A Multiqueue Service Room MAC Protocol for Wireless Networks With Multipacket Reception

Qing Zhao, *Member, IEEE,* and Lang Tong, *Senior Member, IEEE*

***Abstract*—An adaptive medium-access control (MAC) protocol for heterogeneous networks with finite population is proposed. Referred to as the multiqueue service room (MQSR) protocol, this scheme is capable of handling users with different quality-of-service (QoS) constraints. By exploiting the multipacket reception (MPR) capability, the MQSR protocol adaptively grants access to the MPR channel to a number of users such that the expected number of successfully received packets is maximized in each slot. The optimal access protocol avoids unnecessary empty slots for light traffic and excessive collisions for heavy traffic. It has superior throughput and delay performance as compared to, for example, the slotted ALOHA with the optimal retransmission probability. This protocol can be applied to random-access networks with multimedia traffic.**

***Index Terms*—Medium-access control (MAC), multipacket reception (MPR), random-access network.**

I. Introduction

IN MULTIACCESS wireless networks where a common channel is shared by a population of users, a key issue, referred to as medium-access control (MAC), is to coordinate the transmissions of all users so that the common channel is efficiently utilized and the quality-of-service (QOS) requirement of each user is satisfied. The schemes for coordinating transmissions among all users are called MAC protocols.

The conventional assumption on the channel is that any concurrent transmission of two or more packets results in the destruction of all the transmitted information. Based on this assumption, numerous MAC protocols, such as ALOHA [1], [21], the tree algorithm [6], the first-come-first-serve (FCFS) algorithm [10], the window random-access algorithm [20], and a class of adaptive schemes [5], [14], [15],[17], have been proposed. However, with the development of spread spectrum, space-time coding, and new signal processing techniques, this collision channel model does not hold in many important practical communication systems where one or more packets can be successfully received in the presence of other simultaneous transmissions. For instance, the capture phenomenon is common in local area radio networks. Other examples include networks using code-division multiple-access (CDMA) and/or antenna array, multiuser detection techniques, and signal-processing-based collision resolution algorithms [26].

This new channel model which offers the capability of multipacket reception (MPR) presents new challenges for medium-access control in wireless networks. As a commonly seen form of MPR, the capture effect first drew the attention of researchers. The impact of capture on the performance of slotted ALOHA is studied in [2], [9], [13], [19], [24], [27], [28], and references therein. The performance of the FCFS algorithm in channels with capture is analyzed in [23]. In [3], [18], and [25], the window random-access protocol [20] is extended to networks with capture and its performance is evaluated. A hybrid protocol which employs slotted ALOHA and the busy-tone sensing scheme is studied in correlated Rayleigh fading channels with capture in [8].

MPR provided by multiple independent collision channels is studied in [7] and [16], where the contention-free scheme TDMA is extended to a fully connected half-duplex *ad hoc* network. In [22], the authors introduce dynamic slot allocation for cellular systems with antenna arrays. Given a set of active users (users with packets to transmit), the proposed dynamic slot allocation scheme assigns an appropriate number of active users to each time slot to utilize the MPR capability provided by the antenna array. In [11] and [12], a general model for channels with MPR capability is developed. This model can be applied to systems with capture, CDMA, and space-division-multiple-access (SDMA). Under this model, the performance of slotted ALOHA in networks with infinite population is analyzed in [11] and [12].

The above-mentioned studies mainly focus on the impact of MPR on the performance of existing MAC protocols which were originally proposed for the conventional collision channel. The problem of designing random-access protocols explicitly based on a general MPR channel model has rarely been touched. Nevertheless, fully utilizing the MPR capability is a nontrivial problem that calls for further studies. First, MPR provides a new approach to collision resolution. Historically, collision resolution is primarily based on the principle of limiting transmissions in the event of failures. For channels with MPR, this strategy should be reexamined. Consider a channel in which, when there are two simultaneous transmissions, it is highly likely that both transmissions are successful. In the unlikely event of failed transmission, the protocol may want to enable both users to retransmit rather than limit their transmissions using splitting or random backoff. Second, MPR enriches the channel outcome, which makes it more difficult to infer the state of a user from the feedback information. For the conventional channel, a successful reception implies that one and only one user has transmitted; all other users who

Manuscript received January 30, 2001; revised October 28, 2001; approved by IEEE/ACM Transactions on Networking Editor T. Todd. This work was supported in part by the Multidisciplinary University Research Initiative (MURI) under the Office of Naval Research Contract N00014-00-1-0564 and by the Army Research Office under Grant ARO-DAAB19-00-1-0507.

The authors are with the School of Electrical and Computer Engineering, Cornell University, Ithaca, NY 14853 USA (e-mail: ltong@ee.cornell.edu).

Digital Object Identifier 10.1109/TNET.2002.808403

are enabled in the same time slot are idle. For MPR channels, however, a packet can be successfully received in the presence of many simultaneous transmissions. Sophisticated state estimation techniques are required for an efficient utilization of the MPR capability.

In this paper, we propose a MAC protocol designed explicitly for MPR channels. A slotted network with a finite population of users is considered. Users may have different QoS requirements which are characterized by their average packet delay at the heaviest traffic load. Since, in general, packet delay increases with the traffic load, this delay constraint specifies the worst case performance of the network. The proposed protocol maximizes the per-slot throughput (the expected number of successfully received packets in each slot) while ensuring each user's QoS requirement. The key to maximizing per-slot throughput is an optimal estimate of the state of users. By fully exploiting the information provided by previous channel outcomes, the state of each user is updated at the beginning of each slot. Based on the inferred user state, an appropriate access set which consists of users who gain access to the channel is chosen to maximize the expected number of successfully received packets in each slot under the heterogeneous delay constraints. The proposed protocol achieves the maximum possible throughput among all protocols at heavy traffic load and has small delay when the traffic load is light.

This paper is organized as follows. In Section II, we present the model of a communication network with heterogeneous QoS requirements and MPR capability. The existence of MAC protocols that ensures a given set of heterogeneous delay constraints is studied in Section III. In Sections IV and V, we propose the multiqueue service room (MQSR) protocol. Simulation examples are presented in Section VI, where the throughput and delay performance of the MQSR protocol is compared to that of the URN scheme [17] and the slotted ALOHA with optimal retransmission probability.

II. The Model

As illustrated in Fig. 1, the communication network considered here consists of M users who transmit data to a central controller through a common wireless channel. The three basic components of this network—the users, the common wireless channel, and the central controller—are specified, respectively, in Sections II-A–C.

A. Users

Each user generates data in the form of equal-sized packets. Transmission time is slotted, and each packet requires one time slot to transmit. Each user has a single buffer. At the beginning of each slot, a user independently generates a packet with probability p, but only accepts this packet if its buffer is currently empty. Packets generated by a user with a full buffer are assumed lost. Packets generated at the beginning of a slot may be transmitted in this slot, and a successfully transmitted packet leaves its buffer.

Users are partitioned into L groups according to their QoS constraints. The M_l $(l = 1, \ldots, L, \sum_{l=1}^{L} M_l = M)$ users in the lth group require their average packet delay at $p = 1$ to be no greater than d_l, where we define average packet delay as the expected number of slots from the time a packet enters a buffer until the end of its successful transmission. Note that $p = 1$ gives the heaviest traffic load. Since average packet delay generally increases with the traffic load, the delay requirements at $p = 1$ specifies the worst case performance.

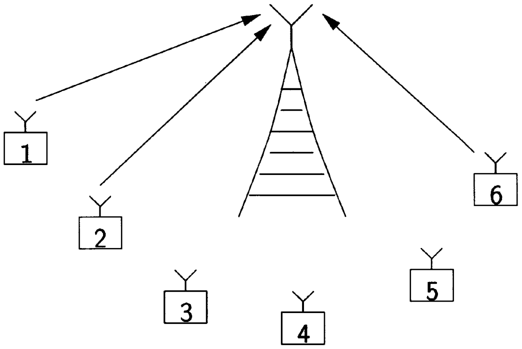


Fig. 1. Network model.

B. Channel

As considered in [4], [11], [12], the slotted channel is such that the probability of having k successes in a slot where there are n transmissions depends only on the number of transmitted packets. Let

$$C_{n,k} = P[k \text{ packets are correctly received } | n \text{ are transmitted}] \quad (1 \leq n \leq M, 0 \leq k \leq n).$$

The multipacket reception matrix of the channel is then defined as

$$\mathbf{C} = \begin{pmatrix} C_{1,0} & C_{1,1} & & & \\ C_{2,0} & C_{2,1} & C_{2,2} & & \\ \vdots & \vdots & \vdots & & \\ C_{M,0} & C_{M,1} & C_{M,2} & \cdots & C_{M,M} \end{pmatrix}. \tag{1}$$

For such an MPR channel, we define the channel capacity as

$$\eta \triangleq \max_{n=1,\ldots,M} \mathcal{C}_n \tag{2}$$

where

$$\mathcal{C}_n \triangleq \sum_{k=1}^{n} k C_{n,k} \tag{3}$$

is the expected number of packets correctly received when n packets are transmitted. Let

$$n_0 \triangleq \min \left\{ \arg \max_{n=1,\ldots,M} \mathcal{C}_n \right\}. \tag{4}$$

We can see that to achieve the channel capacity η, n_0 packets should be transmitted simultaneously. Noticing that the number of simultaneously transmitted packets to achieve η may not be unique, we define n_0 as the minimum to save transmission power. For MPR channels with n_0 greater than 1, contention should be preferred at any traffic load in order to fully exploit the MPR capability.

This general model for MPR channels applies to, as special examples, the conventional collision channel and channels

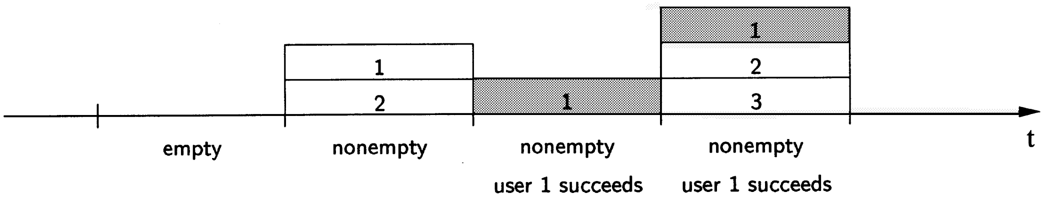


Fig. 2. Possible outcomes of a slot.

with capture. The reception matrix of the conventional collision channel and channels with capture are given by

$$\begin{pmatrix} 0 & 1 & 0 & \cdots & 0 \\ 1 & 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & & \vdots \\ 1 & 0 & 0 & \cdots & 0 \end{pmatrix}, \quad \begin{pmatrix} 1-p_1 & p_1 & 0 & \cdots & 0 \\ 1-p_2 & p_2 & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & & \vdots \\ 1-p_M & p_M & 0 & \cdots & 0 \end{pmatrix} \tag{5}$$

where p_i is the probability of capture given i simultaneous transmissions. With p_1 smaller than 1, this channel model can easily characterize noisy scenarios. Another example of an MPR channel is provided by a CDMA system where a packet is transmitted with a randomly generated code and is successfully received if and only if the number of simultaneously transmitted packets is no larger than P. The reception matrix for such an MPR channel with $P = 2$ is

$$\begin{pmatrix} 0 & 1 & 0 & 0 & \cdots & 0 \\ 0 & 0 & 1 & 0 & \cdots & 0 \\ 1 & 0 & 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & & \vdots \\ 1 & 0 & 0 & 0 & \cdots & 0 \end{pmatrix}. \tag{6}$$

The capacity of this MPR channel is 2 with $n_0 = 2$.

C. Central Controller

Access to the common wireless channel is controlled by the central controller. Specifically, the central controller decides, at the beginning of slot t for each t, an access set $\mathcal{A}(t)$ which contains users enabled to access the channel in slot t. It then broadcast $\mathcal{A}(t)$; users and only users in $\mathcal{A}(t)$ access the channel (if they have packets to transmit). At the end of slot t, the central controller observes the channel outcome $F(t)$ which contains information on whether slot t is empty and whose packets are successfully received in slot t. Here, we assume that the central controller can distinguish without error between empty and nonempty slots. However, if one or more packets are successfully demodulated at the end of slot t, the central controller does not assume the knowledge whether there are other packets transmitted but not successfully received in this slot. We illustrate this point in Fig. 2, where we consider possible outcomes of a slot: empty, nonempty with success, and nonempty without success (successfully received packets are illustrated by shaded rectangles). To the central controller, the two events which happened in the third and the fourth slot are indistinguishable.

After observing the channel outcome of slot t, the central controller acknowledges the sources of successfully received packets (if any) so that they can release their buffer and generate new packets. Users who transmit in slot t but do not receive acknowledgment assume their packets are lost and will retransmit the next time they are enabled. In this paper, we assume that the downlink channel (from the central controller to the users) is error free and the time for acknowledgment and broadcasting $\mathcal{A}(t)$ is negligible.

Our goal is to design a protocol for determining the access set $\mathcal{A}(t)$ for each t. The criterion for choosing $\mathcal{A}(t)$ is to maximize the expected number of successfully received packets in slot t while satisfying each user's delay requirement. The information assumed at the central controller includes the total number M of users, the number M_l $(l = 1, \ldots, L)$ of users in each group, the traffic load p, and the channel reception matrix $\mathbf{C}$. All these network parameters are assumed time invariant.

Before pursuing the protocol design, the first question we should answer is whether it is possible to satisfy a given set of heterogeneous delay constraints with a given channel. In Section III, a necessary and sufficient condition for the existence of a MAC protocol that ensures a given set of delay requirements is derived.

III. Existence Condition

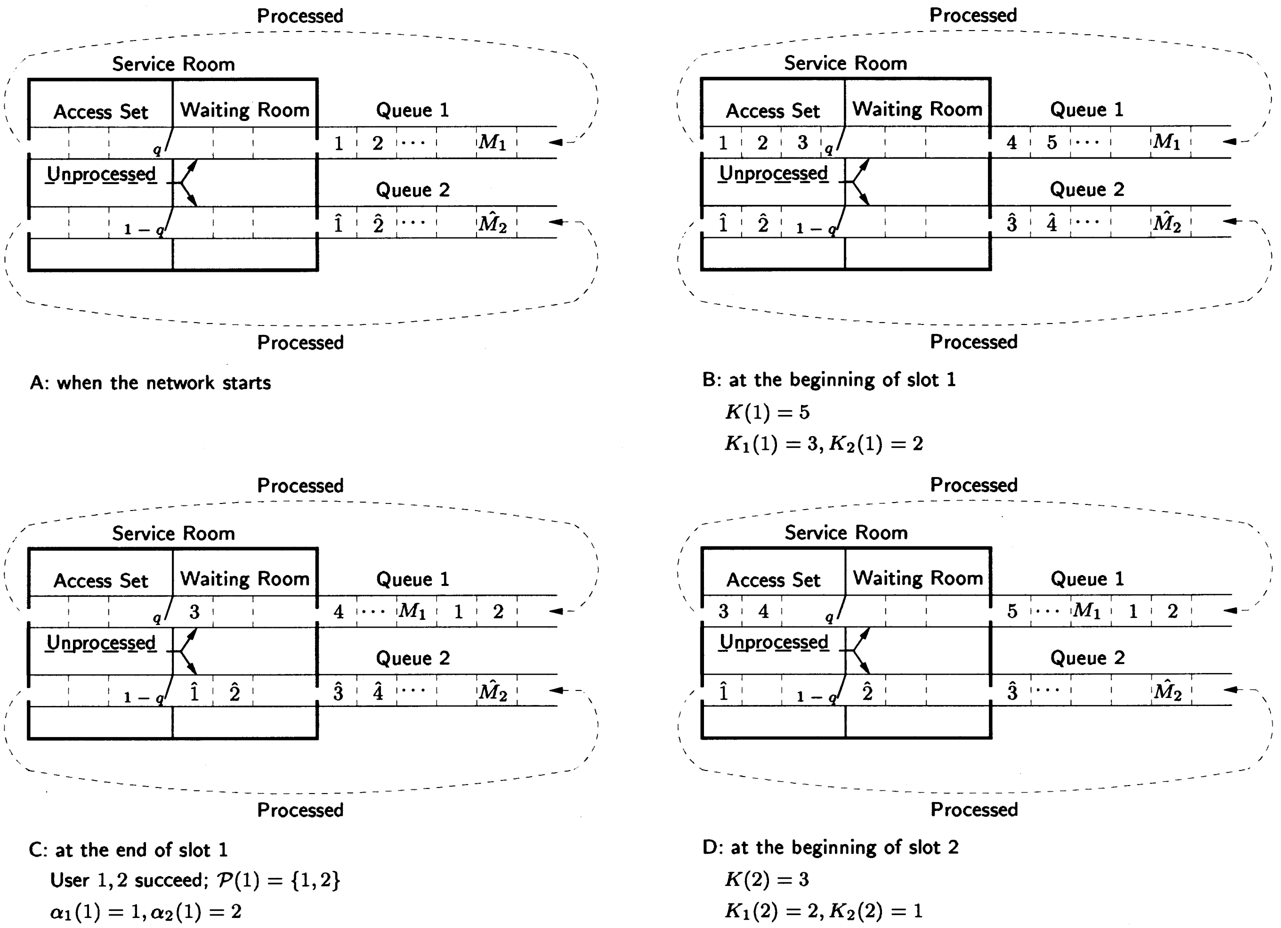
Satisfying a set of heterogeneous delay constraints essentially requires a prioritized allocation of the channel resource. Users with the strongest delay requirement demand a larger share of the channel resource. However, for a channel with limited capacity, we cannot expect that any set of delay constraints can be satisfied. In the following proposition, we give a necessary and sufficient condition for a set of delay requirements being achievable.

Proposition 1: Let M_l $(l = 1, \ldots, L)$ be the number of users who require their packet delay at $p = 1$ to be no larger than d_l. Then for the network model specified in Section II, there exists a MAC protocol that guarantees each user's delay requirement if and only if

$$\sum_{l=1}^{L} \frac{M_l}{d_l} \leq \eta \tag{7}$$

where η is the channel capacity defined in (2).

Proof: The proof of sufficiency is given by the fact that the MQSR protocol proposed in Section IV ensures each user's delay requirement when (7) holds (see Proposition 3). We now consider the proof of necessity. For $p \in (0, 1]$, let $T_l(p)$ denote the throughput of the lth group which is defined as the expected number of packets from the lth group that are successfully received in one slot. For a network where users have homogeneous and independent packet generation processes, we have the



A: when the network starts

B: at the beginning of slot 1
$K(1) = 5$
$K_1(1) = 3, K_2(1) = 2$

C: at the end of slot 1
User 1, 2 succeed; $\mathcal{P}(1) = \{1, 2\}$
$\alpha_1(1) = 1, \alpha_2(1) = 2$

D: at the beginning of slot 2
$K(2) = 3$
$K_1(2) = 2, K_2(2) = 1$

Fig. 3. Basic procedure of the multiqueue service room protocol.

following relation between the throughput $T_l(p)$ and the delay $D_l(p)$ under the equilibrium condition:

$$D_l(p) = 1 + \frac{M_l}{T_l(p)} - \frac{1}{p}, \quad l = 1, \dots, L. \tag{8}$$

A proof of (8) following [15] is provided in Appendix A. At $p = 1$, we have

$$D_l(1) = \frac{M_l}{T_l(1)}, \quad l = 1, \dots, L. \tag{9}$$

Thus, $D_l(1) \le d_l$ implies $T_l(1) \ge M_l/d_l$. Equation (7) then follows from the fact that for any p

$$\sum_{l=1}^{L} T_l(p) \le \eta. \tag{10}$$

□□□

IV. Basic Structure of the MQSR Protocol

We present the MQSR protocol for the case of $L = 2$. Its extension to cases with $L > 2$ is straightforward. We assume that users in the first group require $D_1(1) \le d_1$ and the requirement on $D_2(1)$ by users in the second group is such that condition (7) holds. To avoid the second group making unnecessary sacrifice, we design a protocol which yields $D_1(1) = d_1$.

The basic structure of the MQSR protocol is illustrated in Fig. 3, where users from the first group are indexed by i ($i = 1, \dots, M_1$) and those from the second by $\hat{\imath}$ ($\hat{\imath} = \hat{1}, \dots, \hat{M}_2$). As shown in Fig. 3(a), when the network starts, users of the two groups are waiting, respectively, in two queues to enter the service room for channel access. Users enter the service room in turn and stay ordered inside the service room. The service room consists of an access set and a waiting room. Users in the access set transmit, in the current slot, packets generated before entering the service room while users in the waiting room cannot access the channel until they join the access set. Packets generated by a user when it is inside the service room are held in the user's buffer (if the buffer is empty) and cannot be transmitted until next time this user enters the service room. After entering the service room, a user stays there until the central controller detects that either its packet generated before entering the service room has been successfully transmitted or it enters the service room with an empty buffer. At this time, we say this user is processed. A processed user leaves the service room and goes to the end of its queue.

Let $\mathcal{P}(t)$ denote the set of users who are processed in slot t. At the end of slot t, after determining $\mathcal{P}(t)$, the central controller empties the access set by removing processed users to the end of their queues and unprocessed users to the beginning of the waiting room. The central controller then chooses the access set for slot $t+1$ by specifying the size $K(t+1)(1 \le K(t+1) \le M)$

of the access set. These $K(t+1)$ users who will access the channel in slot $t+1$ are chosen one by one from these two groups. If there are users from both groups waiting outside the access set (either in the waiting room or in the queues), then with probability q, a new user who joins the access set is from the first group and with probability $1-q$ from the second. Otherwise, this user is from the group that still has users waiting outside the access set. Note that $K(t+1) \leq M$. It will never be the case that a new user is needed for the access set while no user is waiting outside. Let $K_l(t+1)$ $(l=1,2)$ be the number of users from the lth group who will access the channel in slot $t+1$. Then, given $K(t+1)$, the possible values of $K_1(t+1)$ are integers from $\max\{0, K(t+1)-M_2\}$ to $\min\{K(t+1), M_1\}$. Let $B(k,q,i)$ denote the probability mass at value i of a binomial distribution with k trials and a success probability q. Then the distribution of $K_1(t+1)$ given $K(t+1)=k$ for $k<M$ is

$$P[K_1(t+1)=k_1|K(t+1)=k] = \begin{cases} \sum_{i=0}^{k_1} B\{k,q,i\}, & \text{if } k_1 = \max\{0, k-M_2\} \\ \sum_{i=k_1}^{k} B\{k,q,i\}, & \text{if } k_1 = \min\{k, M_1\} \\ B(k,q,k_1), & \text{otherwise.} \end{cases} \tag{11}$$

For $k=M$, we have

$$P[K_1(t+1)=k_1|K(t+1)=k] = \begin{cases} 1, & \text{if } k_1 = M_1 \\ 0, & \text{otherwise.} \end{cases} \tag{12}$$

The value of $K_2(t+1)$ is determined by $K_2(t+1) = K(t+1) - K_1(t+1)$. Let $\alpha_l(t)$ $(l=1,2)$ be the number of users from the lth group who remain in the service room (specifically, in the waiting room) after processed users have been removed at the end of slot t. Then, if $K_l(t+1) > \alpha_l(t)$, the first $K_l(t+1) - \alpha_l(t)$ users in Queue l enter the service room and, along with the $\alpha_l(t)$ users in the waiting room, join the access set at the beginning of slot $t+1$. On the other hand, if $K_l(t+1) < \alpha_l(t)$, the first $K_l(t+1)$ users in the waiting room of the lth group enter the access set while the last $\alpha_l(t) - K_l(t+1)$ users remain in the waiting room.

We now consider the example in Fig. 3. The calculation of q, $K(t)$, and $\mathcal{P}(t)$ will be discussed in Section V. For now, we assume arbitrary values for q, $K(t)$, and $\mathcal{P}(t)$ to illustrate the basic procedure of the MQSR protocol.

Suppose that at the beginning of the first slot [Fig. 3(b)], the central controller decides that $K(1)=5$. A coin with bias q is then flipped five times to determine $K_1(1)$ and $K_2(1)$. Assume that $K_1(1)=3$ and $K_2(1)=2$. The central controller then broadcasts the identities of user 1, 2, 3, and $\hat{1}, \hat{2}$. These five users join the access set and transmit their packets (if any) in the first slot. At the end of this slot [Fig. 3(c)], suppose that the central controller successfully receives the packets from user 1 and 2 and decides $\mathcal{P}(1) = \{1,2\}$. The central controller then acknowledges these two users and removes them from the service room to the end of their queues. The unprocessed users 3, $\hat{1}$, $\hat{2}$ go to the waiting room. At the beginning of the second slot [Fig. 3(d)], suppose that $K(2)=3$ with $K_1(2)=2$ and $K_2(2)=1$. Then user 3, 4, and $\hat{1}$ form the access set and user $\hat{2}$ remains in the waiting room. The three users in the access set transmit their packets (if any) generated before their entering the service room. At the end of this slot, suppose that the central controller detects an empty slot. Then $\mathcal{P}(2) = \{3, 4, \hat{1}\}$. User $\hat{2}$ remain unprocessed, i.e., $\alpha_1(2)=0$ and $\alpha_2(2)=1$.

We point out that users do not need to keep track of their positions in the queues or the waiting room. The structure of the service room and waiting queues is kept at the central controller. Users only need to listen to the broadcasting at the beginning of each slot to know whether they are in the access set and users who have transmitted in a particular slot need only listen to the acknowledgment at the end of that slot. Since users in the access set can only transmit packets generated before they enter the service room, the central controller also notifies (when it broadcasts the access set) each user in the access set the time instant that user enters the service room for the last time.

The optimal window protocol proposed in [15] has a similar structure to the MQSR protocol with $L=1$. Relying on exhaustive search, however, the window protocol is only computationally feasible for networks with two or three users and no MPR. Furthermore, homogeneous QoS constraints are assumed in [15].

V. Parameter Design for the MQSR Protocol

In this section, we address the issue of parameter design for the MQSR protocol. The first parameter to be determined is q, an indicator of the priority of users in the first group over users in the second. Since q is constant for each slot, it can be designed off line. The two parameters to be determined on line are $K(t)$, the size of the access set for slot t, and $\mathcal{P}(t)$, the processed set of slot t. The problem of determining q and $K(t)$ is formulated in Section V-A and the determination of $\mathcal{P}(t)$ is detailed in Section V-C.

A. Problem Formulation

At the beginning of slot t, the central controller determines the access set $\mathcal{A}(t)$ by choosing $K(t)$ users from the head of multiple queues with a priority factor q. If we relabel users in each group at the beginning of each slot, starting from the service room to the end of the lth queue, we have

$$\mathcal{A}(t) = \{1, \ldots, K_1(t)\} \bigcup \{\hat{1}, \ldots, \hat{K}_2(t)\}. \tag{13}$$

Let $X_i^{(l)}(t)$ be the state of the ith $(i=1,\ldots,M_l)$ user in the lth $(l=1,2)$ group at the beginning of slot t (after new packet generation), where we define the state of a user as the number of packets it, if enabled, can transmit in slot t. Specifically, when the ith user in the lth group is inside the service room at the beginning of slot t, $X_i^{(l)}(t)$ is the number of packets generated before its entering the service room. When it is waiting in the queue, $X_i^{(l)}(t)$ denotes the number of packets in its buffer at the beginning of slot t. Under the single-buffer assumption, $X_i^{(l)}(t)$ is a random variable with possible values 0 and 1.

Recall that $F(t)$ denotes the channel outcome of slot t. With $\mathbf{C}$ and p given, the information, denoted by $I_{[0,t-1]}$, available at the beginning of slot t for determining $K(t)$ and q is the initial condition of the network in the form of the distribution of $X_i^{(l)}(1)$ $(l=1,2,\ i=1,\ldots,M_l)$, the access sets $\mathcal{A}(1),\ldots,\mathcal{A}(t-1)$, and the channel outcomes $F(1),\ldots,F(t-1)$. The criterion we use for determining $K(t)$ and q is to maximize the per-slot throughput under a set of delay constraints. Specifically, let $S(t)$ denote the number of successfully transmitted packets in slot t. It is a random variable whose distribution conditioned on $I_{[0,t-1]}$ depends on $\mathcal{A}(t)$ and the channel MPR matrix $\mathbf{C}$. The problem of determining $K(t)$ and q can then be formulated as

$$\begin{aligned}&\{q, K(t)\} = \arg \max_{k=1,\ldots,M} E_{I_{[0,t-1]}}[S(t)|K(t)=k]\\&\text{subject to } D_1(1) = d_1\end{aligned} \tag{14}$$

where $E_{I_{[0,t-1]}}[S(t)]$ is a shorthand for $E[S(t)|I_{[0,t-1]}]$. This constrained optimization problem can be decoupled into two steps. We first choose q so that the delay constraint $D_1(1)=d_1$ is satisfied. Then with q determined, choose $K(t)$ for each t so that $E_{I_{[0,t-1]}}[S(t)]$ is maximized. This decoupling is based on the fact that the maximization of $E_{I_{[0,t-1]}}[S(t)]$ at $p=1$ is independent of the delay constraint as indicated by the following proposition.

Proposition 2: For any set of delay constraints that satisfies (7), we have, at $p=1$

- *P2.1* $K(t)=n_0$ maximizes $E_{I_{[0,t-1]}}[S(t)]$ for any t.
- *P2.2* $T(1)=\eta$, where $T(1)$ is the network throughput (defined as the expected number of successfully transmitted packets in one slot) provided by the MQSR protocol at $p=1$.

Proof: At $p=1$, every user has a packet to transmit at the beginning of each slot. We thus have, for any q

$$\begin{aligned}K(t) =&\arg \max_{K(t)=1,\ldots,M} E_{I_{[0,t-1]}}[S(t)]\\=&\arg \max_{K(t)=1,\ldots,M} \mathcal{C}_{K(t)}\\=&n_0\end{aligned} \tag{15}$$

i.e., $K(t)=n_0$ for each t. Since $\mathcal{C}_{n_0}=\eta$, we have

$$T(1)=\eta. \tag{16}$$

□□□

Proposition 2 shows the optimality in terms of channel utilization of the MQSR protocol at $p=1$. It also demonstrates that the optimal size $K(t)$ of the access set and the throughput $T(1)$ of the whole network are independent of q at $p=1$, which enables the decoupling of the constrained optimization problem given in (14). As shown in Section V-B, q, by controlling the average percentage of users from the first group in the access set, determines the allocation of channel capacity between these two groups, which, in turn, determines the packet delay of each group at $p=1$.

B. Determination of q

We now consider the problem of determining q so that the delay constraint $D_1(1)=d_1$ is satisfied.

Proposition 3: Suppose that $M_l \geq n_0$ $(l=1,2)$. To satisfy the delay constraint $D_1(1)=d_1$, the parameter q in the MQSR protocol is given by

$$q=\frac{M_1}{d_1\eta}. \tag{17}$$

Proof: Recall that $K_1(t)$ denote the number of users from group 1 who access the channel in slot t and $S(t)$ the number of successfully received packets in slot t. Let $S_1(t)$ be the number of successfully received packets from group 1 in slot t. Since $K(t)=n_0$ [as shown in (15)] for any t at $p=1$ and q is independent of t, $\{S(t)\}_{t=1}^{\infty}$, $\{S_1(t)\}_{t=1}^{\infty}$, and $\{K_1(t)\}_{t=1}^{\infty}$ are independent and identically distributed (i.i.d.) sequences. Thus, we have, at $p=1$

$$E[S(t)]=\eta, \quad E[S_1(t)]=T_1(1), \quad E[K_1(t)]=qn_0 \tag{18}$$

where the last equation follows from the fact that $K_1(t)$ obeys a binomial distribution with n_0 trials and a success probability q under the condition of $M_l \geq n_0$ $(l=1,2)$. Furthermore, for any $0 \leq s, u \leq n_0$, we have

$$E[S_1(t)|K_1(t)=u, S(t)=s]=\frac{us}{n_0} \tag{19}$$

which follows from the results for the classic problem of "drawing without replacement," where we have total n_0 balls among which u are black and n_0-u are white, and $S_1(t)$ is the number of black balls we get after total s draws without replacement. Averaging over all the realizations of $K_1(t)$ and $S(t)$, and considering the independence between $K_1(t)$ and $S(t)$, we get

$$\begin{aligned}E[S_1(t)] =&\frac{1}{n_0}E[K_1(t)S(t)]\\=&\frac{1}{n_0}E[K_1(t)]E[S(t)]\\=&q\eta\end{aligned} \tag{20}$$

which, along with (18), leads to

$$T_1(1)=q\eta. \tag{21}$$

Combining with (9), we have

$$D_1(1)=\frac{M_1}{q\eta}. \tag{22}$$

To ensure $D_1(1)=d_1$, q should be determined by (17). □□□

When the condition of $M_l \geq n_0$ $(l=1,2)$ is violated, $K_1(t)$ given $K(t)=n_0$ no longer has a binomial distribution and the last equality in (18) does not hold. However, from the distribution given in (11) and (12), the expectation of $K_1(t)$ at $p=1$ can still be obtained as a function of q. With the same derivation as given in the proof of Proposition 3, we obtain q as the solution to

$$E[K_1(t)|K(t)=n_0]=\frac{n_0M_1}{d_1\eta}. \tag{23}$$

C. Determination of $\mathcal{P}(t)$ and $k(t+1)$

We now consider the two parameters to be designed on line. At the end of slot t, the central controller first determines, based on the channel outcome of slot t, the set $\mathcal{P}(t)$ of users that are processed in this slot. It then rearranges the order of users by moving processed users to the end of their queues and unprocessed ones to the head of their waiting rooms. The size $K(t+1)$ of access set for slot $t+1$ is then chosen and $K(t+1)$ users are selected from two groups with a biased coin. Before we get into the formal derivation of computing $\mathcal{P}(t)$ and $K(t+1)$, we present a simple example to provide insights to the basic idea.

1) Example: Consider a network with two users ($M = 2$). Each user with probability $p = 1/2$ independently generates a packet at the beginning of each slot. A noisy channel with capture effect is considered with channel reception matrix $\mathbf{C}$

$$\mathbf{C} = \begin{pmatrix} C_{1,0} & C_{1,1} & \\ C_{2,0} & C_{2,1} & C_{2,2} \end{pmatrix} = \begin{pmatrix} \frac{1}{2} & \frac{1}{2} & \\ \frac{3}{4} & \frac{1}{4} & 0 \end{pmatrix}. \tag{24}$$

We then have

$$\mathcal{C}_1 = \frac{1}{2}, \quad \mathcal{C}_2 = \frac{1}{4}, \quad \eta = \frac{1}{2}, \quad n_0 = 1. \tag{25}$$

Assume that user 1 requires its average packet delay at the heaviest traffic load to be no larger than 3 ($d_1 = 3$) while the delay requirement of user $\hat{1}$ is such that (7) holds, i.e., $M_1 = M_2 = 1$. Based on the delay requirement of user 1, we compute q from (17) as

$$q = \frac{M_1}{d_1 \eta} = \frac{2}{3}. \tag{26}$$

The initial condition I_0 of the network is assumed to be

$$P[X_1^{(1)}(1) = 1] = \frac{3}{4}, \quad P[X_1^{(2)}(1) = 1] = \frac{1}{2} \tag{27}$$

with $X_1^{(1)}(1)$ and $X_1^{(2)}(1)$ being independent. We are now ready to carry out the MQSR protocol.

At the beginning of the first slot, $K(1)$ needs to be determined based on the initial condition I_0 of the network. From (14), we have

$$K(1) = \arg\max_{k=1,2} E_{I_0}[S(1)|K(1) = k] \tag{28}$$

where we have decoupled the delay constraint from the maximization based on Proposition 2. We now compute $E_{I_0}[S(1)|K(1) = k]$ for all possible k's to determine $K(1)$. First, consider $k = 1$. With probability $q = 2/3$, the user who gains access to the channel in slot 1 is user 1. On the condition that user 1 is selected, with probability $P[X_1^{(1)} = 1] = 3/4$, it has a packet to transmit. Taking into account the case when user $\hat{1}$ gains access to the channel, we have

$$\begin{aligned} E_{I_0}[S(1)|K(1) = 1] = &\Big(qP[X_1^{(1)}(1) = 1] \\ &+ (1-q)P[X_1^{(2)}(1) = 1]\Big)\mathcal{C}_1 \\ = &\frac{1}{3}. \end{aligned} \tag{29}$$

Similarly, for $k = 2$, we have

$$\begin{aligned} E_{I_0}[S(1)|K(1) = 2] = &\Big(P[X_1^{(1)}(1) = 1, X_1^{(2)}(1) = 0] \\ &+ P[X_1^{(1)}(1) = 0, X_1^{(2)}(1) = 1]\Big)\mathcal{C}_1 \\ &+ P[X_1^{(1)}(1) = 1, X_1^{(2)}(1) = 1]\mathcal{C}_2 \\ = &\frac{11}{32} \end{aligned} \tag{30}$$

where we have used the independence between $X_1^{(1)}(1)$ and $X_1^{(2)}(1)$. Since $E_{I_0}[S(1)|K(1) = 2] > E_{I_0}[S(1)|K(1) = 1]$, we have $K(1) = 2$ and both users enter the service room to access the channel.

At the end of this slot, assume that the central controller observes a nonempty slot without success. Based on this observation, we need to decide which user or users are processed. Recall that a user is processed in a particular slot if its packet is successfully received in that slot or the central controller detects that it does not have a packet eligible for transmission (i.e., it enters the service room with an empty buffer). Specifically, let t^+ denote the time instance when the packets successfully transmitted in slot t have been removed from their buffer at the end of slot t. We have

$$\begin{aligned} \mathcal{P}(t) = &\left\{i : 1 \le i \le N_1(t), E_{I_{[0,t]}}[X_i^{(1)}(t^+)] = 0\right\} \\ &\bigcup \left\{\hat{i} : 1 \le i \le N_2(t), E_{I_{[0,t]}}[X_i^{(2)}(t^+)] = 0\right\} \end{aligned} \tag{31}$$

where $N_l(t)$ is the number of users from the lth group that are inside the service room (either in the access set or in the waiting room) in slot t. In our case, we have $N_1(1) = N_2(1) = 1$. To evaluate $E_{I_{[0,1]}}[X_i^{(l)}(1^+)]$, we need to compute the distribution of $X_i^{(l)}(1^+)$ from the distribution of $X_i^{(l)}(1)$ and the channel outcome $F(1)$. Though $X_1^{(1)}(1)$ and $X_1^{(2)}(1)$ are independent, these two users' states conditioned on $F(1)$ become correlated after accessing the channel simultaneously in slot 1. To fully capture the information provided by $F(1)$, we compute the joint distribution of $X_1^{(1)}(1^+)$ and $X_1^{(2)}(1^+)$. Let

$$\begin{aligned} P[F(1)] = &P\left[X_1^{(1)}(1) = 1, X_1^{(2)}(1) = 0\right] C_{1,0} \\ &+ P\left[X_1^{(1)}(1) = 0, X_1^{(2)}(1) = 1\right] C_{1,0} \\ &+ P\left[X_1^{(1)}(1) = 1, X_1^{(2)}(1) = 1\right] C_{2,0} \end{aligned} \tag{32}$$

denote the total probability that $F(1)$ occurs. We have, based on Bayes' theorem

$$\begin{aligned} &P_{I_{[0,1]}}\left[X_1^{(1)}(1^+) = 0, X_1^{(2)}(1^+) = 0\right] = 0 \\ &P_{I_{[0,1]}}\left[X_1^{(1)}(1^+) = 1, X_1^{(2)}(1^+) = 0\right] \\ &\quad = \frac{P\left[X_1^{(1)}(1) = 1, X_1^{(2)}(1) = 0\right] C_{1,0}}{P[F(1)]} = \frac{6}{17} \\ &P_{I_{[0,1]}}\left[X_1^{(1)}(1^+) = 0, X_1^{(2)}(1^+) = 1\right] \\ &\quad = \frac{P\left[X_1^{(1)}(1) = 0, X_1^{(2)}(1) = 1\right] C_{1,0}}{P[F(1)]} = \frac{2}{17} \\ &P_{I_{[0,1]}}\left[X_1^{(1)}(1^+) = 1, X_1^{(2)}(1^+) = 1\right] \\ &\quad = \frac{P\left[X_1^{(1)}(1) = 1, X_1^{(2)}(1) = 1\right] C_{2,0}}{P[F(1)]} = \frac{9}{17}. \end{aligned}$$

It is easy to see that neither of these two users is processed, i.e., $\mathcal{P}(1) = \phi$. Hence, both users go to the waiting room.

At the beginning of slot 2, $K(2)$ needs to be determined by comparing $E_{I_{[0,1]}}[S(2)|K(2) = 1]$ with $E_{I_{[0,1]}}[S(2)|K(2) = 2]$ [see (28)]. To compute $E_{I_{[0,1]}}[S(2)|K(2) = k]$, we need the joint distribution of $X_1^{(1)}(2)$ and $X_1^{(2)}(2)$, which can be obtained from the joint distribution of $X_1^{(1)}(1^+)$ and $X_1^{(2)}(1^+)$. With the restriction that packets generated by a user inside the service room cannot be transmitted until the next time this user enters the service room, the state of a user does not change while it is inside the service room. We then have

$$P\left[X_1^{(1)}(2), X_1^{(2)}(2)\right] = P\left[X_1^{(1)}(1^+), X_1^{(2)}(1^+)\right]. \quad (33)$$

Similar to the computation of $E_{I_0}[S(1)|K(1)=k]$ as given in (29) and (30), we have

$$\begin{aligned} E_{I_{[0,1]}}[S(2)|K(2)=1] = &\Big(qP\left[X_1^{(1)}(2)=1\right] \\ &+(1-q)P\left[X_1^{(2)}(2)=1\right]\Big)\mathcal{C}_1 \\ =&0.402 \\ E_{I_{[0,1]}}[S(2)|K(2)=2] = &\Big(P\left[X_1^{(1)}(2)=1, X_1^{(2)}(2)=0\right] \\ &+P\left[X_1^{(1)}(2)=0, X_1^{(2)}(2)=1\right]\Big)\mathcal{C}_1 \\ &+P\left[X_1^{(1)}(2)=1, X_1^{(2)}(2)=1\right]\mathcal{C}_2 \\ =&0.368. \end{aligned}$$

Comparing $E_{I_1}[S(2)|K(2)=1]$ and $E_{I_1}[S(2)|K(2)=2]$, we choose $K(2)$ to be 1. One user is then chosen to join the access set with the priority factor q. Suppose that user 1 is enabled to access the channel in the second slot and an empty slot is observed. Based on this channel outcome, we obtain the state of users at the time instance 2^+ as follows:

$$\begin{aligned} P_{I_{[0,2]}}\left[X_1^{(1)}(2^+)=0, X_1^{(2)}(2^+)=0\right] =&0 \\ P_{I_{[0,2]}}\left[X_1^{(1)}(2^+)=1, X_1^{(2)}(2^+)=0\right] =&0 \\ P_{I_{[0,2]}}\left[X_1^{(1)}(2^+)=0, X_1^{(2)}(2^+)=1\right] =&1 \\ P_{I_{[0,2]}}\left[X_1^{(1)}(2^+)=1, X_1^{(2)}(2^+)=1\right] =&0. \end{aligned}$$

Note that after optimally exploiting the information provided by $F(1)$ and $F(2)$, the state of each user at the end of slot 2 is completely known to the central controller. Since $E_{I_{[0,1]}}\left[X_1^{(1)}(2^+)\right] = 0$, we have $\mathcal{P}(2) = \{1\}$; user 1 leaves the service room and goes to its queue.

We now need to compute the joint distribution of the state of users at the beginning of slot 3 (after new packet generation), based on which $K(3)$ can be obtained. Note that $X_1^{(1)}(3)$ and $X_1^{(2)}(3)$ are independent. Their joint distribution can be obtained from their marginals. First, consider user $\hat{1}$ who is inside the service room. We have

$$P\left[X_1^{(2)}(3)=1\right] = P\left[X_1^{(2)}(2^+)=1\right] = 1. \quad (34)$$

For user 1 who enters the service room in the first slot with an empty buffer, it has been generating packets for two slots. Hence

$$P\left[X_1^{(1)}(3)=1\right] = 1-(1-p)^2 = \frac{3}{4}. \quad (35)$$

We now summarize the insights we gain from this example.

- The state of users at the beginning of each slot is the most crucial information for optimal channel accessing. If this information is known to the central controller, perfect scheduling of transmission can be performed. Without this information, the MQSR protocol controls channel access based on an optimal estimate of the state of users. At the end of slot t for each t, the joint distribution $P_{I_{[0,t]}}\left\{X_i^{(l)}(t^+),\ l=1,2,\ i=1,\ldots,M_l\right\}$ of the state of the users is updated by incorporating the channel outcome $F(t)$. This joint distribution serves as the basis for determining the processed set $\mathcal{P}(t)$ and the size $K(t+1)$ of the access set for slot $t+1$.
- Restricting unprocessed users within the service room makes the state of users outside the service room independent of the state of users inside the service room for the reason that any packet held by a user outside the service room has never been simultaneously transmitted with a packet held by a user inside the service room. This independence enables us to compute $P_{I_{[0,t-1]}}\left\{X_i^{(l)}(t),\ l=1,2,\ i=1,\ldots,M_l\right\}$ from the conditional joint distribution of the state of users inside the service room and the marginal distribution of the state of users outside the service room. Thus, only the conditional joint distribution of the state of users inside the service room needs to be updated at the beginning of each slot.
- Restraining users inside the service room from transmitting packets generated during their current visit to the service room prevents their states from changing while we are updating their conditional joint distribution. This significantly reduces the computational complexity. Furthermore, this time control imposed on packets being eligible for transmission and the circular movement of users in the queues ensure fair channel access and prevent the situation where a user who keeps generating new packets seizes the channel.

2) Determination of $\mathcal{P}(t)$: As discussed in Section V-C1, $\mathcal{P}(t)$ is determined by computing the joint distribution of

$$\begin{aligned} &P\left[\left\{X_i^{(l)}(t^+)=x_i^{(l)},\ l=1,2,\ i=1,\ldots,N_l(t)\right\}|I_{[0,t]}\right] \\ &= \frac{P\left[\left\{X_i^{(l)}(t^+)=x_i^{(l)},\ l=1,2,\ i=1,\ldots,N_l(t)\right\}, F(t)|\mathcal{A}(t), I_{[0,t-1]}\right]}{P[F(t)|\mathcal{A}(t), I_{[0,t-1]}]} \\ &= \begin{cases} 0, & \text{if } x_i^{(l)} \neq 0 \text{ for any } i \leq K_l(t) \\ \dfrac{P\left[\left\{X_i^{(l)}(t)=x_i^{(l)},\ l=1,2,\ i=1,\ldots,N_l(t)\right\}|I_{[0,t-1]}\right]}{\sum\limits_{\left\{x_i^{(l)}=0\ \forall i\leq K_l(t)\right\}} P[\{X_i^{(l)}(t)=x_i^{(l)},\ l=1,2,\ i=1,\ldots,N_l(t)\}|I_{[0,t-1]}]}, & \text{otherwise.} \end{cases} \quad (36) \end{aligned}$$

$X_i^{(l)}(t^+)$ $(l = 1, 2, i = 1, \ldots, N_l(t))$ from the joint distribution of $X_i^{(l)}(t)$ and the channel outcome $F(t)$. If slot t is empty, we have, from Bayes' theorem, (36), shown at the bottom of the previous page. If, on the other hand, slot t is nonempty and $s_i^{(l)}$ packets from the ith user of the lth group are successfully received at the end of slot t, then for $0 \leq x_i^{(l)} \leq 1 - s_i^{(l)}$ $(l = 1, 2,\ i = 1, \ldots, N_l(t))$, we have (37), shown at the bottom of the page, where $z = \sum_{l=1}^{2} \sum_{i=1}^{K_l(t)} \left(x_i^{(l)} + s_i^{(l)}\right)$ and $S = \sum_{l=1}^{2} \sum_{i=1}^{K_l(t)} s_i^{(l)}$ are, respectively, the total number of packets transmitted and the total number of packets successfully received in slot t.

3) Determination of $K(t+1)$*:* As shown in (14), $K(t+1)$ is obtained by maximizing $E_{I_{[0,t]}}[S(t+1)]$ with q determined by (17), i.e.

$$K(t+1) = \arg \max_{k=1,\ldots,M} E_{I_{[0,t]}}[S(t+1)|K(t+1) = k] \quad (38)$$

where $E_{I_{[0,t]}}[S(t+1)|K(t+1) = k]$ is given by

$$E_{I_{[0,t]}}[S(t+1)|K(t+1) = k] = \sum_{k_1=\max(0,k-M_2)}^{\min(k,M_1)} P[K_1(t+1) = k_1|K(t+1) = k] \quad (39)$$

$$E_{I_{[0,t]}}[S(t+1)|K_1(t+1) = k_1,\ K_2(t+1) = k - k_1] \quad (40)$$

with

$$E_{I_{[0,t]}}\left[S(t+1)|K_1(t+1) = k_1,\ K_2(t+1) = k - k_1\right] = \sum_{n=1}^{k} \mathcal{C}_n P\left[\sum_{i=1}^{k_1} X_i^{(1)}(t+1) + \sum_{i=1}^{k-k_1} X_i^{(2)}(t+1) = n|I_{[0,t]}\right]. \quad (41)$$

To obtain $P\left[\sum_{i=1}^{k_1} X_i^{(1)}(t+1) + \sum_{i=1}^{k-k_1} X_i^{(2)}(t+1) = n|I_{[0,t]}\right]$ for all possible k and k_1, we compute the conditional joint distribution of $\left\{X_i^{(l)}(t+1),\ l = 1, 2,\ i = 1, \ldots, M_l\right\}$ by classifying users into two sets: users inside the service room and users waiting in the queues at the beginning of slot $t+1$.

We first consider users inside the service room at the beginning of slot $t+1$. Recall that $N_l(t)$ denotes the number of users from the lth group that are inside the service room in slot t and $\alpha_l(t)$ the number of unprocessed users from the lth group in slot t (without loss of generality, we assume these unprocessed users are the first $\alpha_l(t)$ of the $N_l(t)$ users). These unprocessed users in slot t will remain in the service room in slot $t+1$. Since packets generated by them at the beginning of slot $t+1$ cannot be transmitted until the next time they enter the service room, we have $X_i^{(l)}(t+1) = X_i^{(l)}(t^+)$ for $l = 1, 2,\ i = 1, \ldots, \alpha_l(t)$. Hence, the conditional joint distribution of $\left\{X_i^{(l)}(t+1),\ l = 1, 2,\ i = 1, \ldots, \alpha_l(t)\right\}$ can be easily obtained from the conditional joint distribution of $\{X_i^{(l)}(t^+),\ l = 1, 2,\ i = 1, \ldots, N_l(t)\}$ given by (36) and (37) by summing over all possible values taken by $X_i^{(l)}(t^+)$ $(l = 1, 2,\ i = \alpha_l(t)+1, \ldots, N_l(t))$. See (42), shown at the bottom of the page.

We now consider users waiting in the queues at the beginning of slot $t+1$. The marginal distribution of $X_i^{(l)}(t+1)$ $(l = 1, 2,\ i = \alpha_l(t) + 1, \ldots, M_l(t))$ is given by

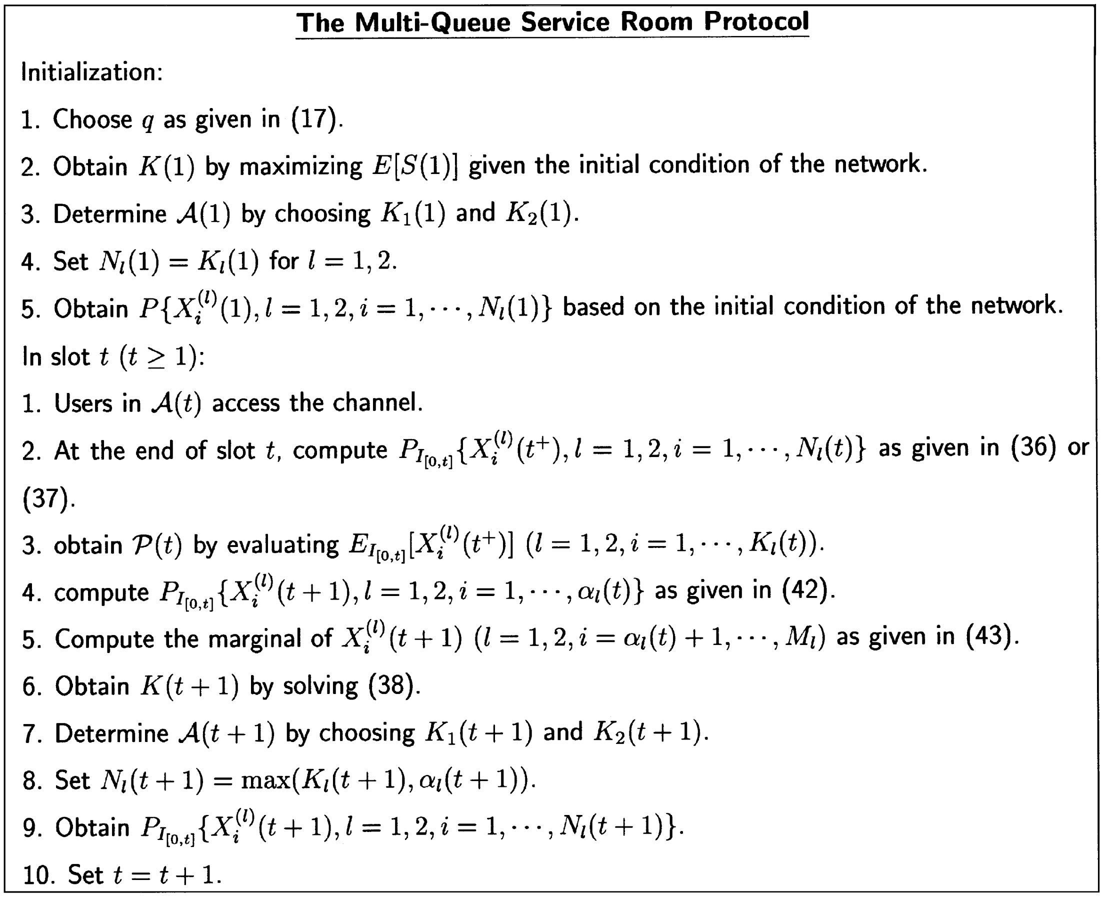
$$P\left[X_i^{(l)}(t+1) = x\right] = \begin{cases} (1-p)^{W_i^{(l)}(t+1)}, & \text{if } x = 0 \\ 1 - (1-p)^{W_i^{(l)}(t+1)}, & \text{if } x = 1 \\ 0, & \text{otherwise} \end{cases} \quad (43)$$

where $W_i^{(l)}(t+1) = t + 1 - \tau_i^{(l)}$ with $\tau_i^{(l)}$ defined as the index of the slot in which the ith user in the lth group last entered the service room or the index of the slot in which this user last successfully transmitted a packet, whichever is larger.

By the independence of traffic generation among all users, the conditional joint distribution of $\left\{X_i^{(l)}(t+1), l = 1, 2,\ i = 1, \ldots, M_l\right\}$ is obtained as the product of the conditional joint distribution of $\left\{X_i^{(l)}(t+1), l = 1, 2, i = 1, \ldots, \alpha_l(t)\right\}$ given in (42) and the marginal distribution of $X_i^{(l)}(t+1)(l = 1, 2, i = \alpha_l(t) + 1, \ldots, M_l(t))$ given in (43). With this joint distribution, $E_{I_{[0,t]}}[S(t+1)|K(t+1) = k]$ can be computed

$$P\left[\left\{X_i^{(l)}(t^+) = x_i^{(l)},\ l = 1, 2,\ i = 1, \ldots, N_l(t)\right\}|I_{[0,t]}\right] = \frac{\frac{z!}{(S!(z-S)!)} C_{z,S} P\left[\left\{X_i^{(l)}(t) = x_i^{(l)} + s_i^{(l)}, l = 1, 2, i = 1, \ldots, N_l(t)\right\}|I_{[0,t-1]}\right]}{\sum_{\left\{0 \leq x_i^{(l)} \leq 1 - s_i^{(l)},\ l=1,2,\ i=1,\ldots,N_l(t)\right\}} \frac{z!}{(S!(z-S)!)} C_{z,S} P\left[\left\{X_i^{(l)}(t) = x_i^{(l)} + s_i^{(l)},\ l = 1, 2,\ i = 1, \ldots, N_l(t)\right\}|I_{[0,t-1]}\right]} \quad (37)$$

$$P\left[\left\{X_i^{(l)}(t+1) = x_i^{(l)},\ l = 1, 2,\ i = 1, \ldots, \alpha_l(t)\right\}|I_{[0,t]}\right] = \sum_{\left\{x_i^{(1)}, x_j^{(2)}, i, \hat{j} \in \mathcal{P}(t)\right\}} P\left[\left\{X_i^{(l)}(t^+) = x_i^{(l)},\ l = 1, 2,\ i = 1, \ldots, N_l(t)\right\}|I_{[0,t]}\right]. \quad (42)$$



The Multi-Queue Service Room Protocol

```
Initialization:
1. Choose $q$ as given in (17).
2. Obtain $K(1)$ by maximizing $E[S(1)]$ given the initial condition of the network.
3. Determine $\mathcal{A}(1)$ by choosing $K_1(1)$ and $K_2(1)$.
4. Set $N_l(1) = K_l(1)$ for $l = 1, 2$.
5. Obtain $P\{X_i^{(l)}(1), l = 1, 2, i = 1, \cdots, N_l(1)\}$ based on the initial condition of the network.
In slot $t$ $(t \geq 1)$:
1. Users in $\mathcal{A}(t)$ access the channel.
2. At the end of slot $t$, compute $P_{I_{[0,t]}}\{X_i^{(l)}(t^+), l = 1, 2, i = 1, \cdots, N_l(t)\}$ as given in (36) or (37).
3. obtain $\mathcal{P}(t)$ by evaluating $E_{I_{[0,t]}}[X_i^{(l)}(t^+)]$ $(l = 1, 2, i = 1, \cdots, K_l(t))$.
4. compute $P_{I_{[0,t]}}\{X_i^{(l)}(t+1), l = 1, 2, i = 1, \cdots, \alpha_l(t)\}$ as given in (42).
5. Compute the marginal of $X_i^{(l)}(t+1)$ $(l = 1, 2, i = \alpha_l(t) + 1, \cdots, M_l)$ as given in (43).
6. Obtain $K(t+1)$ by solving (38).
7. Determine $\mathcal{A}(t+1)$ by choosing $K_1(t+1)$ and $K_2(t+1)$.
8. Set $N_l(t+1) = \max(K_l(t+1), \alpha_l(t+1))$.
9. Obtain $P_{I_{[0,t]}}\{X_i^{(l)}(t+1), l = 1, 2, i = 1, \cdots, N_l(t+1)\}$.
10. Set $t = t + 1$.
```

Fig. 4. MQSR protocol.

for all possible k and the optimal size $K(t+1)$ of the access set can be determined.

Up to now, all parameters in the MQSR protocol have been specified. The basic procedure of the MQSR protocol is summarized in Fig. 4.

VI. Simulation Examples

Presented in this section are simulation studies on the throughput and delay performance of the proposed MQSR protocol in a CDMA network with $M = 10$ users. The channel reception matrix is given in (6), which shows that the capacity of this channel is 2 with $n_0 = 2$.

A. Performance Comparison Under Homogeneous Delay Constraints

We first consider the scenario of homogeneous QoS requirement $(L = 1)$ and compare the performance of the proposed MQSR protocol with that of the URN scheme [17] and the slotted ALOHA with optimal retransmission probability. As shown in [15], for a network model specified in Section II, the performance measures—throughput, delay, and packet drop rate—are equivalent. A higher throughput implies a smaller delay and a smaller packet drop rate. In this simulation example, we use throughput as our measure to evaluate the performance of the MQSR protocol.

The URN scheme was originally proposed for the conventional collision channel. Given the total number of active users (users with packet to transmit) at the beginning of slot t, this protocol randomly picks $K(t)$ users to access the channel in slot t so that the probability of having one active user in the access set is maximized. Here, we extend the URN scheme to networks with MPR capability, where the size of the access set for each slot is chosen to maximize the probability of having n_0 active users in the access set. In the simulation examples, we assumed that the total number of active users at the beginning of each slot was known in the URN scheme. The throughput of the MQSR protocol and the URN scheme was obtained by simulations while that of the slotted ALOHA was a theoretical result obtained by analyzing its Markov chain representation. At each tested traffic load, the throughput of slotted ALOHA with all possible retransmission probability (from 0 to 1 with a grid of 0.05) was analyzed and the maximum was chosen as its performance at that traffic load.

As shown in Fig. 5, the MQSR protocol achieved significant improvement in throughput over the slotted ALOHA with optimal retransmission probability. As compared to the URN scheme, the MQSR protocol performed better for $p \geq 0.2$ and slightly worse for $p \leq 0.1$. The reason for this lies in the fact that the knowledge of the number of active users at the beginning of each slot was assumed by the URN scheme. At light traffic load with $p < 0.2$, the probability of having no more than $n_0 = 2$ active users in the network at the beginning

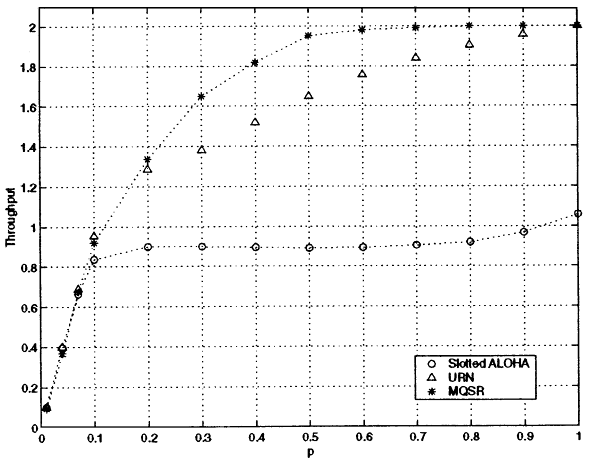


Fig. 5. Throughput comparison.

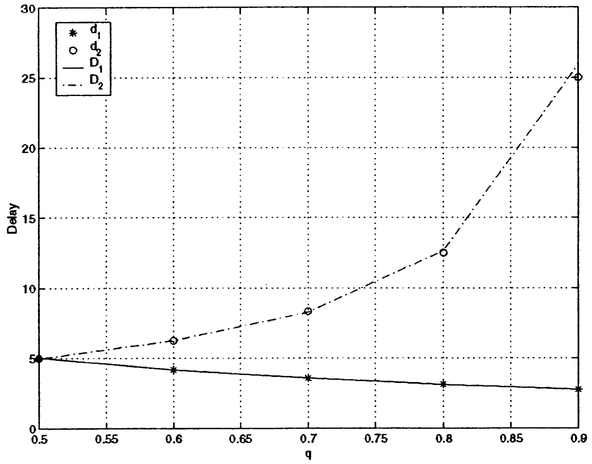


Fig. 6. Delay performance of the MQSR protocol at $p = 1$.

of each slot is large. For example, this probability is no less than $\sum_{i=0}^{2} B(10, 0.1, i) = 0.9298$ at $p = 0.1$. When the total number of active users is no more than n_0, the knowledge of the number of active users is equivalent to the knowledge of each user's state, in the sense that both lead to the optimal (in terms of per-slot throughput) decision $K(t) = M$. Hence, with large probability, the URN scheme at light traffic load maximizes the per-slot throughput with the knowledge of each user's state while the MQSR protocol does so without this knowledge. It then becomes clear why the MQSR protocol performed worse than the URN scheme at light traffic load. Actually, a close performance to that of the URN scheme at light traffic load demonstrates the MQSR protocol's capability of fully exploiting the information provided by the channel outcomes. At moderate and heavy traffic load, even with the knowledge of the total number of active users at the beginning of each slot, the URN scheme yielded a performance inferior to that of the MQSR protocol. This indicates that instead of the total number of active users, the joint distribution of all users' state conditioned on all previous channel outcomes should be defined as the network state for designing optimal access control schemes.

Fig. 5 also shows that the MQSR protocol and the URN scheme achieved the channel capacity at heavy traffic load, as expected. Note that the MQSR protocol already achieved the capacity at moderate traffic load $p = 0.5$, while the URN scheme did so at $p = 1$.

B. Performance Under Heterogeneous Delay Constraints

We now consider the case of $L = 2$, $M_1 = M_2 = 5$, where users of the first group require their packet delay D_1 at $p = 1$ to be no larger than d_1. We considered different delay requirement of the first group, as illustrated by asterisks in Fig. 6. The corresponding q was obtained by (17). The simulated delay of the first group was indicated by the solid line in Fig. 6. The circles and dashed line indicate, respectively, the calculated delay and simulated delay of the second group for a given q. Fig. 6 shows that the delay requirement of the first group was satisfied for the choice of q given in (17).

VII. Conclusion

In this paper, we have proposed a MQSR MAC protocol designed explicitly for multiaccess networks with MPR capability. By optimally exploiting all available information up to the current slot, the proposed MQSR protocol dynamically controls the size of the access set according to the traffic load and the channel MPR capability so that the expected number of successfully transmitted packets is maximized under a set of heterogeneous delay constraints. As a consequence, the channel MPR capability is efficiently exploited and the channel capacity is achieved at heavy traffic load.

A heuristic analysis on the packet delay provided by the MQSR protocol at any traffic load is given in Appendix B. While deriving an upper bound on the packet delay, we provide insights into the behavior of the MQSR protocol and answer the question whether it is possible that a user stays in the service room for an infinitely long period. Upper bounds on the expected number of slots that an active user (a user who enters the service room with a packet) and an idle user (a user who enters the service room without packet) spend in the service room during one visit are obtained.

Appendix A

A. Proof of (8)

Here, we abbreviate $T_l(p)$ to T_l. The same applies to $D_l(p)$.

Let B_l denote the expected number of backlogged users in the lth group, where a user is backlogged if its buffer is unable to accept an arriving packet. Let N_l denote the expected number of packets held by users in the lth group. By noting that a user with a buffered packet is only backlogged if it is unable to successfully transmit this packet, we have

$$N_l = B_l + T_l. \tag{44}$$

Since under equilibrium conditions, the expected number of successfully transmitted packets in one slot equals to the expected number of packets generated by unbacklogged users, we have

$$T_l = p(M_l - B_l). \tag{45}$$

Solving for B_l from (45) and substituting into (44), we get

$$N_l = T_l + M_l - \frac{T_l}{p}. \tag{46}$$

From Little's Theorem, we also have

$$D_l = \frac{N_l}{T_l}. \tag{47}$$

Equation (8) then follows by substituting (46) into (47). □□□

Appendix B

A. Analysis of Packet Delay

Here, we give an upper bound of the packet delay provided by the MQSR protocol at any traffic load under the equilibrium condition. The case of $L=1$ and an MPR channel with $C_{n,k}>0$ for $n=1,\ldots,M$ and $k=0,\ldots,n$ is considered.

At $p=1$, we readily have, from (9) and Proposition 2

$$D(1) = \frac{M}{\eta}. \tag{48}$$

We now provide an upper bound on $D(p)$ for $p \in (0,1)$. For simplicity, we abbreviate $D(p)$ to D.

Let $E[\tau]$ denote the average number of slots a user stays in the service room during one visit. Since in any slot, there is at least one user inside the service room, we have

$$D \leq ME[\tau]. \tag{49}$$

In order to bound $E[\tau]$, we consider two cases: the user of interest (UoI) is active (it enters the service room with a packet) or it is idle (it enters the service room without a packet). Define

$$m_1 \triangleq E[\tau \mid \text{the UoI is active}], \quad m_2 \triangleq E[\tau \mid \text{the UoI is idle}]. \tag{50}$$

We now derive upper bounds on m_1 and m_2.

Case 1: The UoI is Active: Let $E[\tau_A]$ denote the average number of slots that an active user stays in the access set during one visit to the service room. Since $K(t) \geq 1$ for any t, and an idle user, besides slots during which it stays in the access set with other active users, can only stay in the access set alone for at most one slot, we have

$$m_1 \leq ME[\tau_A]. \tag{51}$$

We now bound $E[\tau_A]$ as follows.

$$\begin{aligned} E[\tau_A] &= \sum_{n=1}^{\infty} P[\tau_A \geq n] \\ &\leq \sum_{n=1}^{\infty} P\left[\text{in each of } (n-1) \text{ slots, not all transmitted} \right. \\ &\qquad\qquad \left. \text{packets are successfully received}\right] \\ &\leq \sum_{n=1}^{\infty} \left(1 - \min_{l=1,\ldots,M} \mathcal{C}_{l,l}\right)^{n-1} \\ &= \frac{1}{\min_{l=1,\ldots,M} \mathcal{C}_{l,l}}. \end{aligned} \tag{52}$$

Note that $\mathcal{C}_{l,l} > 0$ for all l, a consequence of the condition that $C_{n,k} > 0$ for $n=1,\ldots,M$ and $k=0,\ldots,n$. Thus, from (51) and (52), we have

$$m_1 \leq \frac{M}{\min_{l=1,\ldots,M} \mathcal{C}_{l,l}}. \tag{53}$$

Case 2: The UoI is Idle: Suppose that when the UoI enters the service room, there are, before it, j $(j \geq 0)$ active users inside the service room. With $C_{n,k} > 0$ for $n=1,\ldots,M$ and $k=0,\ldots,n$, the idle UoI can only leave the service room after it is involved in an empty slot, which can only happen after these j active users are processed. Hence, after at most jm_1 slots on the average, there are no active users before the UoI. We have a situation where there are $k-1$ $(k \geq 1)$ idle users before the UoI and total $i-1$ $(i \geq 1)$ idle users in the access set with the UoI. Let $E\left[\tau_0^{(i)}\right]$ denote the average number of slots from the time instant that this situation occurs to the time instant that the UoI is processed. We then have, with $j \leq M-1$

$$m_2 \leq (M-1)m_1 + \max_{i=1,\ldots,M} E\left[\tau_0^{(i)}\right]. \tag{54}$$

We now bound $E\left[\tau_0^{(i)}\right]$ for $i=1,\ldots,M$. It is clear that $E\left[\tau_0^{(M)}\right] = 1$. Suppose that the UoI is the first user in the access set. In this case, the UoI is processed when the first empty slot occurs. Thus, the worst case for $\tau_0^{(i)}$ is that no empty slots occur until the number of idle users in the access set reaches M. Let $E\left[\xi^{(i)}\right]$ denote the average number of slots needed for the number of idle users in the access set increasing from i to $i+1$ given that no empty slot occurs. We have

$$\begin{aligned} E\left[\tau_0^{(i)}\right] &\leq E\left[\xi^{(i)}\right] + E\left[\tau_0^{(i+1)}\right] \\ &\leq \sum_{r=i}^{M-1} E\left[\xi^{(r)}\right] + E\left[\tau_0^{(M)}\right] \\ &= \sum_{r=i}^{M-1} E\left[\xi^{(r)}\right] + 1. \end{aligned} \tag{55}$$

Now consider the general case where the UoI is the kth $(k=1,\ldots,i)$ idle user in the access set. In this case, the worst situation, which involves only the first user in the access set, for $\tau_0^{(i)}$ is that $k-1$ empty slotsoccur before the number of idle users in the access set reaches M. Thus, with $k \leq i$, we have

$$E\left[\tau_0^{(i)}\right] \leq i \sum_{r=i}^{M-1} E\left[\xi^{(r)}\right] + 1. \tag{56}$$

Now consider the user, denoted User A, who will be the $(i+1)$th idle user in the access set. Given that User A becomes the $(i+1)$th idle user at its nth visit to the service room, $E\left[\xi_n^{(i)}\right]$ denotes the average number of slots until its nth visit to the service room. Let p_n be the probability that it is the nth visit to the service room that User A becomes idle. We then have

$$E\left[\xi^{(i)}\right] = \sum_{n=1}^{\infty} E\left[\xi_n^{(i)}\right] p_n. \tag{57}$$

We now need to bound $E\left[\xi_n^{(i)}\right]$ and p_n. Let $E[H]$ denote the average duration of the period between two consecutive visits by User A to the service room among the first n visits. Then

$$E[H] \leq (M-i)m_1 \tag{58}$$

which follows from the fact that all the $M-i$ users are active during any visit to the service room before the nth visit of User A. We can then bound $E\left[\xi_n^{(i)}\right]$ as follows:

$$E\left[\xi_n^{(i)}\right] \leq n(M-i)m_1. \tag{59}$$

It can be shown, with the help of Jensen's Inequality, that p_n is upper bounded by

$$p_n \leq \left(1-(1-p)^{(M-i)m_1}\right)^{n-1}. \tag{60}$$

Since $(1-(1-p)^{(M-i)m_1}) < 1$, we have, from (57)

$$\begin{aligned} E\left[\xi^{(i)}\right] \leq& (M-i)m_1 \sum_{n=1}^{\infty} n\left(1-(1-p)^{(M-i)m_1}\right)^{n-1} \\ =& \frac{(M-i)m_1}{(1-p)^{2(M-i)m_1}}. \end{aligned} \tag{61}$$

Thus, from (56) and (61), we have

$$E\left[\tau_0^{(i)}\right] \leq 1 + i \sum_{r=i}^{M-1} \frac{(M-r)m_1}{(1-p)^{2(M-r)m_1}}. \tag{62}$$

With (54), we then have

$$m_2 \leq 1 + (M-1)m_1 + \max_{i=1,\ldots,M-1} i \sum_{r=i}^{M-1} \frac{(M-r)m_1}{(1-p)^{2(M-r)m_1}} \tag{63}$$

which, combining with (49) and (53), leads to an upper bound of D.

References

[1] N. Abramson, "The ALOHA system—Another alternative for computer communications," in *Proc. Fall Joint Comput. Conf., AFIPS Conf.*, 1970, p. 37.

[2] ——, "The throughput of packet broadcasting channels," *IEEE Trans. Commun.*, vol. COM-25, pp. 117–128, Jan. 1977.

[3] D. E. Ayyildiz and H. Delic, "Adaptive random-access algorithm with improved delay performance," *Int. J. Commun. Syst.*, vol. 14, pp. 531–539, 2001.

[4] J. Q. Bao and L. Tong, "A performance comparison between *ad hoc* and centrally controlled CDMA wireless LANs," *IEEE Trans. Wireless Commun.*, vol. 1, pp. 829–841, Oct. 2002.

[5] J. I. Capetanakis, "Generalized TDMA: The multiaccessing tree protocol," *IEEE Trans. Commun.*, vol. COM-27, pp. 1476–1484, Oct. 1979.

[6] ——, "Tree algorithms for packet broadcast channels," *IEEE Trans. Inform. Theory*, vol. IT-25, pp. 505–515, Sept. 1979.

[7] I. Chlamtac and A. Farago, "An optimal channel access protocol with multiple reception capacity," *IEEE Trans. Comput.*, vol. 43, pp. 480–484, Apr. 1994.

[8] A. Chockalingam, M. Zorzi, L. B. Milstein, and P. Venkataram, "Performance of a wireless access protocol on correlated Rayleigh fading channels with capture," *IEEE Trans. Commun.*, vol. 46, pp. 644–655, May 1998.

[9] G. del Angel and T. L. Fine, "Randomized power control strategies for optimization of multiple access radio systems," in *Proc. 38th Allerton Conf. Communication, Control and Computing*, Oct. 2000.

[10] R. Gallager, "Conflict resolution in random-access broadcast networks," in *Proc. AFOSR Workshop Communications Theory and Applications*, Sept. 1978, pp. 74–76.

[11] S. Ghez, S. Verdú, and S. C. Schwartz, "Stability properties of slotted ALOHA with multipacket reception capability," *IEEE Trans. Automat. Contr.*, vol. AC-33, pp. 640–649, July 1988.

[12] ——, "Optimal decentralized control in the random-access multipacket channel," *IEEE Trans. Automat. Contr.*, vol. 34, pp. 1153–1163, Nov. 1989.

[13] I. M. I. Habbab *et al.*, "ALOHA with capture over slow and fast fading radio channels with coding and diversity," *IEEE J. Select. Areas Commun.*, vol. 7, pp. 79–88, Jan. 1989.

[14] M. G. Hluchyj, "Multiple access window protocol: Analysis for large finite populations," in *Proc. IEEE Conf. Decision and Control*, New York, 1982, pp. 589–595.

[15] M. G. Hluchyj and R. G. Gallager, "Multiaccess of a slotted channel by finitely many users," in *Proc. Nat. Telecommunications Conf.*, New Orleans, LA, Aug. 1981, pp. D4.2.1–D4.2.7.

[16] S. Kim and J. Yeo, "Optimal scheduling in CDMA pakcet radio networks," *Comput. Oper. Res.*, vol. 25, pp. 219–227, Mar. 1998.

[17] L. Kleinrock and Y. Yemini, "An optimal adaptive scheme for multiple access broadcast communication," in *Proc. Int. Conf. Communications*, June 1978, pp. 7.2.1–7.2.5.

[18] D. F. Lyons and P. Papantoni-Kazakos, "A window random-access algorithm for environments with capture," *IEEE Trans. Commun.*, vol. 37, pp. 766–770, July 1989.

[19] J. J. Metzner, "On improving utilization in ALOHA networks," *IEEE Trans. Commun.*, vol. COM-24, pp. 447–448, Apr. 1976.

[20] M. Paterakis and P. Papantoni-Kazakos, "A simple window random-access algorithm with advantageous properties," *IEEE Trans. Inform. Theory*, vol. 35, pp. 1124–1130, Sept. 1989.

[21] L. G. Roberts, ALOHA packet system with and without slots and capture, in *ASS Note 8*, Stanford Res. Inst., Adv. Res. Projects Agency, Network Inform. Ctr., Stanford, CA, 1972.

[22] F. Shad, T. D. Todd, V. Kezys, and J. Litva, "Dynamic slot allocation (DSA) in indoor SDMA/TDMA using a smart antenna basestation," *IEEE/ACM Trans. Networking*, vol. 9, pp. 69–81, Feb. 2001.

[23] M. Sidi and I. Cidon, "Splitting protocols in presence of capture," *IEEE Trans. Inform. Theory*, vol. IT-31, pp. 295–301, Mar. 1985.

[24] C. Vanderplas and J. P. M. Linnartz, "Stability of mobile slotted ALOHA network with Rayleigh fading, shadowing, and near-far effect," *IEEE Trans. Veh. Technol.*, vol. 39, pp. 359–366, Nov. 1990.

[25] B. Yucel and H. Delic, "Mobile radio window random-access algorithm with diversity," *IEEE Trans. Veh. Technol.*, vol. 49, pp. 2060–2070, Nov. 2000.

[26] Q. Zhao and L. Tong, "Semi-blind collision resolution in random-access wireless *ad hoc* networks," *IEEE Trans. Signal Processing*, vol. 48, pp. 2910–2920, Oct. 2000.

[27] M. Zorzi, "Mobile radio slotted ALOHA with capture and diversity," *Wireless Networks*, vol. 1, pp. 227–239, May 1995.

[28] M. Zorzi and R. R. Rao, "Capture and retransmission control in mobile radio," *IEEE J. Select. Areas Commun.*, vol. 12, pp. 1289–1298, Oct. 1994.

Qing Zhao (S'97–M'02) received the B.S. degree from Sichuan University, Chengdu, China, in 1994, the M.S. degree from Fudan University, Shanghai, China, in 1997, and the Ph.D. degree from Cornell University, Ithaca, NY, in 2001, all in electrical engineering.

She was a Communication System Engineer with Aware, Inc., Bedford, MA, in 2001–2002. She is currently a Postdoctoral Research Associate with the School of Electrical and Computer Engineering, Cornell University, Ithaca, NY. Her research interests include signal processing for communication systems, protocol design for wireless communication networks, and DSL technology.

Dr. Zhao received the IEEE Signal Processing Society Young Author Best Paper Award in 2000.

Lang Tong (S'87–M'91–SM'01) received the B.E. degree from Tsinghua University, Beijing, China, in 1985, and the M.S. and Ph.D. degrees in electrical engineering from the University of Notre Dame, Notre Dame, IN, in 1987 and 1990, respectively.

He was a Postdoctoral Research Affiliate with the Information Systems Laboratory, Stanford University, Stanford, CA, in 1991. Currently, he is an Associate Professor with the School of Electrical and Computer Engineering, Cornell University, Ithaca, NY. His areas of interest include statistical signal processing, adaptive receiver design for communication systems, signal processing for communication networks, and information theory.

Dr. Tong received the Young Investigator Award from the Office of Naval Research in 1996 and the Outstanding Young Author Award from the IEEE Circuits and Systems Society.