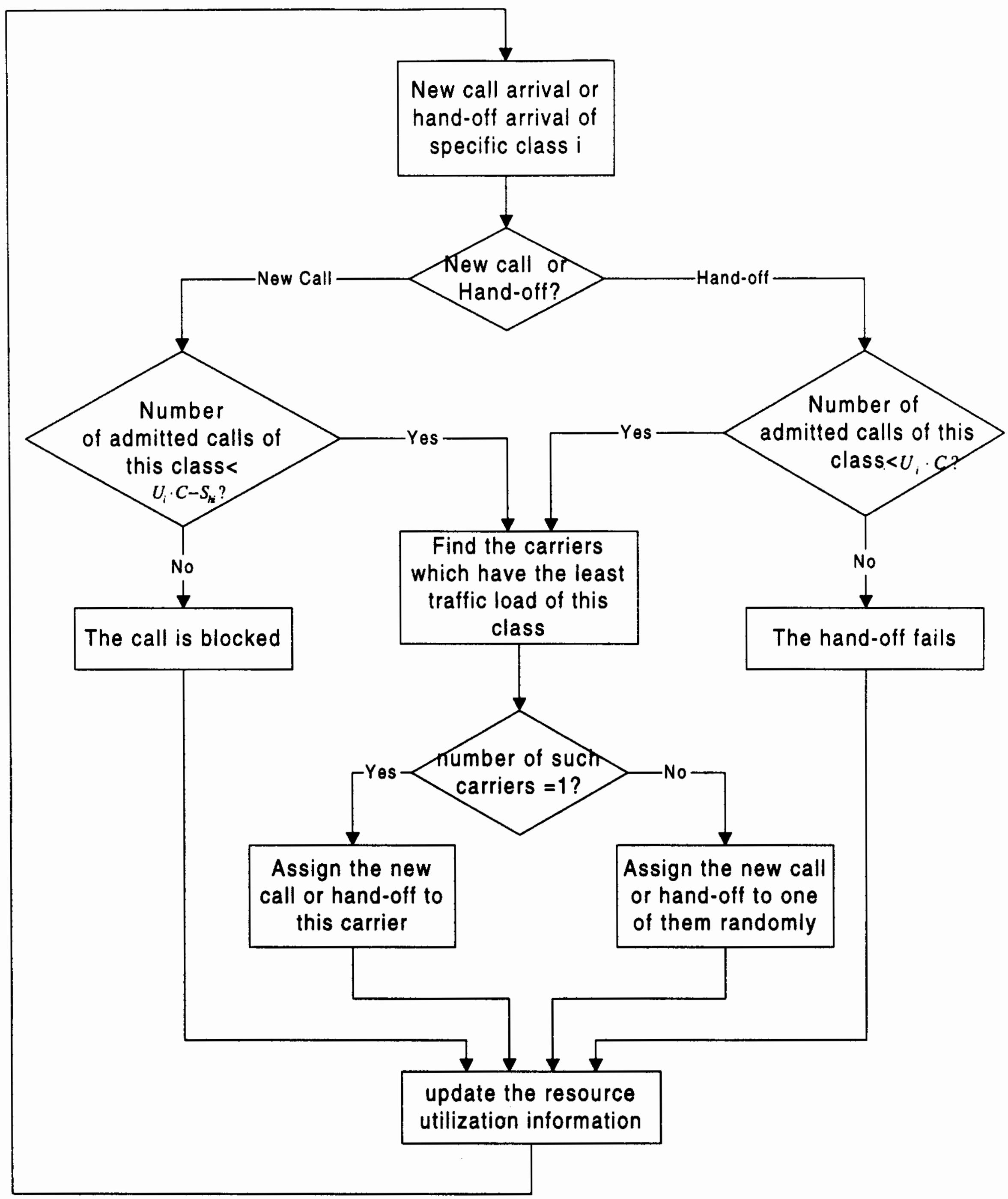

Fig. 3. The flowchart of call admission functions

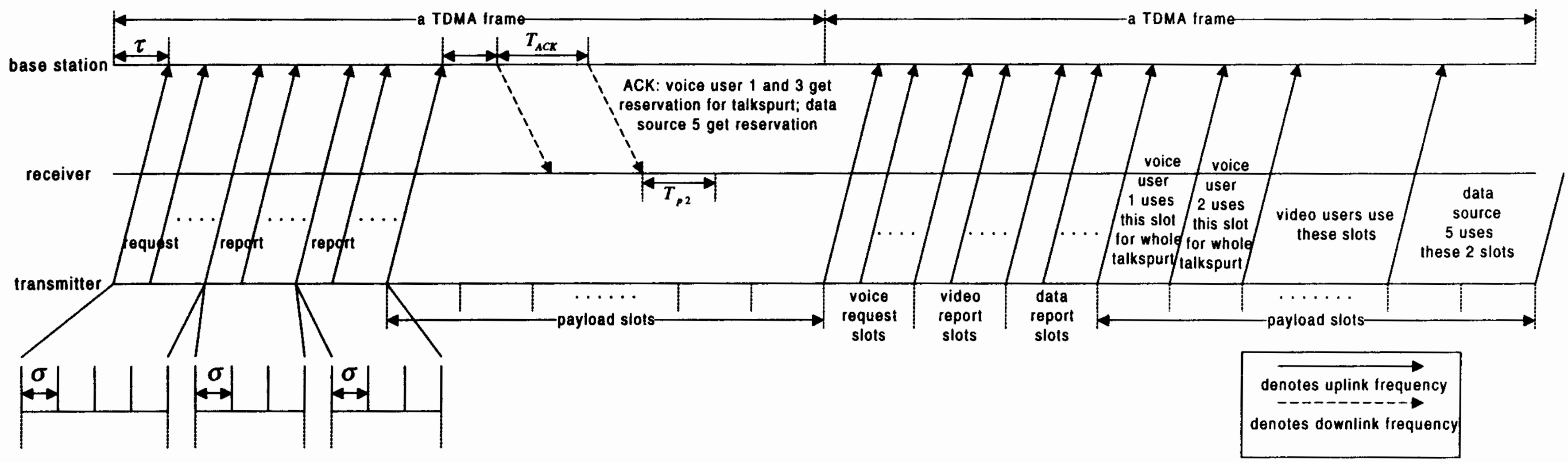

Fig. 4. The timing of the medium access scheme.

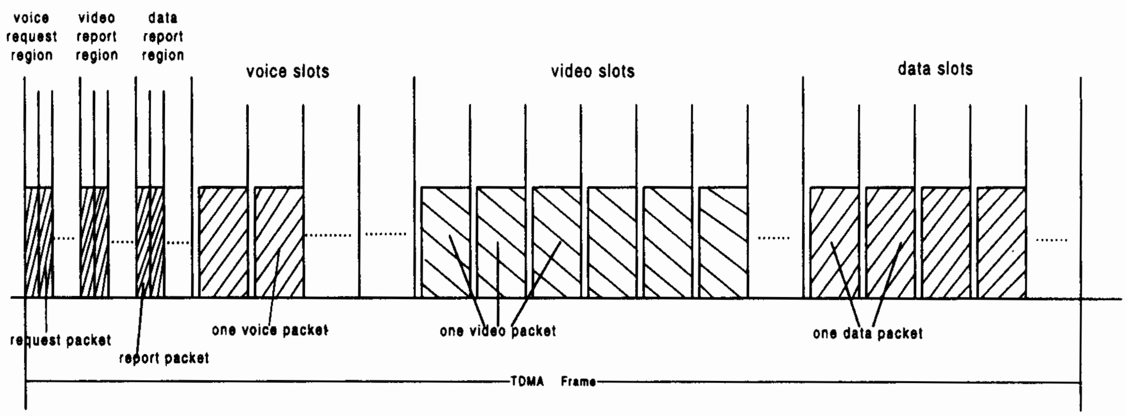

Fig. 5. The structure of R&R-DSA TDMA frame.

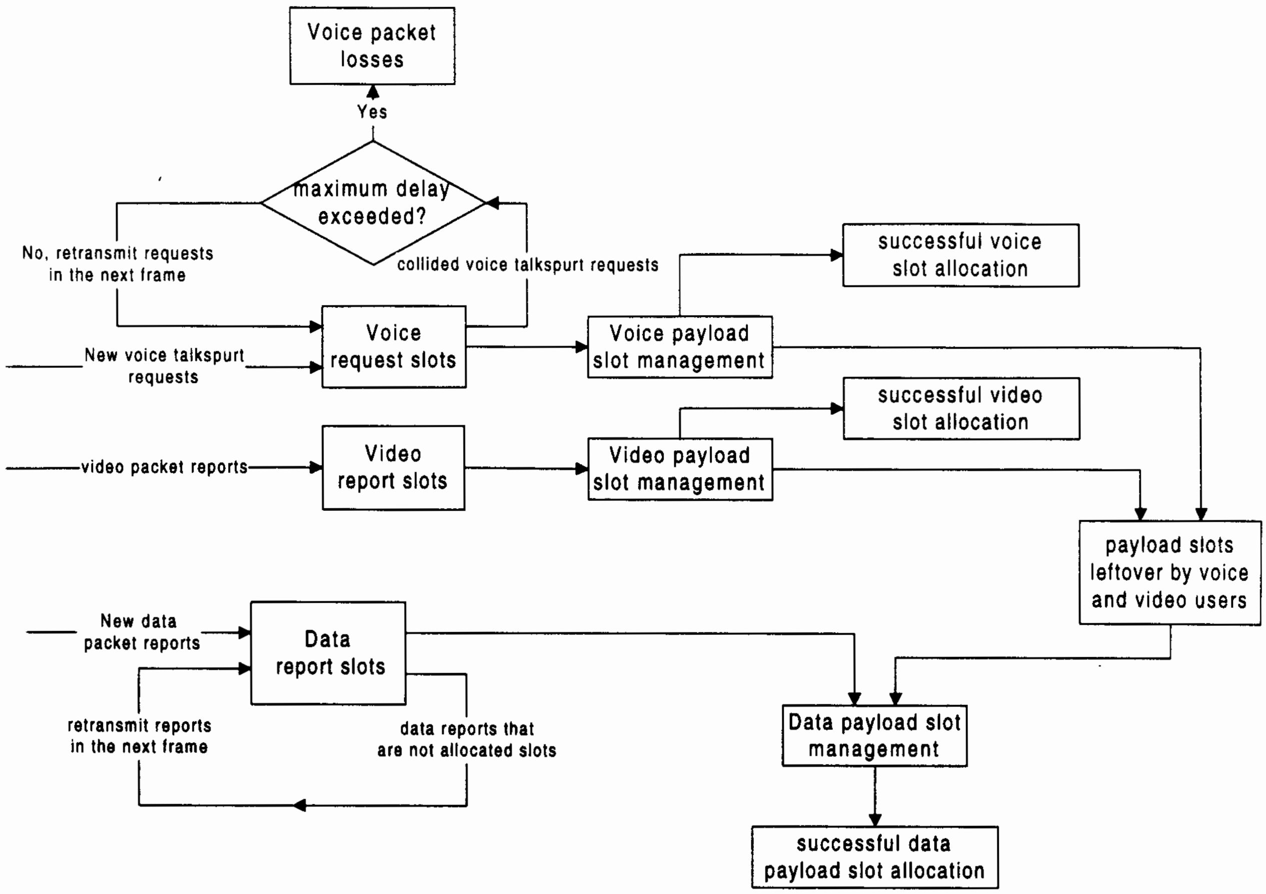

Fig. 6. R&R-DSA TDMA scheme

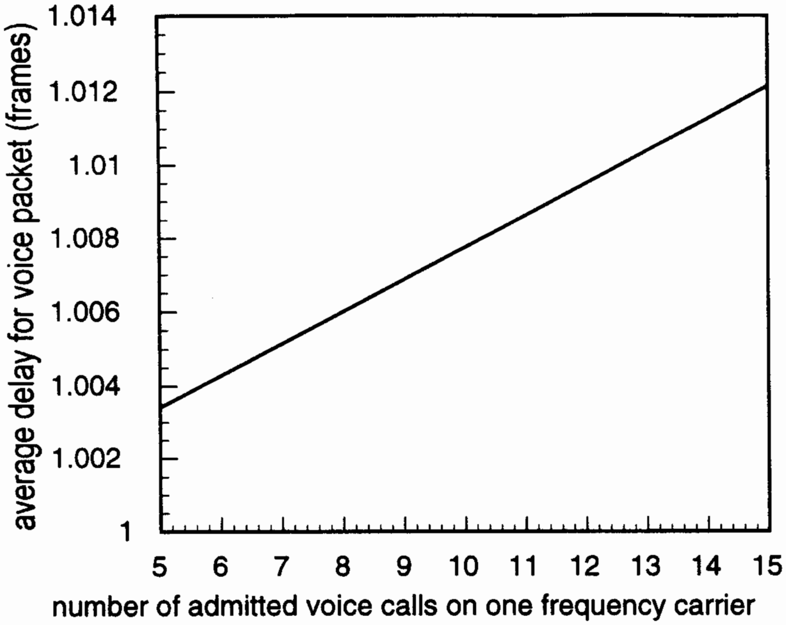

Fig. 7. Average delay of voice packets: N_s =5.

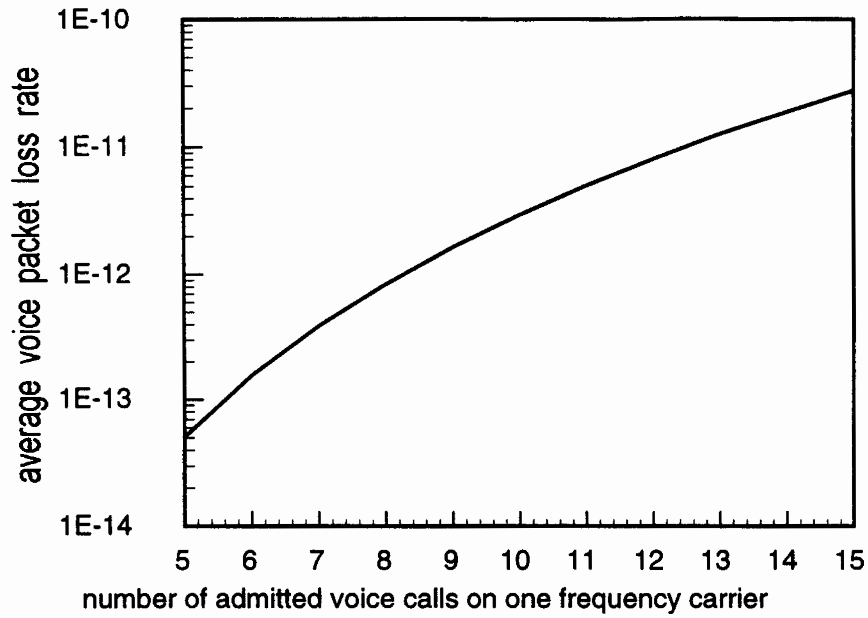

Fig. 8. Average packet loss rate of voice packets: N_s =5.

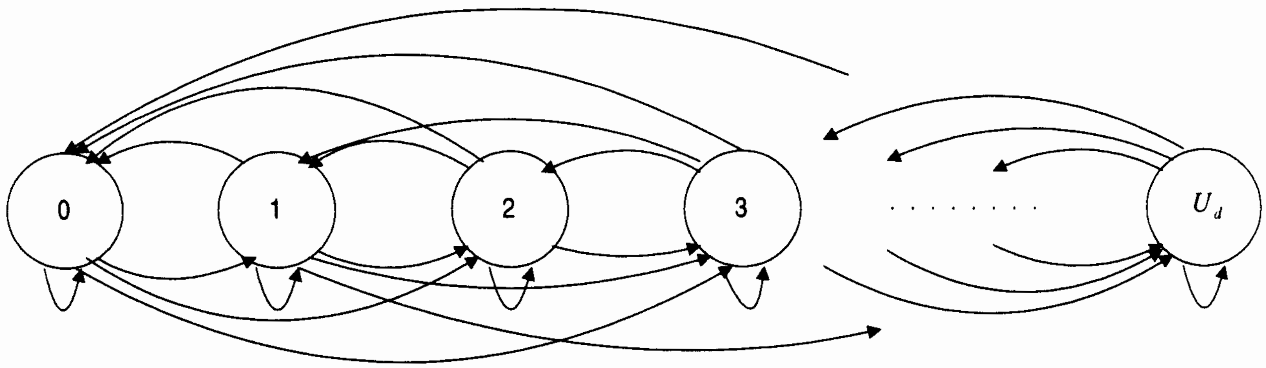

Fig. 9. The discrete time Markov chain model of the distributed queue of data reports.

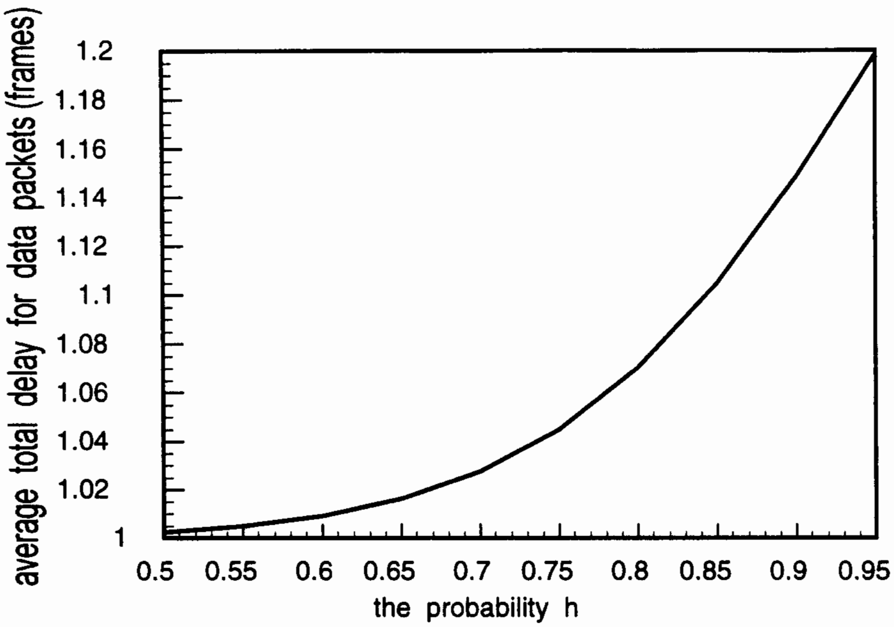

Fig. 10. The average delay for data packets: 10 data sources with generation probability *h* but no voice and video traffic on the frequency carrier.

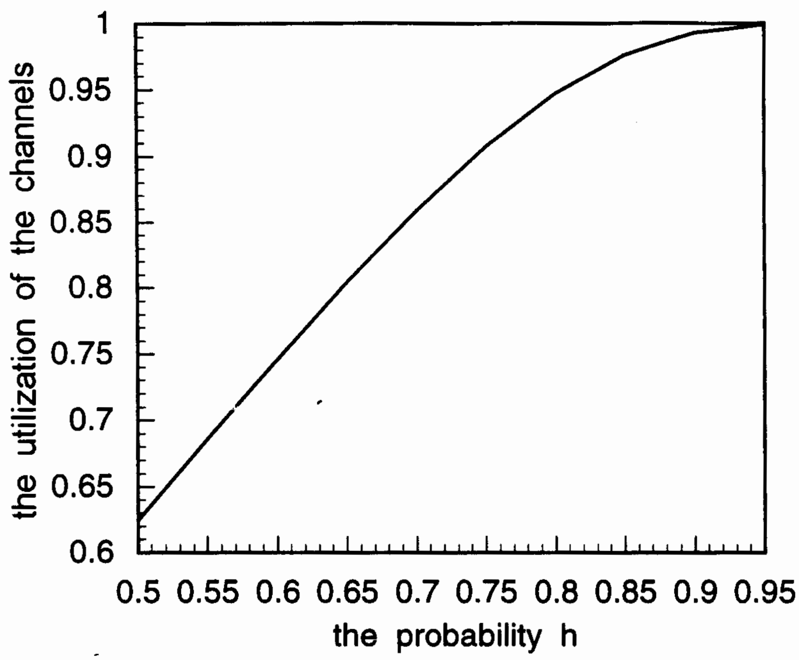

Fig. 11. The utilization of channel resources: 10 data sources with generation probability *h* but no voice and video traffic on the frequency carrier.

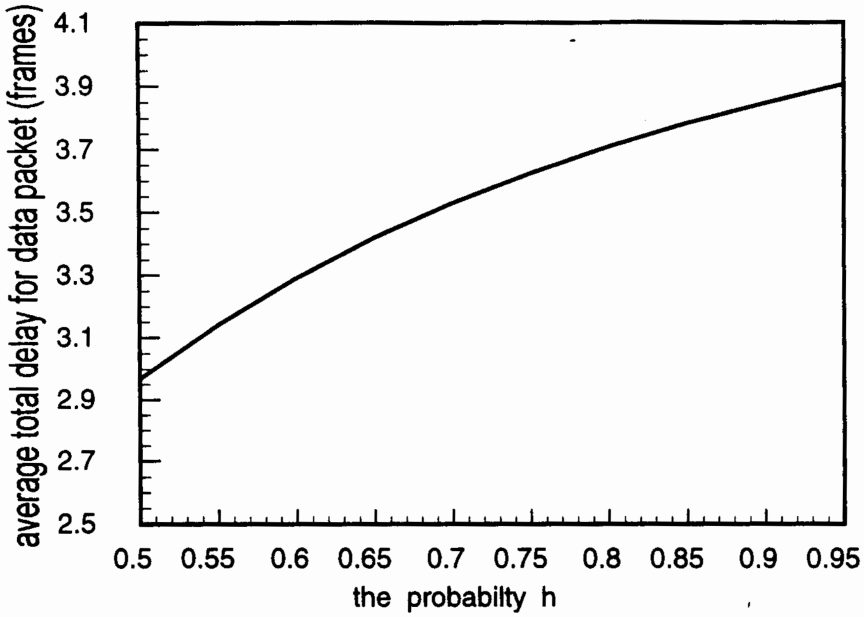

Fig. 12. The average total delay of data packet: 4 voice calls, 3 video calls and 10 data sources with generation probability h on the frequency carrier.

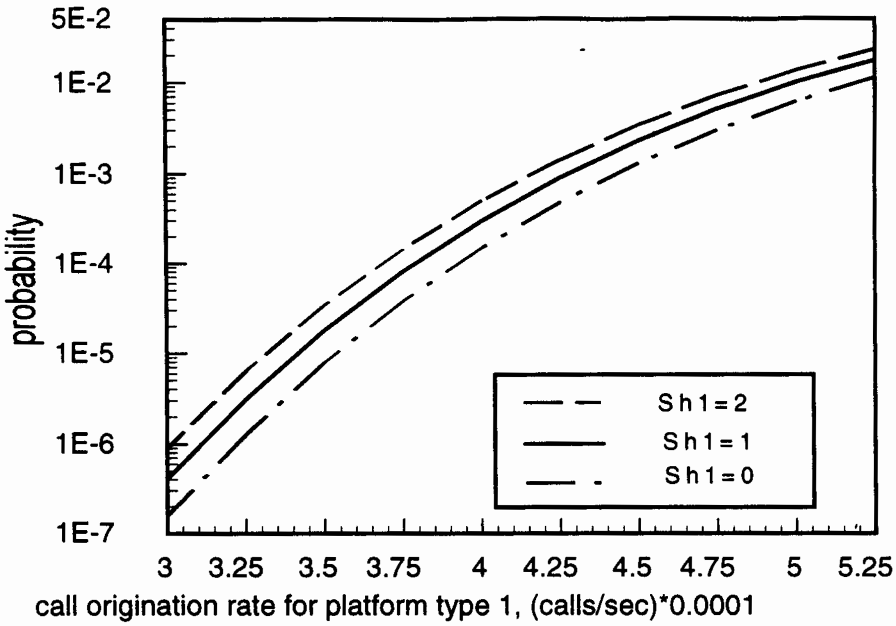

Fig. 13. Blocking probability for voice calls: C=4, G=2, I=2, U_s=4, $\nu(1,0)=\nu(2,0)=350$, $\Lambda_n(1,1)/\Lambda_n(2,1)=1.0$, $\overline{T}(1)=100$s, $\overline{T}_D(1)=1000$s, $\overline{T_D}(2)=200$s.

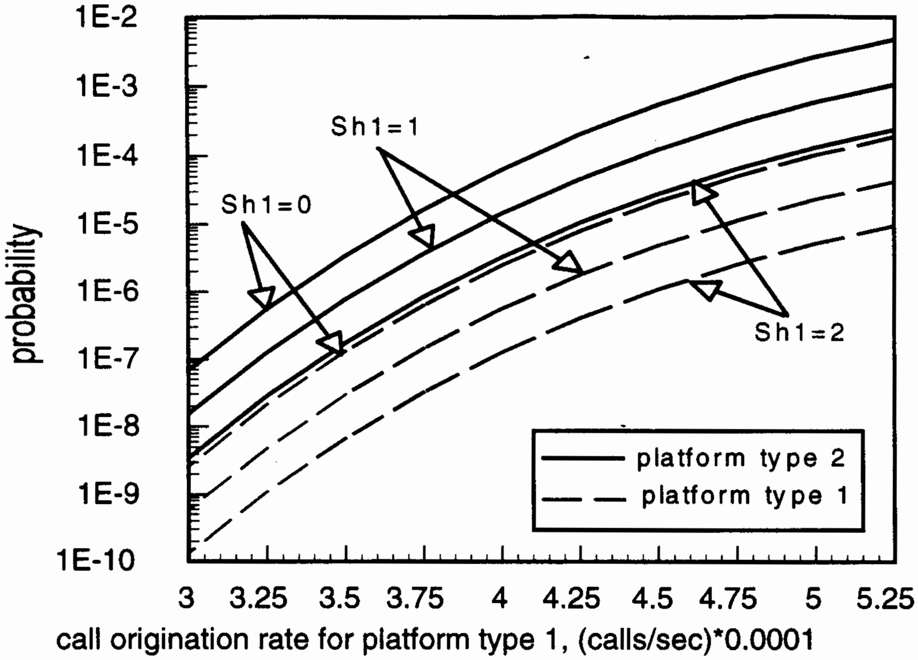

Fig. 14. Forced termination probability for video calls: C=4, G=2, I=2, U_s=4, $v(1,0)=v(2,0)$=350, $\Lambda_n(1,1)/\Lambda_n(2,1)$=1.0, $\overline{T}(1)$=100s, $\overline{T}_D(1)$=1000s, $\overline{T_D}(2)$=200s.

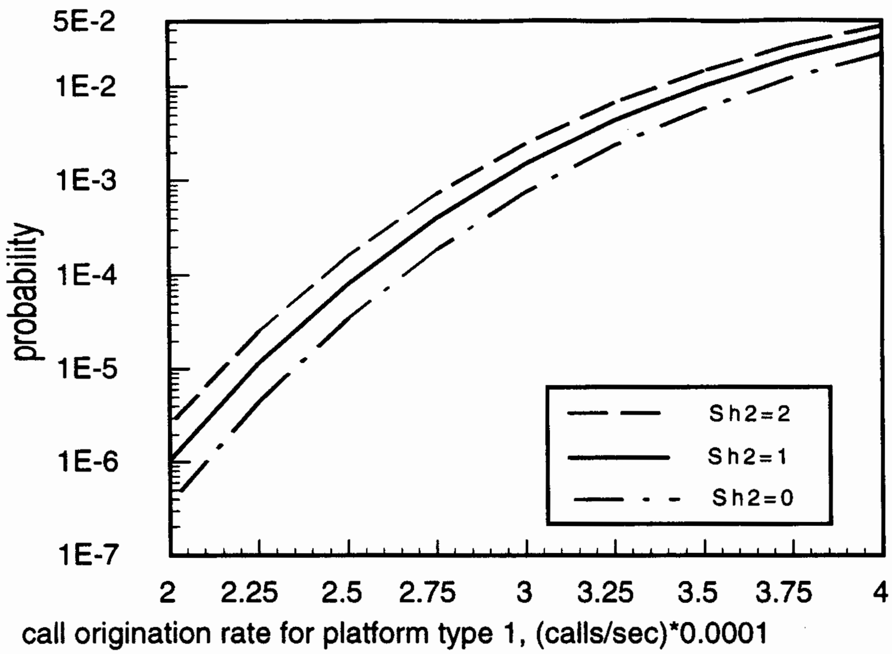

Fig. 15. Blocking probability of video calls: C=4, G=2, I=2, U_v=3, $v(1,0) = v(2,0)$=350, $\Lambda_n(1,2) / \Lambda_n(2,2)$=1.0, $\overline{T}(2)$=100s, $\overline{T}_D(1)$=1000s, $\overline{T_D}(2)$=200s.

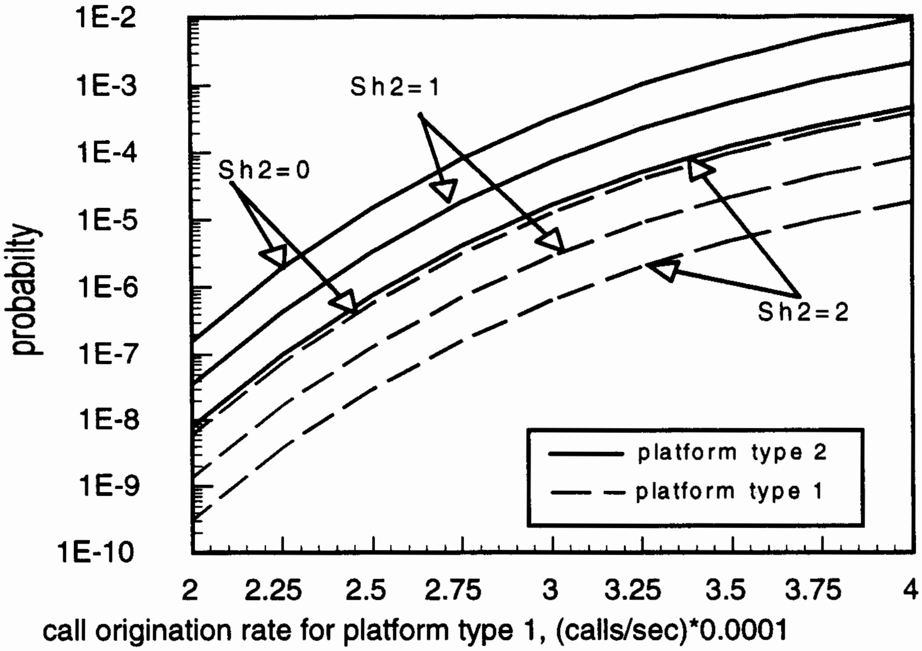

Fig. 16. Forced termination probability for video calls: C=4, G=2, I=2, U_v=3, $v(1,0)=v(2,0)$=350, $\Lambda_n(1,2)/\Lambda_n(2,2)$=1.0, $\overline{T}(2)$=100s, $\overline{T}_D(1)$=1000s, $\overline{T_D}(2)$=200s.