11-1-2000

AN ACCESS PROTOCOL TO PROVIDE QOS IN WIRELESS :NETWORKS

Xin Liu
Purdue University School of Electrical and Computer Engineering

Edwin K. P. Chong
Purdue University School of Electrical and Computer Engineering

Ness B. Shroff
Purdue University School of Electrical and Computer Engineering

AN ACCESS PROTOCOL TO PROVIDE QOS IN WIRELESS NETWORKS

XIN LIU
EDWIN K. P. CHONG
NESS B. SHROFF

TR-ECE 00-14
NOVEMBER 2000

SCHOOL OF ELECTRICAL
AND COMPUTER ENGINEERING
PURDUE UNIVERSITY
WEST LAFAYETTE, INDIANA 47907-1285

AN ACCESS PROTOCOL TO PROVIDE QOS IN WIRELESS :NETWORKS*

Xin Liu, Edwin K. P. Chong, and Ness B. Shroff

School of Eletrical and Computer Engineering

1285 Electrical Engineering Building

Purdue University

West Lafayette, IN 47907-1285

'The work was supported by the National Science Foundation under grants ECS-9410313, ECS-9501652, CDA 96-17388, NCR-9624525, and ANI-9805441.

TABLE OF CONTENTS

LIST OF FIGURES

ABSTRACT

In wireless networks, contention-and-reservation schemes provide promising implementations of packet switching, which efficiently multiplexes different classes of traffic. In this report, we present an access scheme to satisfy the QoS requirements for two classes of traffic during the contention-based communication. In this algorithm, different classes of users contend with other users for resources based on controlled class-dependent permission probabilities. We prove that our algorithm is stable for a large class of arrival processes. Under certain QoS requirements, we derive an upper-bound for the throughput for a general class of random access algorithms. We show that the throughput of our algorithm asymptotically approaches this upper-bound. We also consider the algorithm with a capture model in the presence of near/far effects and Rayleigh fading with lognormal shadowing. We present a class-distance-dependent permission probability, which provides location fairness, certain delay guarantees, and a good throughput.

1. INTRODUCTION

The development of wireless communication networks attracts intense interests from academia and industry. The goal of wireless communications is to provide a convenient and economical way for all people to transfer all kinds of information, such as voice and data. Compared with circuit switching, packet switching provides more efficient multiplexing of different classes of traffic. In circuit switched networks, when a user is admitted to the network, a certain amount of network resource is assigned to the user and exclusively used by the user until its communication finishes, regardless of whether the user has information to transmit during this period. In packet switched networks, when a new user is admitted, no specific resource is assigned to it. Resources are shared by users in the system. A user only occupies the network resource when it has information to transmit. Consider a phone call as an example. When the user talks, voice packets are generated at a certain rate; when the user is silent, no voice packet is generated. On average, the user talks less than half of the whole call duration. In circuit switched networks, the networks assign the voice user the resource equivalent to its packet rate during talking, so about half of the resources is wasted. In packet switched networks, when a user does not talk, no resource is assigned to this user; when the user begins talking after a period of silence, the network assigns resource to this user again. Hence, packet switching utilizes network resources more efficiently than circuit switching in general. Efficiency is very important for wireless networks because wireless bandwidth is scarce. However, wireless packet switching scheme suffers access problems in the uplink. In other words, when a user becomes active, it has packets to transmit and no network resource is assigned to it, the user has to compete with other users to gain the access to network resources. To solve this problem, a variety of contention and reservation medium

access control (MAC) protocols have been widely used in the area of communication networks [1, 2, **3,** 4]. Typically, there are two transmission phases:

1. Newly activated users compete to gain access to the networks. The first packet of a newly activated user is transmitted through the network using some random access protocols; i.e., contention-based communications. This first packet may be a packet in a special form or a normal data packet. In this report, we call the first packet a *request.* If the first packet is lost during transmission, or is received in error, then it is retransmitted until successful.

2. Following the first successful contention-based transmission, subsequent transmissions are scheduled contention-free using a scheduling strategy.

We call the first phase the *contention phase* and the second phase the scheduling phase. In this report, we focus on the contention phase of communications. In packet switched wireless networks, the contention phase may exist throughout the whole communication period, and not only during the admission period. Every time a user becomes active (say, a user begins talking after being silent), at that very moment, because no resource is assigned to the user, the user has to inform the base station about its resource requirement through contention-based communication. Hence, contention-based communication plays an important role in packet-switched wireless networks.

In packet switched networks, admission control and resource allocation are used to provide QoS. In general, admission control is based on the resource allocation scheme. In wired networks, resource allocation is implemented by smart scheduling schemes. However, smart scheduling is not enough to provide QoS for wireless networks, where contention plays an important part. For example, we want to provide delay guarantee to real-time traffic in wireless networks. When a user begins talking, it first sends its request to the base station through random access; i.e., contention-based transmission. Then the base station schedules the traffic after it receives resource request from the user. Therefore, the user experiences delay caused by contention

plus the delay caused by scheduling. To guarantee the delay experienced by the user, we need to guarantee the delay in both contention phase and scheduling phase. During the scheduling phase, smart scheduling strategies can be used to provide delay guarantees. However, we also need to algorithms in the contention phase to provide delay guarantees to users. To provide QoS in the contention phase is intrinsic difficult due to the nature of random access. While there is a significant body of work on the development of effective scheduling and admission control policies to ensure QoS, there is very little work done in implementing QoS during the contention phase of communication.

In this report, we present an algorithm that implements QoS requirements for two classes of traffic in the contention phase of packet switched time-slotted wireless networks. Controlled time-slotted ALOHA is the random access algorithm considered in this report. Two traffic classes, voice and data, are considered. We consider only two classes for the convenience of calculation and explanations, although more classes can be considered similarly. We assume that voice users have *delay requirements* and data users do not have such requirements.

We consider the QoS algorithm under two conditions: with and without exploiting capture. In wire-line networks, if two or more users transmit at the same time through the same media, usually all of them are assumed to be failed. However, this assumption may be unnecessarily pessimistic in the mobile radio environment, where the received packets at the base station are subject to the near/far effect and channel fading. Packets from different users in the same slot may arrive at the base station with different power levels and the base station may successfully decode one or more packet. This is referred to as *capture*. It is obvious that the system throughput will be improved if the system explores capture. However, unfairness exists between near and far users due to the nature of radio transmission. In this report, we present a distance-dependent permission probability scheme, which provide distance fairness with a good throughput.

In summary, if we do not consider the ability of capture, the QoS requirement is presented in terms of delay. When we consider capture, the QoS requirement is explained in terms of delay and distance fairness.

This report is organized as follows. In Section 2, we describe the system model. We present and analyze the QoS algorithm without capture in Section **3.** An upper-bound for the throughput is derived, under certain QoS requirements, for a general class of random access algorithms. The throughput of our algorithm asymptotically approaches this upper-bound. In Section 4, we adopt a SIR capture criterion and a propagation model considering the near/far effect and slow Rayleigh fading with log-normal shadowing. We present a distance-dependent permission probability scheme. In this scheme, users at different distances from the base station transmit with different probabilities to achieve fairness and good throughput. We provide numerical results for distance-dependent permission probability functions because there is no close form. Simulation results are provided in Section 5. Conclusion and future work are presented in Section 6.

2. SYSTEM MODEL

In this section, we describe the system model. There is a base station with mobile users in its coverage area. We consider the uplink of a time-slotted system and focus on the contention phase of communication. We assume that time is divided into frames and each frame consists of M request slots. Each request slot is large enough to contain a fixed size request. The base station monitors and controls the contention phase in the system. *In the following, when we mention users we mean newly activated users with requests to transmit, except otherwise specified.*

At the beginning of a frame, the base station broadcasts a permission probability for each class of users through a non-collision error-free signaling channel. A user decides whether or not to transmit in a request slot in the frame according to the permission probability of its class broadcasted by the base station. Different classes of users may have different permission probabilities.

We assume that a user can transmit at most once in a frame. There are M request slots in each frame. The parameter, M, determines how often. the base station updates its control parameters, and how long a user waits for before it retransmits. In practice, the larger the value of M, the less the signaling, the better the estimation of the number of users, and the longer the delay.

In some cases, we prefer a large value of M. An example of a practical application is in satellite communications. After the contention of a time slot, a user cannot know immediately whether its request is successfully received by the hub station. In satellite communications, the round trip delay is relatively large. For instance, the propagation delay is around 20–25ms for LEO (low earth orbit) systems [5]. An immediate ack from the the hub is impossible. Furthermore, the coverage area of satellite communications is relatively large, it is difficult for an earth station to

detect whether its transmission is successful. Hence, a large value of M may be suitable for such a case. In other cases, a small value of M could be favored. A good example of such a case is a local wireless network, where the sum of the round trip delay, and processing time, etc., is small. A user transmits, then waits for acknowledgment. If the user does not receive an acknowledgment from the base station in the predetermined waiting time, it assumes that the transmission failed. The user could retransmit it in the next frame. The extreme case is where $M = 1$; i.e., a user can retransmit its request in the next request slot. In the extreme case $M = 1$, the scheme studied in this report becomes the pure priority scheme; i.e., when there are voice users, no data user transmits, and when there is no voice user, data users transmit. However, even in wireless LAN, it is not necessary to adopt a very small value of M (say, $M = 1$). Usually, the requests are much shorter than normal data packets. Hence, the delay caused by several request slots are tolerable in order to reduce the cost of extensive signaling-s.

In Section 3, we assume that the system is not capable of correctly deciphering any transmissions when two or more overlapping transmissions arrive in the same slot; i.e., if two or more users transmit their requests through the same request slot in a frame, neither of them can be successfully received. This situation is called collision. In Section 4, we consider a system that exploits capture. When two or more packets are transmitted at the same time slot, it is possible that one or more packet could be successfully received. The capture model that we use in this report is based on signal to interference ratio (SIR). The SIR capture ratio, R, is predetermined. If the SIR (ratio of the received power of one user to the sum of power of all others) is larger than the capture ratio R, the packet is assumed to be received successfully. The propagation model considered includes the near/far effect and fast fading with shadowing. The capture ratio R is an important parameter that reflects the physical layer requirement for reliable communication. The following is some typical SIR requirements in different analog/digital cellular systems. For example, AMPS (Advanced Mobile Phone System) requires $R \approx 17 - 18$dB. U.S. IS-54 and IS-136 TDMA

reduces the requirement to 14 dB because they employ digital techniques. Due to its more robust modulation scheme, GSM (Global System for Mobile communications), however, can tolerate SIR as low as 6.5 to 9 dB. The capture ratio determines how difficult it is for capture to occur. When the capture ratio is relatively large, it is unlikely that a packet can succeed when two or more users are transmitting. Hence, the case of large capture ratio can be approximated by the model without capture. We discuss the capture ability in detail in Section 4.

We assume that a request is never discarded; i.e., a user always retransmits its request until it is acknowledged by the base station that its request has been received successfully. While the request of a user is delayed, some packets may be buffered at the user. In real-time applications, human factors may decide whether to send a delayed packet or to drop it. This issue is irrelevant to our scheme. Furthermore, we assume that the acknowledgment is error-free and the base station uses a scheduling strategy to decide when the active user should transmit in the reservation phase of communication.

3. QOS ALGORITHM WITHOUT CAPTURE

In this section, we assume that the system does not exploit capture; i.e., when only one user transmits in a request slot, the transmission succeeds; when two or more users transmit in the same request slot, neither of them succeed. We first present the QoS algorithm with restriction to the delay requirement of voice users. Then we analyze the throughput and stable condition. Finally, we derive a throughput upper-bound under the QoS requirement for a large class of random access algorithms.

3.1 Algorithm

Denote p_v (p_d) as the permission probability that a voice (data) user transmits in a request slot in a frame. In this report, the permission probabilities, p_v and p_d, are used to stabilize the ALOHA system, to achieve good throughput, and to provide QoS guarantees. The use of permission probabilities to stabilize ALOHA is not a new idea. Permission probabilities are also used to provide priority to voice users in [3, 6]. In the literature, there are algorithms centralized or decentralized to estimate the number of users in the system. All these algorithms can be used in our scheme. Hence, we focus on how to use the permission probabilities to satisfy QoS instead of how to estimate the number of users. During the analysis we assume that the base station knows the precise numbers of voice users and data users in each frame. Knowing this information is the ideal condition of the algorithm. Practically, we use a Kalman filter to estimate the numbers of voice users and. data users with requests in each frame. We show through simulations that using a Kalman filter for the estimation provides very good results.

As mentioned before, a user can transmit at most once in a frame. We do not distinguish between newly arrived users and retransmitted users. 'The base station

broadcasts p, and p_d at the beginning of frame i. A voice user randomly selects a request slot to transmit in this frame with probability p,, as would a data user with probability p_d. All users select and transmit independently. 'The base station acknowledges those users whose requests have been successfully accepted at the end of frame i. Users that have not been acknowledged assume that their requests have not been successfully transmitted. They retransmit in the next frame. The base station estimates the number of users in the system, calculates p, and p_d for frame i + 1, and so on. It is easy to prove that the throughput is maximized when M users transmit in each frame (Appendix A). However, this throughput may come at the cost of excessive delay for voice users. Hence, we need to develop a scheme that attempts to maximize throughput subject to a given level of delay requirement for voice users.

A good measure of QoS is the delay experienced by a user before its request is successfully received by the base station. However, the precise delay distribution of voice users is very difficult to find in this context. Thus, we define an average success probability, $\bar{P}_s$, as the QoS measure used in this report. Suppose the system has reached steady state. When a voice user becomes active, on average, it transmits its request successfully with probability $\bar{P}_s$, given by

$$\bar{P}_s \;:=\; E\left[p_s(N_v, N_d)\right] = \sum_{i,j} p_s(i,j)\pi(i,j), \tag{3.1}$$

where $p_s(i,j)$ is the probability that a voice user transmits its request successfully in a frame in steady state when there are i voice users and j data users in the system, and $\pi(i,j)$ is the steady state distribution that i voice users and j data users are in the system.

Our QoS requirement for voice users is $\bar{P}_s \geq A_0$, where A_0 is the given delay threshold. Roughly speaking, the contention delay of a voice user is geometrically distributed with parameter $\bar{P}_s$; i.e., the distribution of access delay D is approximated by $P(D = x) = \bar{P}_s(1 - \bar{P}_s)^{x-1}$. The larger the M, the better the approximation. In Section 5, we show the distribution of voice users from simulations is well

approximated by a geometric distribution (see Figure 5.1).

The QoS algorithm is described as follows. Suppose that the base station knows that N_v voice users and N_d data users are in the system. Then, the permission probabilities of voice users and data users are

$$\begin{aligned} p_v &= \min(1, \frac{M}{N_v}), \\ p_d &= \begin{cases} \min\left(1, \frac{(C-N_v)^+}{N_d}\right) & : \text{ if } N_v > 0, \\ \min\left(1, \frac{M}{N_d}\right) & : \text{ if } N_v = 0, \end{cases} \end{aligned} \tag{3.2}$$

where

$$(x)^+ = \begin{cases} x & : \text{ if } x \geq 0, \\ 0 & : \text{ otherwise.} \end{cases}$$

Note that C is a tuning parameter used to satisfy the QoS requirements of voice users. So the algorithm does the following. If the number of voice users in the system is less than M, all voice users can transmit freely. In this case, data users may or may not be allowed to transmit. If the number of voice users in the system is greater than M, then a voice user is allowed to transmit based on the outcome of the toss of a biased coin with probability M/N_v of success. In this case, no data users are allowed to transmit. Before we illustrate how to calculate C, we first make a few observations:

- Data users yield to voice users the right to access request slots.
- The parameter C satisfies $0 \leq C \leq M$. The expected number of data users to transmit is $(C - N_v)^+$. The total throughput is maximized when $C = M$. The larger the value of C, the higher the throughput, and the larger the delay of voice users. Hence, there is a tradeoff between the throughput of the system and the delay requirement of voice users. When the QoS requirement is stringent, C is small, data users are allowed to access request slots with lower probability, and voice users have a higher probability to succeed in a frame.
- When there is no voice user; i.e., $N_v = 0$, the value of p_d is set to maximize the throughput.

The tuning parameter C can be calculated theoretically (see Appendix B). Practically, there is a very simple approximation for C. Let K_0 satisfy

$$\left(1-\frac{1}{M}\right)^{K_0-1} = A_0. \tag{3.3}$$

If K_0 is not too small compared to M and the fraction of voice users is not too large, then K_0 is a good approximation of C. In this case, the number of voice users in the system in steady state is seldom larger than K_0. Therefore, the average delay $\bar{P}_s$ is:

$$\begin{aligned}\bar{P}_s &= E(p_s) = E(p_s(i)|i \le C)p(i \le C) + E(p_s(i)|i > C)p(i > C) \\ &\approx (1-\frac{1}{M})^{C-1} = (1-\frac{1}{M})^{K_0-1} = A_0.\end{aligned}$$

In fact, if $K_0 \ge 0.5M$ and the fraction of voice users is less than 70%, C $\approx K_0$ is a good approximation. We set C $= K_0$ in simulations in Section 5 and find that it works well.

We, next, analyze the algorithm. First, we calculate the throughput. Second, we prove that the algorithm is stable for a large class of arrival processes. Then, we derive an upper bound on the throughput of random access algorithms under the QoS requirement $\bar{P}_s \ge A_0$. We show that the throughput of our algorithm asymptotically approaches the upper-bound.

3.2 Throughput

Suppose there are k users transmitting in a frame. Each user selects one of the request slots randomly and independently. In this section, neither capture ability nor transmission error is considered. The throughput, T_k, is defined as the average number of requests that are successfully transmitted in a frame and p_k is the probability that a user transmits successfully. We then have

$$\begin{aligned}T_k &= k\left(1-\frac{1}{M}\right)^{k-1}, \\ p_k &= \frac{T_k}{k} = \left(1-\frac{1}{M}\right)^{k-1}.\end{aligned}$$

We consider the throughput under three conditions:

1. When $N_v \geq C$, each voice user transmits in a request slot with probability $p_v = \min(1, M/N_v)$ and no data user transmits. The throughput is:

$$\begin{aligned} T(N_v, N_d) &= \sum_{i=0}^{N_v} T_i P(i \text{ voice users transmit in this frame}) \\ &= \sum_{i=0}^{N_v} i \left(1 - \frac{1}{M}\right)^{i-1} \binom{N_v}{i} p_v^i (1-p_v)^{N_v - i} \\ &= N_v p_v \sum_{i=0}^{N_v} \binom{N_v - 1}{i-1} \left(\left(1 - \frac{1}{M}\right) p_v\right)^{i-1} (1-p_v)^{(N_v-1)-(i-1)} \\ &= N_v p_v \left(1 - \frac{p_v}{M}\right)^{N_v - 1}. \end{aligned} \tag{3.4}$$

2. When $N_v < C$, each voice user transmits in a request slot with probability 1 and each data users transmits with probability $p_d = (C - N_v)/N_d$. Therefore,

$$\begin{aligned} p_s(N_v, N_d) &= \sum_{i=0}^{N_d} p_{i+N_v} P(i \text{ data users transmit in this frame}) \\ &= \sum_{i=0}^{N_d} \left(1 - \frac{1}{M}\right)^{N_v + i - 1} \binom{N_d}{i} (1-p_d)^{N_d - i} {p_d}^i \\ &= \left(1 - \frac{1}{M}\right)^{N_v - 1} \left(1 - \frac{p_d}{M}\right)^{N_d}. \end{aligned}$$

The throughput consists of successfully transmitted voice and data requests:

$$\begin{aligned} T(N_v, N_d) &= \sum_{i=0}^{N_d} T_{i+N_v} P(i \text{ data users transmit in this frame}) \\ &= \sum_{i=0}^{N_d} (N_v + i) \left(1 - \frac{1}{M}\right)^{N_v + i - 1} \binom{N_d}{i} (1-p_d)^{N_d - i} {p_d}^i \\ &= N_v \left(1 - \frac{1}{M}\right)^{N_v - 1} \left(1 - \frac{p_d}{M}\right)^{N_d} \\ &\quad + (C - N_v) \left(1 - \frac{1}{M}\right)^{N_v} \left(1 - \frac{p_d}{M}\right)^{N_d - 1}. \end{aligned} \tag{3.5}$$

3. When $N_v = 0$, data users transmit with probability p_d, $p_d = min(1, N_d/M)$, to maximize the throughput.

$$T(0, N_d) = N_d p_d \left(1 - \frac{p_d}{M}\right)^{N_d - 1}. \tag{3.6}$$

3.3 Stability Analysis

We now prove that our algorithm is stable with a fairly weak assumption on the arrival process. We consider a system with a unique stationary clistribution as a stable system. We use Pake's Lemma to find a sufficient condition for the system to be stable [7].

Lemma 1 (Pake's Lemma) Let $\{X_k, k = 0, 1, 2, \cdots\}$ be an irreducible, aperiodic homogeneous Markov chain with state space $\{0, 1, 2, \cdot.\cdot\}$. The following two conditions are *sufficient* for the Markov chain to be ergodic.

$$\text{a)} \quad |E(X_{k+1} - X_k | X_k = i)| < \infty, \ \forall\, i,$$

$$b) \quad \limsup_{i\to\infty} E(X_{k+1} - X_k | X_k = \mathrm{i}) < 0.$$

Note that an irreducible, aperiodic, ergodic Markov chain has a unique stationary distribution.

Let A_k be the total number of users that arrive in the kth frame. Suppose that $\{A_k, k = 0, 1, 2, \cdots\}$ are random variables with mean value λ. Let X_k be the number of users (voice users and data users) at the beginning of the kth frame, then $X_k = N_v + N_d$. Let $B(X_k)$ be the number of users whose request are successfully transmitted in the kth frame. We now prove that $\{X_k, k = 0, 1, 2, \cdots\}$ is ergodic using Pake's lemma. We have

$$X_{k+1} = X_k + A_k - B(X_k).$$

So, for any i,

$$|E(X_{k+1} - X_k | X_k = i)| = |E(A_k - B(X_k) | X_k = i)| = |E(A) - E[B(i)]|$$

$$\leq \ |E(A)| + |E[B(i)]| \leq \lambda + M < \infty.$$

Hence, condition (a) of Pake's lemma is satisfied.

To satisfy condition (b) of Pake's lemma, we require that

$$\begin{aligned} & \limsup_{i\to\infty} \mathrm{E}(X_{k+1} - X_k | X_k = \mathrm{i}) \\ = \ & \limsup_{i\to\infty} E(A_k - B(X_k) | X_k = \mathrm{i}) \\ = \ & \limsup_{i\to\infty} (\lambda - E[B(i)]) < 0. \end{aligned}$$

So

$$\lambda \leq \liminf_{i \to \infty} \mathrm{E}[\mathrm{B}(i)] \tag{3.7}$$

is a sufficient condition for the system to be stable.

In our QoS algorithm, when there are N_v voice users and N_d data users, the total number of users is $i = N_v + N_d$. Then, 3 L such that for all $i \geq$ L, we have (see Appendix C)

$$T(N_v, N_d) \geq C\left(1 - \frac{C}{iM}\right)^{i-1}.$$

Hence,

$$B(i) \geq C\left(1 - \frac{C}{iM}\right)^{i-1}$$

Then

$$\liminf_{i \to \infty} E[B(i)] \geq \liminf_{i \to \infty} C\left(1 - \frac{C}{iM}\right)^{i-1} = Ce^{-\frac{C}{M}}. \tag{3.8}$$

Hence, from (3.8), $\lambda \leq Ce^{-C/M}$ is the sufficient condition for the system to be stable under the QoS requirement $\bar{P}_s \geq A_0$, where λ is the arrival rate. Note that in the special case $C =$ M; i.e., the system is designed to achieve the maximum achievable throughput, (3.8) becomes:

$$\liminf_{i \to \infty} E[B(i)] = \liminf_{i \to \infty} M\left(1 - \frac{1}{i}\right)^{i-1} = Me^{-1} \tag{3.9}$$

The sufficient stable condition is λ .< Me-', which is exactly the stable condition for slotted ALOHA. Furthermore, there is no bistable point in the system because the throughput does not decrease when the number of blocked users in the system increases.

3.4 Upper Bound on Throughput

We consider the QoS requirement as $\bar{P}_s \geq A_0$. With this restriction, we derive an upper-bound on the throughput for random access algorithms satisfying the following two assumptions. First, all users transmit in request slots randomly and independently. Second, each user transmits in at most one request slot in each frame. Let Ω be the set of all such random access algorithms.

We consider the throughput under two conditions. Condition 1: there is at least one voice user in the system. Condition *2:* there is no voice user in the system. First, we consider the throughput under Condition 1. Let X denote the total number of users that transmit in this frame, $0 \leq X \leq \infty$. The probability that the voice user successfully transmits its request in this frame is p.

$$p = \begin{cases} p_X & : \text{ if the user transmits in this frame,} \\ 0 & : \text{ otherwise,} \end{cases}$$

where

$$p_X := \left(1 - \frac{1}{M}\right)^{(X-1)}.$$

Note that

$$E(p_X) = E\left(\left(1 - \frac{1}{M}\right)^{(X-1)}\right) \geq E(p) = \bar{P}_s \geq A_0. \tag{3.10}$$

Let T_1 be the throughput given that there is at least one voice user in the system. Then,

$$T_1 = E\left(X\left(1 - \frac{1}{M}\right)^{(X-1)}\right). \tag{3.11}$$

We want to maximize *(3.11)* with the constraint *(3.10)*. Let $Y = (1 - 1/M)^{(X-1)}$. So

$$E(Y) \geq \bar{P}_s = A_0 = \left(1 - \frac{1}{M}\right)^{K_0 - 1}.$$

Let $f(y) = -y\left(\frac{\ln y}{\ln\left(1-\frac{1}{M}\right)} + 1\right)$, which is a strictly convex function. By Jensen's inequality [8],

$$T_1 = \mathrm{E}(-f(Y)) \leq -f(E(Y)) = K_0\left(1 - \frac{1}{M}\right)^{K_0 - 1} =: T_C. \tag{3.12}$$

Next, we consider the condition *2;* i.e., no voice user is in the system. Let T_0^α be the throughput of a random access algorithm a when there is no voice user in the system. Let $T_0^m = \max\{T_0^\alpha,\ a \in \Omega\}$. Let q_α denote the probability that no voice user is in the system of a random access algorithm a. Let $P_0 = \max\{q_\alpha,\ a \in \Omega\}$.

For algorithm a, let P_1^α be the probability that there is at least one voice user in the system. Hence, $1 - P_1^\alpha \leq P_0$. The throughput T of algorithm a is given by:

$$\begin{aligned} \mathrm{T} &= T_1 P_1^\alpha + T_0^\alpha (1 - P_1^\alpha) \leq T_C P_1^\alpha + T_0^m (1 - P_1^\alpha) = T_C + (1 - P_1^\alpha)(T_0^m - T_C) \\ &\leq T_C + P_0 (T_0^m - T_C) =: T_{max}. \end{aligned} \tag{3.13}$$

Therefore, T_{max} is the upper-bound on the throughput of random access algorithms in Ω (algorithms such that all users transmit for request slots randomly and independently, and each user transmits for at most one time slot in a frame). This upper-bound is not restricted to the (p_v, p_d) strategy used in this report.

The above upper-bound, T_{max}, may not tight. We compare T_1 with T_C. Since f is a strictly convex function, (3.12) achieves equality when $\mathrm{Y} = E(Y)$ with probability 1. Hence, the upper-bound T_C is only achievable if $\mathrm{X} = \mathrm{C}$ with probability 1; i.e., there are always exactly C users transmitting in each frame. However, C may not be an integer and it may not be possible to let exactly C users transmit in random access algorithms. So T_{max} may not a tight upper-bound. We try to approach the upper-bound by assigning p_v and p_d such that $E(N_v p_v + N_d p_d) = \mathrm{C}$ in our scheme.

Next, we show that

$$\lim_{M \to \infty} \frac{T(N_v, N_d)}{T_{max}} = 1,$$

when there are enough users in the system.

With some tedious algebra (Appendix D), we can show that

$$T(N_v, N_d) \geq T_C \left(1 - \frac{1}{M}\right),$$

when

$$N_v + N_d \geq C.$$

Recall that P_0 is the maximum probability that there is no data user in the system. Let p_0 be the probability that there is no new voice user with a request in a frame. Then, $P_0 = p_0 P(\text{all voice users with requests transmit successfully by the end of a frame in steady state})$. In practice, P_0 is small when M is large; i.e., in a large frame, it is unlikely there is no voice user in the frame. For example, if the arrival process

of voice users is a Poisson process with mean vM, then $P_0 \leq p_0 = e^{-vM}$. Suppose $P_0 \to 0$ as M $\to \infty$. We have

$$\frac{T(N_v, N_d)}{T_{max}} \geq \frac{T_C\left(1 - \frac{1}{M}\right)}{T_C + P_0(T_0^m - T_C)} \to 1.$$

Hence, the throughput of the presented QoS algorithm asymptotically approaches the upper-bound. In other words, when there is at least one voice user in the system, the throughput of our QoS algorithm approaches T_C. Furthermore, as M goes large, the probability that there is no voice user in the system goes to zero. So the throughput of our QoS algorithm asymptotically approaches the upper-bound.

4. QOS ALGORITHM WITH CAPTURE

In this section, we consider a system that exploit capture. We first describe the SIR capture model. Then we present the idea of distance-dependent permission probability. Users at different distance transmit with different probabilities to achieve a good throughput with the restriction of distance fairness. Finally, we describe the QoS algorithm with capture.

4.1 Capture model

We explain the capture model used in this report. Various capture models for mobile radio networks have been presented. In most models, the capture measure employed is the signal-to-interference power ratio (SIR) at the base station [9, 10, 11]. The receiver forms the ratio of the power of a reference packet to the total sum of all other colliding packets and thermal noise. The reference packet is assumed error-free and survives the collision if this ratio is higher than a predefined capture ratio. Another class of capture models is based on transmission reliability requirements. Two such capture models have been investigated in [12, 13, 14], where the capture measures are the target bit error rate and the correct reception of the packet header.

The capture model used in this report is based on SIR. With minor modifications, the presented algorithm can be used with other capture models. 'The propagation model considers the near/far effect and Rayleigh fading with log-normal shadowing. For simplicity, we consider a round cell with radius one instead of a hexagonal one and we assume that the base station is at the center of the cell. We also assume that users have omni-directional antennas and the frequency reuse distance is large enough so that the co-channel interference can be ignored.

The packet transmitted from a user to the base station experiences the near/far

effect and Rayleigh fading with lognormal shadowing. Furthermore, different users experience independent identical distributed (i.i.d.) shadowing and Rayleigh fading. Slow Rayleigh fading is assumed so that all the bits throughout a whole packet experience the same fading. Thus, the received power of a packet from user i is expressed as

$$y_i = a_i^2 s_i K r_i^{-\eta} P_T,$$

where a_i^2 is exponentially distributed, accounting for Rayleigh fading, $Kr_i^{-\eta}$ accounts for the power-loss law, r_i is the distance between user i and the base station, K is a constant, s_i is a random variable with a lognormal distribution, accounting for the shadowing, and P_T, the transmitted power, is assumed to be the same for all users. We assume that power control is not used as is typical in many practical TDMA systems.

We consider the probability that the packet from user 1 can be received successfully. Let R denote the predetermined SIR capture ratio. Given that I users transmit in the same time slot from distances $[r_1, \cdots, r_I]$ and they experience independent fading $[a_1^2, \ldots, a_I^2]$ and shadowing $[s_1, \cdots, s_I]$, the success probability of a packet from user 1 is

$$P\left\{\frac{a_1^2 s_1 K r_1^{-\eta} P_T}{\sum_{i=2}^{I} a_i^2 s_i K r_i^{-\eta} P_T + N_0} > R\right\},$$

where N_0 accounts for thermal noise. When two or more users transmit at the same time, the interference from other users is usually much larger than that of thermal noise. Thus, to focus on the nature of interference and to simplify calculation, thermal noise is neglected. By averaging over all Rayleigh fading random variables, the probability that user 1 is successfully received given $I-1$ interference packets is expressed as (see [11]):

$$P_I(r_1|s_1, \vec{r}, \vec{s}) = \prod_{i=2}^{I} \frac{1}{R\frac{s_i}{s_1}\left(\frac{r_i}{r_1}\right)^{-\eta} + 1}, \tag{4.1}$$

where $\vec{r}$ and $\vec{s}$ are $(I-1)$-component random vectors of the form $\vec{r} = (r_2, \cdots, r_I)^T$ and $\vec{s} = (s_2, \cdots, s_I)^T$, accounting for distance and shadowing of interference packets

respectively. From (4.1), we note that the success probability of user 1 is decreased by $1/(R\frac{s_i}{s_1}(\frac{r_i}{r_1})^{-\eta}+1$ times due to the transmission of user i (at distance r_i with shadowing s_i).

4.2 Distance-Dependent Permission Probability

In this section, we consider the problem that how n users should transmit in a frame with M request slots. The permission probability used to maximize the throughput in networks without capture does not maximize the throughput in wireless networks that exploit capture. Furthermore, in order to provide fairness to near and far users, permission probability should relate to a user's distance from the base station. In this report, we introduce a distance-dependent permission probability $q_M(n,r)$:

$q_M(n,r)$:= P(a user at distance r transmits in a request slot when there are n users in the system and there are M request slots in each frame).

The permission probability of a user depends on both the number of users in the system and the user's distance from the base station. Each user needs to detect its distance from the base station. Using distance measurement is reasonable, because in the future mobiles will likely be equipped with GPS device. Also, typically the distance measurement method is extendible to be based on path-loss and shadowing. Furthermore, we show that the scheme is robust to estimation errors of distances via numerical results.

Suppose that according to the signaling from the base station, n users are in the system. A user at distance r randomly selects one request slot, with probability $q_M(n,\mathrm{r})$, among M request slots and transmits in the selected request slot. All users select and transmit independently. Let's consider the success probability of user 1. When user 1 transmits in a request slot, user i at distance r_i transmits in the same time slot with probability $q_M(n,r_i)/M$. When user i transmits in the same request slot, it decreases the success probability of user 1 by $1/(R\frac{s_i}{s_1}(\frac{r_i}{r_1})^{-\eta}+1)$ times. With probability $1-q_M(n,r_i)/M$, user i does not transmit in the same time slot with user

1, which causes no interference to user 1. Hence, user i affects the success probability of user 1 with a factor v_i:

$$v_i = \begin{cases} \frac{1}{R\frac{s_i}{s_1}\left(\frac{r_i}{r_1}\right)^{-\eta}+1} & : \text{ with probability } \frac{q_M(n,r_i)}{M} \\ 1 & : \text{ with probability } 1 - \end{cases} \tag{4.2}$$

Given $\vec{r}$ and $\vec{s}$, the success probability of user 1 is

$$P(n, r_1|s_1, \vec{r}, \vec{s}) = q_M(n, r_1)E_{v_2} \cdots E_{v_I}(\prod_{i=2}^{I} v_i).$$

Since different users transmit independently,

$$P(n, r_1|s_1, \vec{r}, \vec{s}) = q_M(n, r_1)\prod_{i=2}^{I} E_{v_i}(v_i).$$

Denote

$$w_i = E_{v_i}(v_i) = \frac{1}{R\frac{s_i}{s_1}\left(\frac{r_i}{r_1}\right)^{-\eta}+1} \times \frac{q_M(n, r_i)}{M} + 1 \times \left(1 - \frac{q_M(n, r_i)}{M}\right)$$

Then

$$P(n, r_1|s_1, \vec{r}, \vec{s}) = q_M(n, r_1)\prod_{i=2}^{I} w_i.$$

Averaging over 7 and $\vec{s}$, we have

$$P(n, r_1|s_1) \quad = \quad q_M(n, r_1)E_{\vec{r}}E_{\vec{s}}\left(\prod_{i=2}^{I} w_i\right)$$

Denote

$$w = \frac{1}{R\frac{s}{s_1}\left(\frac{r}{r_1}\right)^{-\eta}+1}\frac{q_M(n, r)}{M} + \left(1 - \frac{q_M(n, r)}{M}\right),$$

where r is a random variable equivalent to r_i in distribution and so does s to s_i. Since all users experience i.i.d. shadowing and all users' distances from the base station are i.i.d. random variables, we have

$$P(n, r_1|s_1) = q_M(n, r_1)(E_r E_s(w))^{I-1}, \tag{4.3}$$

By averaging over the shadowing experienced by user 1, the probability of correct reception of a packet transmitted from a distance r_1 when there are n users in the system, is

$$P(n, r_1) = E_{s_1}(P(n, r_1|s_1)). \tag{4.4}$$

In this report, $P(n, \mathrm{r})$ is also called the individual throughput. Note that $P(n, \mathrm{r})$ is a function of the permission probability function $q_M(n, \mathrm{r})$.

Next, we discuss how to determine the distance-dependent permission probability and its importance. In the following discussion, we assume that new arrivals of users are uniformly distributed in the cell. We do not distinguish between newly-arrived users and retransmitted users. Hence, the distribution of users with requests is decided by the distribution of new arrivals and the distribution of retransmitted users.

In this report, we design the distance-dependent permission probability, $q_M(\mathrm{n}, \mathrm{r})$, such that the individual throughput is maximized with the restriction of distance fairness. Distance fairness means that users at different distance:; have the same individual throughput. The constrained optimization problem is expressed as

$$\begin{aligned} \underset{q_M}{\text{maximize}} \quad & P(n, \mathrm{r}), \\ \text{subject to} \quad & g(q) \leq 0, \end{aligned} \tag{4.5}$$

where $g(q) = \sup_{0 \leq r_i r^0 \gtrsim 1} |P(n, \mathrm{r}) - P(n, r_0)| - \epsilon$, and $\epsilon \geq 0$. When $\epsilon = 0$, absolutely fairness is required. In practice, a small unfairness is usually tolerable and ϵ is the measure of tolerance. Distance fairness is a good service quality and it justifies that users are uniformly distributed. Note that the distance distribution is required in the calculation of (4.3).

In [13], distance fairness is assumed to be obtained by power control. In other words, the near/far effect is compensated by maintaining an equal mean arrival power level, which is used to ensure an equal individual throughput among all users from different distances. In [10], the authors mention that users with lower received power

should have higher transmission probabilities to achieve fairness. In the simulations in [10], users at three different distances are assigned three different permission probabilities to achieve fairness. Furthermore, the authors propose a joint control strategy that adjust both received powers and transmission probabilities to achieve fairness and a good throughput. They assume that users at a lower received power level have no interference to the success probabilities of users at a higher received power level.

In this report, however, we assume no power control is used; i.e., all users transmit with the same power. There are several reasons for this assumption. First, power control complicates the system. Many typical TDMA systems do not implement power control. Second, we consider the contention phase of transmission, which is at the very beginning of each traffic burst, close-loop power control may not be available. Finally, power control may weaken the capture effects and decrease the system throughput. The effect of power control greatly depends on the SIR capture ratio, R, of the capture model. When $R > 1$, it is possible to capture a few packets at the same time. When $R < 1$, at most one packet can be captured. For example, in a CDMA system, processing gain is large, thus $R < 1$. If one user has a very large received power, the probability of other users being captured decrease dramatically. Hence, the use of power control to compensate the near/far effect and shadowing benefits the system throughput. However, in many TDMA systems, processing gain is small, which requires $R > 1$. Then, the larger the variance of the received power, the higher the probability that the power of one user is larger than R times the interference of all other users. An ideal condition in CDMA systems that all received powers are at the same level actually causes no capture in a system with a large value of R ($R > 1$). In this report, we consider the system with a small processing gain, thus a large value of R. In such a system, power control (used to compensate the near/far effect and shadowing) decreases the system throughput [13]. So fairness has to be maintained through other ways. In this report, the permission probability in (4.5) is used to achieve fairness with a good throughput.

We note that we need to know the location distribution of user:; to calculate the

unconditional capture probability in (4.3). In this report, we assume that users are uniformly distributed in the cell due to fair individual throughput at different distance. It is intuitively true that users are uniformly distributed in the cell if i) new arrivals are uniformly distributed, ii) individual throughput are distance independent, and iii) movements of users are random, independent of distance and omni-directional. The rigorous proof of this property requires complicated mathematics and it is not closely related to our work. Hence, we omit the proof.

In [9, 14], users' locations are assumed to follow a uniform distribution, or approximations of uniform distribution, for the convenience of calculation. With this assumption, distance-dependent success probabilities are calculated. Users closer to the base station have higher probability of success; i.e., higher individual throughput. However, if the individual throughput of users is a decreasing function of distance, then the distribution of users' locations is usually not uniform, further areas have higher density of users. To calculate the exact distribution is very hard.

In summary, distance fairness is a good quality of service and justifies the assumption of uniform distribution of users. Hence, in this report? we present the distance-dependent permission probability to achieve distance fairness with a good throughput. Next, we show some numerical results related to distance-dependent permission probability functions.

4.3 Numerical Results on Permission Probabilities

Since $q_M(n, \mathrm{r})$ has no close form solution, we use numerical results to show how the scheme works. In this subsection, we show the permission probabilities as a function of the number of users and the distances between users and the base station. We also show the throughput under the permission probability scheme. Furthermore, we test whether the scheme is robust to estimation errors of the number of users and the distances between users and the BS. We use the SIR capture model with the following set of parameters: $\eta = 4$ (path loss law parameter), $\mathrm{R} = 2$ (capture ratio), and 4dB shadowing. We also assume that user's distance from the base station is larger than $r_0 = 0.05$; i.e., about 0.2% of the area in the cell is prohibited. There are

two reasons. First, users are usually prohibited to be too close to the base station in practice. Second, if the user is too close to the base station, the propagation model is quiet different. For the convenience of calculation, we ignore this small area.

Figure 4.1 shows the distance-dependent permission probability function $q_1(\mathrm{n},\mathrm{r})$ for $2 \leq \mathrm{n} \leq 10$. These functions are the numerical solutions of (4.5) with $\epsilon = 0.02$ and $\mathrm{M} = 1$. Users at longer distance have larger permission probabilities. Figure 4.2 shows the individual throughput with $q_1(n,\mathrm{r})$, $2 \leq \mathrm{n} \leq 10$, in Figure 4.1. Users at different distances have fairly equal individual throughput.

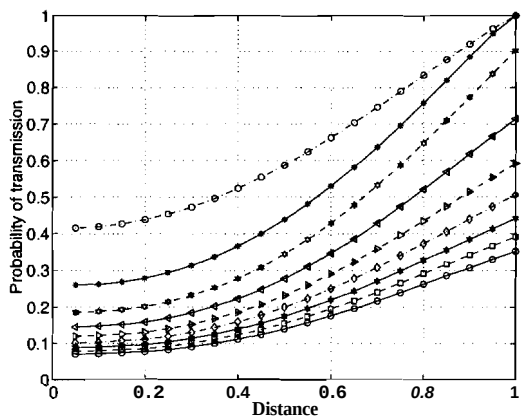

Fig. 4.1. Distance-dependent permission probability functions for one request slot. From the top to the bottom, they are $q_1(2,r)$, $q_1(3,\mathrm{r}), \ldots, q_1(10,\mathrm{r})$.

Figure 4.3 compares the throughput under the distance fairness constraint with three other cases. In case 1, users at different distances have the same permission probability $p_c(n)$; i.e., $q_1(n,r) = p_c(n)$, $0 \leq \mathrm{r} \leq 1$. With the assumption that users are uniformly distributed, $p_c(n)$ is used to maximize the throughput. When all users have the same $p_c(n)$, nearer users have higher probabilities of success. Hence, the density of users is higher at further areas. The throughput calculated with the uniform distribution assumption is an upper-bound on the actual throughput with uniform permission probability. We note that the throughput with distance fairness is only slightly less than this upper-bound. We should mention that the system is simpler if all users use the same permission probability. Hence, such a scheme

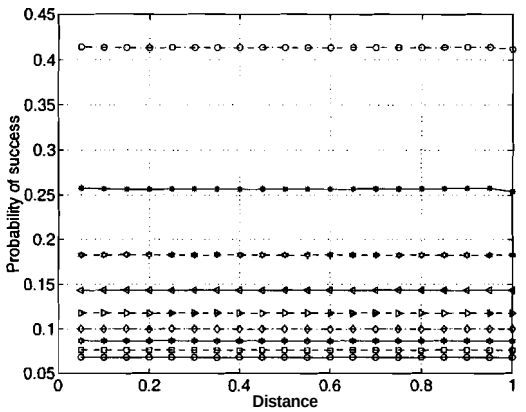

Fig. **4.2.** Individual throughput for one request slot. From the top to the bottom, they are $P(2,\mathrm{r}), P(3,\mathrm{r}), \cdots, P(10,\mathrm{r})$.

should be adopted when the simplicity of the system is very important. In case 2, $q_1(n,\mathrm{r})$ is used to maximize the throughput without the distance fairness constraint. In the calculation, we assume that users are uniformly distributed. The maximum throughput is achievable in some special cases. For example, all users move very fast and randomly. Once a user fails, it retransmits in next frame. At that time, the user locates anywhere in the cell with the same probability due to its fast and random movement. Hence, users are uniformly distributed in the cell. Users attribute fairness to their very fast and random movements and permission probabilities are only used to maximize the throughput. In case 3, no capture ability is considered. Obviously, throughput is much less without capture.

In the ideal condition, we assume that the base station knows the number of users, n, in the system. However, in practice, n has to be estimated. Figures **4.4** show the throughput with estimation errors. we assume n is the actual number of users in the system. Its estimate, v, given by the base station, is a binomial distributed random variable with parameter (p,N). The mean is $\mathrm{n} = \mathrm{Np}$, the variance is $Np(1-\mathrm{p})$, and the normalized variance is $\mathrm{e} = 1 - \mathrm{p}$. Thus, $q_1(\mathrm{v},\mathrm{r})$ is used in the system instead of $q_1(n,\mathrm{r})$. Figure **4.4** shows that the system is robust under estimation errors of the number of users.

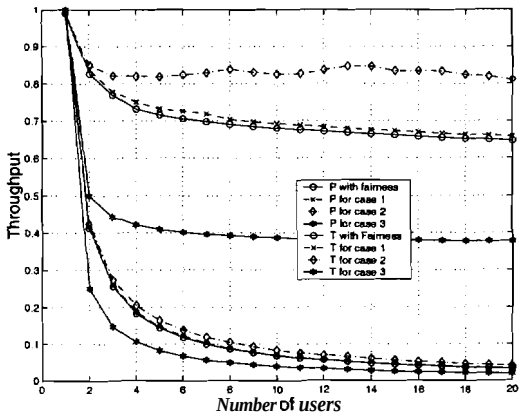

Fig. 4.3. Throughput comparison for one request slot. In the legend, P denotes the individual throughput and T denotes the overall throughput. Case 1: all users have the same permission probability. Case 2: maximum throughput without fairness constraint. Case 3: no capture.

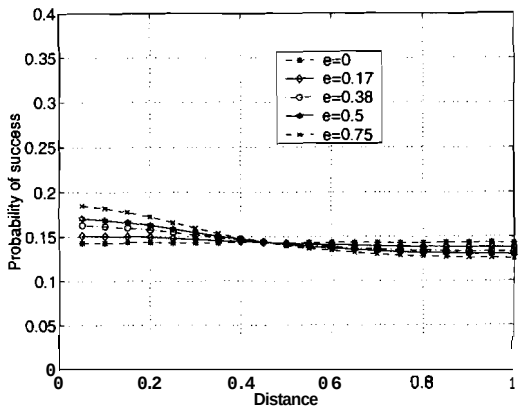

Fig. 4.4. Throughput with estimation errors of the number of users. In the legend, e indicates normalized variance; that is, e = variance/mean . For a binomial distribution, $e = 1 - p$.

At last, we show the system behavior with estimation errors of the distances between users and the base station. Assume r is the actual distance of a user from the base station. Its estimate, r', is a truncated Gaussian distributed random variable with mean r and variance $\sigma^2 = er$, where e is the normalized variance. Figure **4.5** shows that the system with estimation errors. The shape of the curve is due to the distribution of the estimation error. Since we use normalized variance, for very small r, the variance is small. Hence, the individual throughput does not change dramatically. For medium r, the permission probability in the neighborhood is almost linear. The trend to transmit with higher permission probability is compensated by the trend to transmit with lower permission probability. However, for large r, say r = 1, the variance is large and the user always transmit with lower permission probability because the estimate $r' \leq$ r with probability 1. Hence, the individual throughput drops most.

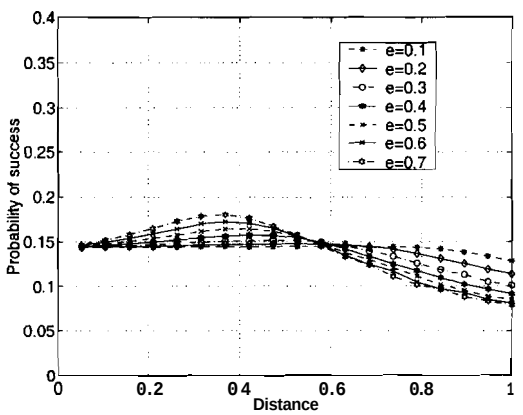

Fig. **4.5.** Throughput with estimation errors of the distance of users from the base station. In the legend, e indicates normalized variance; i.e., e =variance/mean.

4.4 QoS Algorithm with Capture

We consider an extreme case that $M = 1$ before we consider the general case that $M \geq 1$. When $M = 1$, users can retransmit in the next frame; i.e., the next request slot. At the beginning of each frame, the base station broadcasts the numbers of voice

users and data users in the system through a non-collision error-free signaling channel. After transmission in each request slot, users in the system can know immediately whether the transmission is successful. The successful user does not retransmit and unsuccessful users may retransmit in later frames. At a certain time, assume there are N_v voice users and N_d data users. Then for a voice user at distance r_i, its permission probability is

$$p_v(r_i) = q_1(\mathrm{Nu}, r_i).$$

For a data user at distance r_i, its permission probability is

$$p_d(r_i) = \begin{cases} q_1(N_d, r_i) & : \quad \text{if } N_v = 0 \\ 0 & : \quad \text{if } N_v > 0. \end{cases}$$

In this case, the delay of voice users is as small as possible. In other words, from the point of view of a voice user, there is no data users in the system in the ideal condition. (Ideal condition means that the base station knows exactly the numbers of users in the system, which is impractical. In practice, the numbers of voice users and data users are estimated. All estimation algorithms in literature can be implemented here.) Hence, in this case, whether the delay requirement of voice users can be satisfied is determined by the arrival process of voice users. The system always offers the best to voice users as it can. At the same time, the throughput of the system is also maximized. The scheme is actually a pure priority scheme. The system always serves users with the highest priority currently in the system.

We, next, consider the general case that $\mathrm{M} \geq 1$. At the beginning of each frame, the base station broadcasts the numbers of voice users and data users in the system through a non-collision error-free signaling channel. No user can retransmit in the same frame. Each user randomly selects a request slot and transmits in it with a certain class-dependent distance-dependent permission probability. There exists the tradeoff between the throughput of the system and the delay performance of voice users. Let N_M denote the numbers of users transmitting with the maximum throughput, T_M be the maximum throughput of one frame. If $N_v < N_M$, data users

should transmit with certain probability of maximize the throughput. However, the delay performance of voice users suffers.

To obtain the delay requirement of voice users, we set a threshold C. Let N_v and N_d be the numbers of voice users and data users. Let $K = \max(\min(C, N_v + N_d), N_v)$. The permission probabilities of voice users and data users are:

$$\begin{aligned} p_v(r_i) &= q_M(K, r_i) \\ p_d(r_i) &= \frac{K - N_v}{N_d} q_M(K, r_i). \end{aligned} \tag{4.6}$$

The above algorithm is quiet similar to the formula (3.2). The difference is that the permission probabilities are also functions of the distance in (4.6). In this scheme, there is no simple way to calculate the threshold C. We can adopt the theoretical method (Appendix B), which involves a large amount of computation because of capture. Otherwise, we estimate a suitable value of C via simulations.

5. SIMULATION RESULTS

In this section, we provide simulation results of the studied algorithms. We assume in the simulation that the arrival processes of voice users and data users are independent Poisson processes.

Figure **5.1** indicates the delay distribution of a voice user. We can see that the delay distribution of a voice user is well approximated by a geometric distribution.

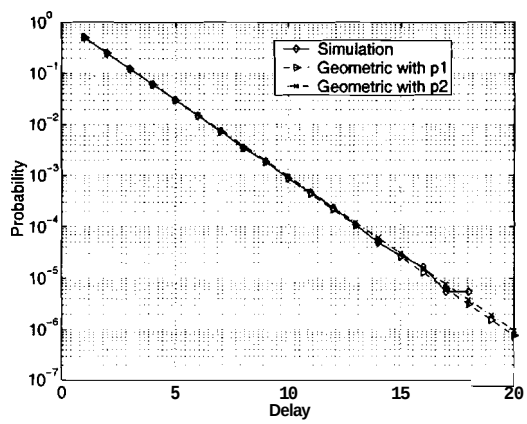

Fig. **5.1.** Delay distribution of a voice user when $M = 20$, **pl** is the reciprocal of the average delay of voice users, and $p2$ is the average probability of success of voice users.

Figures 5.2 and **5.3** illustrate the performance of the QoS algorithm without capture. In the simulation, the fraction of voice users is **50%.** Figure 5.2 indicates the delay performance of voice users. The delay performance is in terms of the average probability of success. Simulations are run under both the ideal condition and the practical condition. By the ideal condition, we mean that the base station knows the exact numbers of voice and data users in the system. In practice, a Kalman filter is used to estimate the numbers of users. The Kalman filter approach is implemented

with two threshold values. We use (3.3) to approximate C. In the ideal condition, (3.3) offers a pretty good approximation. With $C = K_0$ in the Kalman filter approach, $\bar{P}_s$ is less than the QoS requirement because of estimation errors. Thus, in practice, we should use a smaller threshold value than the one calculated under the ideal condition, which is represented by the curve with $C = 0.9K_0$. Figure 5.3 shows the throughput performance. We compare the throughput in the ideal condition with the practical approaches. As expected, the Kalman filter approach with the smaller C has less throughput, illustrating the tradeoff between the throughput and QoS. We use the probability of no new voice user in a frame, $p_0 = e^{-r\lambda}$, as the upper-bound of the probability of no voice user in a frame, P_0. Hence, p_0 is used in (3.13) to calculate the upper-bound of throughput, which is also shown in Figure 5.3.

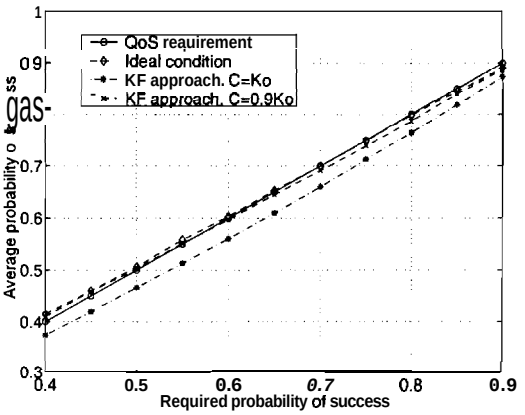

Fig. 5.2. Delay performance without capture for M = 20 with 50% voice users. In the legend, KF denotes Kalman filter.

Next, we show the simulation results of the QoS algorithm with capture. We still assume in the simulation that the arrival processes of voice users and data users are independent Poisson processes. During simulations, we use the SIR model with the same parameter set as the one used for numerical results in last section: $\eta = 4$, R = 2(capture ratio), and 4dB shadowing. Each simulation run 100000 times. The ratio of voice traffic is 0.5 in all simulations. We estimate the numbers of voice users and data users in the simulation, which is not the ideal condition. However, we

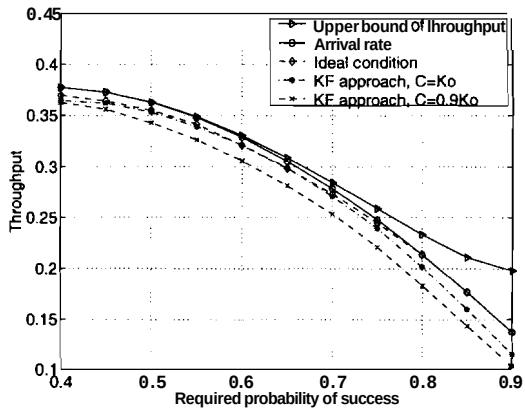

Fig. 5.3. Throughput without capture for $M = 20$ with 50% voice users. In the legend, KF denotes Kalman filter.

assume that each user knows its exact distance from the base station.

Figure 5.4 shows the simulation result when users can retransmit in the next request slot; i.e., $M = 1$. Since the number of users is estimated, there exists a little unfairness between near and far users. However, this unfairness is small. With better estimation algorithms, we expect lower delay and better fairness. Note the unit of delay is a frame with $M = 1$.

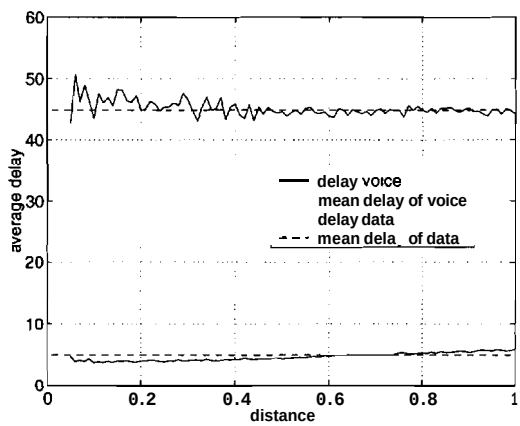

Fig. 5.4. Average delay of voice users and data users:.

Figure 5.5 compares the delay of voice users when $C = 3$ and $C = \infty$. In the

simulation, we set the buffer number of data users 60. When $C = 3$, data users yield voice users and the maximal throughput is not obtained. When $C = \infty$, voice users and data users are treated the same and the maximal throughput is obtained. Figure 5.5 shows the tradeoff between the delay performance of voice users and the total throughput of the system. The smaller the threshold, the less the delay of voice users, and the less the overall throughput. When $C = 3$, the average delay of voice users is about 0.5 frame less than that of $C = \infty$. However, the percentage of the buffer of data users overflow is 1% while it is 0 when $C = \infty$. When $C = \infty$, the data delay is the same as the voice delay, so it is omitted in this figure.

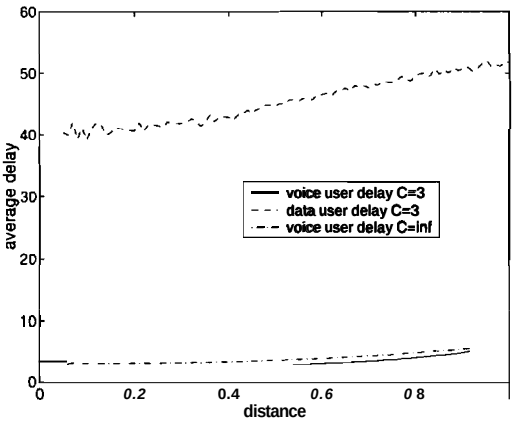

Fig. 5.5. Compare delay performance when $C = 3$ and $C = \infty$. When $C = \infty$, the data delay is the same as the voice delay, so it is omitted in this figure.

6. CONCLUSIONS

There are mainly two contributions in this report. First, we present a random access scheme that provides certain QoS guarantees during the contention phase of communication. Permission probabilities are used to provide QoS for two traffic classes, voice users and data users. The same idea can be extended to multi-class users. The QoS requirement of voice users is defined as $\bar{P}_s$, the average success probability of voice users. For a predetermined QoS measure $\bar{P}_s$, a threshold C is calculated such that a voice users has an average success probability larger or equal to $\bar{P}_s$. We prove that the algorithm is stable with a weak assumption. We derive the upper-bound of a general class of random access algorithms under the QoS requirement in term of $\bar{P}_s$ and show that the studied algorithm asymptotically approaches the upper-bound. The analysis is based on the QoS algorithm without capture and we consider systems with capture in Section 4. The QoS algorithms with and without capture are the same in essence except that the individual throughput is higher when capture is considered.

Second, we introduce a distance-dependent permission probability scheme with the SIR capture model in this report. With the assumption of uniform arrival distribution and random movement, we show that our algorithm provide distance fairness with a good throughput and it justifies the uniform distribution of users. Furthermore, permission probabilities are used to provide the QoS requirement in terms of delay and distance fairness.

In wireless networks, providing QoS during contention phase is important to support bursty traffic. It is quite different from the wire-line scenario. So existing methods such as using in ATM do not apply directly. There would be large research space for this topic.

APPENDIX

APPENDIX

A.1 Appendix A

Let $T_k = k\left(1 - \frac{1}{M}\right)^{k-1}$. Then $T_M = \max(T_k)$; i.e., $T_M \geq T_k$ for all k.

Proof: we prove that $T_M/T_k \geq 1$ for all k.

1. When $k < \mathrm{M}$,

$$\frac{T_k}{T_{k-1}} = \frac{k}{k-1}\left(1 - \frac{1}{M}\right) = \frac{kM - k}{kM - M} \geq 1.$$

So

$$\frac{T_M}{T_k} = \frac{T_M}{T_{M-1}}\frac{T_{M-1}}{T_{M-2}} \cdots \frac{T_{k+1}}{T_k} \geq 1.$$

2. When $k > \mathrm{M}$,

$$\frac{T_{k+1}}{T_k} = \frac{k}{(k+1)\left(1 - \frac{1}{M}\right)} = \frac{kM}{kM + M - k - 1} \geq 1.$$

So

$$\frac{T_M}{T_k} = \frac{T_M}{T_{M+1}}\frac{T_{M+1}}{T_{M+2}} \cdots \frac{T_{k-1}}{T_k} \geq 1.$$

3. When $k = \mathrm{M}$, it is trivial.

Hence, $T_M/T_k \geq 1$ for all k. □

A.2 Appendix B

We now explain how to determine C. A two dimensional Markov chain is used to calculate the steady-state distribution. Suppose that we know the distribution of the arrival process. Given C = x, transmission probabilities between states are determined by (3.2) and the arrival process. Hence, $\pi(i,k)$ can be calculated and so can $\bar{P}_s(x)$. Since $\bar{P}_s(x)$ is a monotone decreasing function of x and $0 \leq \mathrm{C} \leq \mathrm{M}$, the parameter C is the unique solution of $\bar{P}_s(x) = A_0$, which can be obtained easily

using a standard zero-finding algorithm such as Newton's method. If $\bar{P}_s(0) < A_0$, the QoS requirement cannot be satisfied. In other words, even without data users, the delay caused by the contention among voice users are still larger than required if $\bar{P}_s(0) < A_0$.

A.3 Appendix C

When the scheme in (3.2) is used, $T(N_v, N_d)$ is the throughput given by (3.4), (3.5), and (3.6). Suppose C $\leq$ M. We need to prove that

$$T(N_v, N_d) \geq C\left(1 - \frac{C}{iM}\right)^{i-1},$$

where i $= N_v + N_d$ and $i \geq$ C.

Proof:

We first prove that $f_k = (1 - \frac{1}{k})^{k-1}$ is a monotonically decreasing function of k and $g_k = (1 - \frac{1}{k})^k$ is a monotonically increasing function of k

First we prove that $\ln(1 - x) + x \leq 0$ for $0 < x < 1$. Because $\ln(1 - 0) - 0 = 0$ and

$$\frac{d(\ln(1-x) + x)}{dx} = -\frac{1}{1-x} + 1 < 0,$$

so $\ln(1 - x) - x \leq 0$ for $0 < x < 1$. Then,

$$\begin{aligned}
\frac{df_k}{dk} &= (1 - \tfrac{1}{k})^{k-1}\left((k-1)'\ln(1 - \tfrac{1}{k}) + \frac{(1 - \frac{1}{k})'}{(1 - \frac{1}{k})}(k-1)\right) \\
&= (1 - \tfrac{1}{k})^{k-1}\left(\ln(1 - \tfrac{1}{k}) + \frac{\frac{1}{k^2}}{(1 - \frac{1}{k})}(k-1)\right) \\
&= (1 - \tfrac{1}{k})^{k-1}\left(\ln(1 - \tfrac{1}{k}) + \frac{1}{k}\right) \\
&\leq 0.
\end{aligned}$$

Hence, f_k is a a monotonically decreasing function of k.

We prove that $g_k = (1 - \frac{1}{k})^k$ is a monotonically increasing function of k similarly. First we prove $\ln(1 - \frac{1}{k}) + \frac{1}{k-1} \geq 0$ for $k > 1$.

$$\left(\ln(1 - \tfrac{1}{k}) + \frac{1}{k-1}\right)' = \frac{1}{k(k-1)} - \frac{1}{(k-1)^2} \leq 0.$$

Since $\lim_{k\to\infty} \ln\left(1-\frac{1}{k}\right)+\frac{1}{k-1}=0$, so

$$\ln\left(1-\tfrac{1}{k}\right)+\frac{1}{k-1}\geq 0.$$

Let $g(k)=(1-\frac{1}{k})^{k}$. Then,

$$\begin{aligned}\frac{dg_k}{dk} &= \left(1-\tfrac{1}{k}\right)^k\left(\ln\left(1-\tfrac{1}{k}\right)+\frac{\frac{1}{k^2}}{\left(1-\frac{1}{k}\right)}k\right)\\ &= \left(1-\tfrac{1}{k}\right)^k\left(\ln\left(1-\tfrac{1}{k}\right)+\frac{1}{k-1}\right)\geq 0\end{aligned}$$

So $g(k)=(1-\frac{1}{k})^{k}$ is a monotonically increasing function of k.

Furthermore.

$$\lim_{k\to\infty}\left(1-\frac{1}{k}\right)^k=\lim_{k\to\infty}\left(1-\frac{1}{k}\right)^{k-1}=e^{-1},$$

$$\left(1-\frac{1}{k}\right)^k\leq e^{-1}\leq\left(1-\frac{1}{k}\right)^{k-1}.$$

Next, we prove that

$$T(N_v,N_d)\geq C\left(1-\frac{C}{iM}\right)^{i-1},$$

where $i=N_v+N_d$. Denote $F_C(i)=C\,(1-\frac{C}{iM})^{i-1}$.

1. $N_v>M$,

$$T(N_v,N_d)=M\left(1-\frac{1}{N_v})^{(N_v-1)}\right)\geq Me^{-1}$$

If $C=M$, $F_C(i)=M(1-1/i)^{i-1}$. Since $i\geq N_v$, as we proved above,

$$\left(1-\frac{1}{i}\right)^{i-1}\leq\left(1-\frac{1}{N_v}\right)^{(N_v-1)}.$$

So $T(N_v,N_d)\geq F_C(i)$.

If $C<M$, we have

$$\begin{aligned}\frac{dF_C(i)}{dc} &= \left(1-\frac{C}{iM}\right)^{i-1}-\frac{1}{Mi}C\left(1-\frac{C}{iM}\right)^{i-1}\\ &= \left(1-\frac{C}{Mi}\right)\left(1-\frac{C}{iM}\right)^{i-1}\\ &\geq 0\end{aligned}\tag{A.1}$$

So $F_C(i) \leq F_M(i)$. Since $T(N_v, N_d) \geq F_M(i)$, we have $T(N_v, N_d) \geq F_C(i)$ for $C \leq M$. So when $N_v > M$, $T(N_v, N_d) \geq F_C(i)$.

2. When $C \leq N_v \leq M$,

$$T(N_v, N_d) = N_v p_v \left(1 - \frac{p_v}{M}\right)^{N_v - 1}.$$

We first prove that $T(C, N_d) \geq F_C(i)$. We note $i \geq C$. We need to prove

$$T(C, N_d) = C\left(1 - \tfrac{1}{M}\right)^{C-1} \geq F_C(i) = C\left(1 - \frac{C}{iM}\right)^{i-1}$$

i.e., we need to prove

$$\left(1 - \tfrac{1}{M}\right)^{C-1} \geq \left(1 - \frac{C}{iM}\right)^{i-1}$$

i.e..

$$\left(1 - \tfrac{1}{M}\right)^{(M-1)\frac{C-1}{M-1}} \geq \left(1 - \frac{C}{iM}\right)^{\left(\frac{iM}{C}-1\right)\frac{i-1}{\frac{iM}{C}-1}}.$$

Since $M \leq \frac{iM}{C}$, as proved above

$$\left(1 - \tfrac{1}{M}\right)^{(M-1)} \geq \left(1 - \frac{C}{iM}\right)^{\left(\frac{iM}{C}-1\right)}.$$

So we only need to prove

$$\frac{C-1}{M-1} \leq \frac{i-1}{\frac{iM}{C}-1}.$$

since $(C-1)(\frac{iM}{C}-1) = iM - C - \frac{iM}{C} + 1$ and $(M-1)(i-1) = iM - M - i + 1$,

$$(M-1)(i-1) - (C-1)(\tfrac{iM}{C}-1) = \tfrac{iM}{C} + C - i - M = i\left(\frac{M}{C} - 1)\right) - C\left(\frac{M}{C} - 1)\right) \geq 0.$$

So

$$\frac{C-1}{M-1} \leq \frac{i-1}{\frac{iM}{C}-1}.$$

Hence,

$$\left(1 - \tfrac{1}{M}\right)^{(M-1)\frac{C-1}{M-1}} \geq \left(1 - \frac{C}{iM}\right)^{\left(\frac{iM}{C}-1\right)\frac{i-1}{\frac{iM}{C}-1}}.$$

We have proved that $T(C, N_d) = C\left(1 - \frac{1}{M}\right)^{C-1} \geq F_C(i) = C\,(1 - \frac{C}{iM})^{i-1}$

As showed in Appendix A, $T(N_v, N_d)$ is an increasing function for $C \leq N_v \leq M$, we have proved that $T(N_v, N_d) > F_C(i)$ for $C \leq N_v \leq M$.

3. $0 \le N_v < C$,

Suppose there are x voice users and $i - x$ data users, they transmit according to (3.2), the throughput is

$$\begin{aligned} T(N_v, N_d) &= N_v \left(1 - \frac{1}{M}\right)^{N_v - 1} \left(1 - \frac{p_d}{M}\right)^{N_d} \\ &\quad + (C - N_v) \left(1 - \frac{1}{M}\right)^{N_v} \left(1 - \frac{p_d}{M}\right)^{N_d - 1}. \end{aligned}$$

Denote

$$F_i(x) = \begin{cases} T(x, i - x) & : \quad x > 0 \\ C\left(1 - \frac{C}{iM}\right)^{i-1} & : \quad x = 0 \end{cases}$$

So $F_i(x) \le T(x, i - x)$.

So we only need to prove that

$$F_i(x) \ge F_C(i).$$

Since $F_i(0) = F_C(i)$, and $F_i(x)$ is a continues function of x, so we only need to prove that $dF_i(x)/dx \ge 0$.

$$\begin{aligned} F_i(x) &= x\left(1 - \tfrac{1}{M}\right)^{x-1} \left(1 - \frac{p_d}{M}\right)^{(i-x)} + (C - x)\left(1 - \tfrac{1}{M}\right)^{x} \left(1 - \frac{p_d}{M}\right)^{(i-x)-1} \\ &= \underbrace{\left(1 - \tfrac{1}{M}\right)^{x-1}}_{a_1} \underbrace{\left(1 - \tfrac{p_d}{M}\right)^{i-x-1}}_{a_2} \underbrace{\left[x\left(1 - \tfrac{p_d}{M}\right) + (c - x)\left(1 - \tfrac{1}{M}\right)\right]}_{a_3}, \\ \frac{dF_i(x)}{dx} &= a_1' a_2 a_3 + a_1 a_2' a_3 + a_1 a_2 a_3', \end{aligned}$$

where $p_d = \frac{C-x}{i-x}$ and i > C. Note

$$\begin{aligned} \frac{dp_d}{dx} &= \left(\frac{C - x}{i - x}\right)' = \frac{C - i}{(i - x)^2} < 0, \\ a_2' &= \frac{d}{dx}\left[\left(1 - \tfrac{p_d}{M}\right)^{i-x-1}\right] \\ &= \left(1 - \tfrac{p_d}{M}\right)^{i-x-1} \left(-\ln\left(1 - \tfrac{p_d}{M}\right) - \frac{1}{M}\frac{C - i}{(i - x)^2}(i - x - 1)\right), \\ a_1' &= a_1 \ln\left(1 - \tfrac{1}{M}\right). \end{aligned}$$

Then

$$\begin{aligned}
\frac{dF_i(x)}{dx} &= a_1 \ln\left(1-\tfrac{1}{M}\right)a_2a_3 \\
&\quad + a_1a_2\left(-\ln\left(1-\tfrac{p_d}{M}\right) - \frac{1}{M}\frac{C-i}{(i-x)^2}(i-x-1)\right)a_3 \\
&\quad + a_1a_2\left[\left(1-\tfrac{p_d}{M}\right) + \frac{i-C}{M(i-x)^2} - \left(1-\tfrac{1}{M}\right)\right] \\
&= a_1a_2\left\{\ln\left(1-\tfrac{1}{M}\right)a_3 + \left(-\ln\left(1-\tfrac{p_d}{M}\right) - \frac{1}{M}\frac{C-i}{(i-x)^2}(i-x-1)\right)a_3\right. \\
&\quad \left. + \left(1-\tfrac{p_d}{M}\right) + \frac{i-C}{M(i-x)^2} - \left(1-\tfrac{1}{M}\right)\right\}.
\end{aligned}$$

Since a_1 and a_2 are positive, we only need to proof that the content in the bracket are positive. Since a_3 are positive and the only negative term in the bracket is $\ln\left(1-\frac{1}{M}\right)a_3$, we only need to proof that

$$\frac{1}{M}\frac{C-i}{(i-x)^2}(i-x-1)a_3 \geq \ln\left(1-\tfrac{1}{M}\right)a_3;$$

i.e.,

$$\frac{1}{M}\frac{C-i}{(i-x)^2}(i-x-1) \geq \ln\left(1-\tfrac{1}{M}\right).$$

Since $\ln(1-x) < -x$ for $0 < x < 1$, we only need to prove that

$$-\frac{C-i}{(i-x)^2}(i-x-1) \leq 1;$$

i.e.,

$$(i-C)(i-x-1) - (i-x)^2 \leq 0.$$

We have $x < C$ and $i > C$, so

$$\begin{aligned}
(i-C)(i-x-1) - (i-x)^2 &= i^2 - Ci - ix + Cx - i + C - i^2 + 2ix - x^2 \\
&= (i-x)(x-C) - (i-C) \\
&\leq 0 \qquad \text{(A.2)}
\end{aligned}$$

Hence,

$$\frac{1}{M}\frac{C-i}{(i-x)^2}(i-x-1) \geq \ln\left(1-\tfrac{1}{M}\right)$$

So far, we have proved that $dF_i(x)/dx$ is positive. Since $F_i(0) = F_C(i)$, for $x \geq 0$ we have $F_i(x) > F_C(i)$ for $0 \leq x < C$.

Hence, $T(N_v, N_d) \geq C\,(1 - \frac{C}{iM})^{i-1}$, where $i = N_v + N_d \geq C$ and $C \leq M$. □

A.4 Appendix D

Suppose $i = N_v + N_d \geq C$ and $C \leq M$, then $T(N_v, N_d) \geq T_C\big(1 - \frac{1}{M}\big)$. Proof we have proved in Appendix C that

$$T(N_v, N_d) \geq C\left(1 - \frac{C}{iM}\right)^{i-1},$$

where $C \leq M$ and $i \geq C$. Next we prove that $F_C(i)$ is a decreasing function of i.

$$\begin{aligned}\frac{dF_C(i)}{di} &= F_C(i)\left(\ln\left(1 - \frac{C}{iM}\right) + (i-1)\frac{C}{i^2M - iC}\right)\\ &= F_C(i)\left(\ln\left(1 - \frac{C}{iM}\right) + \frac{C}{iM} - \frac{C}{iM} + (i-1)\frac{C}{i^2M - iC}\right)\end{aligned}$$

Since $\ln\big(1 - \frac{C}{iM}\big) + \frac{C}{iM} \leq 0$, we only need to prove $-\frac{C}{iM} + (i-1)\frac{C}{i^2M - iC} \leq 0$ to show that $F_C(i)$ is a decreasing function. Since

$$-\frac{C}{iM} + (i-1)\frac{C}{i^2M - iC} = \frac{C}{i}\left(\frac{i-1}{M(i-1)} - \frac{i-1}{iM - C}\right) \leq 0,$$

So $F_C(i)$ is a decreasing function. Hence,

$$T(N_v, N_d) \geq F_C(i) \geq \lim_{i\to\infty} F_C(i) = Ce^{-\frac{C}{M}}.$$

So

$$T(N_v, N_d) \geq Ce^{-\frac{C}{M}} \geq C\big(1 - \tfrac{1}{M}\big)^{M\frac{C}{M}} = T_C\big(1 - \tfrac{1}{M}\big).$$

□

LIST OF REFERENCES

[1] C. Bisdikian, "A review of random access algorithms," IBM Res. Rep., Jan. 1996.

[2] 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, no. 8, pp. 885–890, 1989.

[3] W. S. Jeon, D. G. Jeong, and C.-H. Choi, "An integrated service MAC protocol for local wireless communications," IEEE Trans. Veh. Technol., vol. 47, no. 1, pp. 352–363, 1998.

[4] R. LaMaire, A. Krishna, and H. Ahmadi, "Analysis of a wireless MAC protocol with client-server traffic and capture," IEEE J. *Sel.* Areas Commun., vol. 12, no. 8, pp. 1299–1313, 1994.

[5] H. Peyravi, "Medium access control protocols performance in satellite communications," IEEE Communications Magazine, vol. 37, no. 3, pp. 62–71, 1999.

[6] K. Mori and K. Ogura, "An adaptive permission probability control method for integrated voice/data CDMA packet communications," *IEICE* Trans. Fundamentals, vol. E81-A, no. 7, pp. 1339–1348, 1998.

[7] R. Rom and M. Sidi, Multiple access protocols : performance *and* analysis. New York : Springer-Verlag, 1990.

[8] P. Billingsley, Probability and Measure. Wiley, 1985.

[9] J.-P. M. Linnartz, R. Hekmat, and R.-J. Venema, "Near-far effects in land mobile random access networks with narrow-band Rayleigh fading channels," IEEE Trans. Veh. Technol., vol. 41, no. 1, pp. 77–89, 1992.

[10] B. Wu and Q. Wang, "Maximization of the channel utilization in wireless heterogeneous multiaccess networks," IEEE Trans. Veh. Technol'., vol. 46, no. 2, pp. 437–444, 1997.

[11] M. Zorzi, "Mobile radio slotted ALOHA with capture, diversity and retransmission control in the presence of shadowing," *ACM/Baltzer* Wireless Networks, vol. 4, pp. 379–388, 1998.

[12] I. M. I. Habbab, M. Kavehrad, and C. E. W. Sundberg, "ALOHA with caputure over slow and fast fading radio channels with coding and diversity," IEEE J. *Sel.* Areas Commun., vol. 7, no. 1, pp. 79–88, 1989.

[13] X. Qiu and V. O. K. Li, "On the capacity of packet reservation multiple access with capture in personal communication systems," IEEE Trans. Veh. Technol., vol. 45, no. 4, pp. 666–675, 1996.

[14] H. Zhou and R. H. Deng, "Capture model for mobile radio slotted ALOHA systems," IEE Proceedings, vol. 145, pp. 91–97, April 1998.