A Multi-Queue Service Room MAC Protocol for Wireless Networks with Multipacket Reception

*Qing Zhao and Lang Tong**
School of Electrical and Computer Engineering
384 Frank H.T. Rhodes Hall
Cornell University
Ithaca, NY 14853
Tel:(607) 255-3500. Fax: (607) 255-9072
Email: ltong@ee.cornell.edu

* Corresponding Author

Abstract

An adaptive medium access control (MAC) protocol for heterogeneous networks with finite population is proposed. Referred to as the Multi-Queue 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.

Keywords

Medium Access Control. Random Access Network. Multipacket Reception.

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], [14], the tree algorithm [5], the first-come first-serve (FCFS) algorithm [7], and a class of adaptive schemes [12], [4], [11], [10], 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 CDMA and/or antenna array, multiuser detection techniques, and signal processing based collision resolution algorithms [16].

A general model for channels with multipacket reception (MPR) capability has been developed in [8], [9], where the MPR capability is characterized by the probability $C_{n,k}$ of successfully receiving k packets when total n are transmitted. Based on this channel model, the authors of

[8], [9] analyzed the performance of slotted ALOHA in networks with infinite population. As a special form of MPR, the impact of capture on the slotted ALOHA and the FCFS algorithm was studied in [13], [2], [6] and [15].

Although many protocols originally proposed for the conventional collision channel can be adapted to an MPR channel, two unique features of MPR channels make the optimality of their extension questionable. First, the MPR capability enriches the channel outcome, which makes it more difficult to infer the state of a user from the feedback information, where we define the state of a user as the number of packets held by this user (under the single buffer assumption, a user is either in state 0 (idle) or state 1 (active)). For example, in the conventional collision channel, a successfully received packet implies that all users who were enabled to access the channel simultaneously with the source of this packet are idle. However, when a packet is successfully received in an MPR channel, the number of simultaneously transmitted packets can be any n that is no larger than the number of enabled users and satisfies $C_{n,1} > 0$. Second, the MPR capability opens new options for resolving packet collision. In the conventional channel, a collision has to be resolved by splitting users in a certain way, either dividing users into several subsets (as in the tree algorithm) or setting a retransmission probability which is smaller than 1 (as in ALOHA). For an MPR channel, splitting is not always necessary or even sensible for collision resolution. Take, for example, an MPR channel with $C_{2,0} = \epsilon$ and $C_{2,2} = 1 - \epsilon$ where $0 < \epsilon \ll 1$. When two packets are simultaneously transmitted and none successfully received, in stead of splitting, the same set of users should be enabled again to fully exploit the channel MPR capability. Combining these two points, we see that two issues are crucial to the efficient utilization of an MPR channel: (1) how to infer the state of users from previous channel outcomes; (2) based on the inferred user state, whether and how splitting of users should be performed for a given MPR channel.

In this paper, we propose a MAC protocol which provides optimal answers to these questions. We consider a slotted network with MPR capability and a finite population of users. Users may have different QoS requirements which are characterized by the 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. Our goal is to design a MAC protocol that maximizes the per-slot throughput (the expected number of successfully received packets in each slot) while ensuring each user's QoS requirement. To achieve this, the state of each user is updated at the beginning of each slot by optimally exploiting the information provided by previous channel outcomes. 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. This access set may be a subset of users who have accessed the channel in the previous slot, corresponding to an optimal splitting of users in terms of per-slot throughput. 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-A, we present the model of a communication network with heterogeneous delay constraints and MPR capability. In Section II-B, the problems of satisfying heterogeneous delay constraints, inferring state of users from channel outcomes, and determining the optimal access set are addressed. In Section III, we propose the multi-queue service room (MQSR) protocol. Simulation examples are presented in Section IV, where the throughput and delay performance of the MQSR protocol is compared to that of the URN scheme [12] and the slotted ALOHA with optimal retransmission probability.

II. The Problem Statements

A. The Model

We consider a communication network with M users who transmit data to a central controller through a common wireless channel. 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. A packet generated at the beginning of a slot may be transmitted in this slot, and a successfully transmitted packet leaves its buffer. Packets generated by a user with a full buffer are assumed lost.

Users are partitioned into L groups according to their QoS constraints. The M_l ($l = 1, \cdots, L, \sum_{l=1}^{L} M_l = M$) users in the lth group require their packet delay at $p = 1$ no greater than d_l, where we define packet delay as the expected number of slots from the time a packet enters a buffer until the end of its successful transmission.

As considered in [8], [9], [3], 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

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

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

$$\eta \stackrel{\Delta}{=} \max_{n=1,\cdots,M} \mathcal{C}_n, \tag{2}$$

where

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

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

$$n_0 \stackrel{\Delta}{=} \min\{\arg \max_{n=1,\cdots,M} \mathcal{C}_n\}. \tag{4}$$

We can see that at heavy traffic load, n_0 packets should be transmitted simultaneously to achieve the channel capacity η. 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 with capture. The reception matrix of the conventional collision channel and channels with capture are given by

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

where p_i is the probability of capture given i simultaneous transmissions. 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 given below.

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

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

We assume that the central controller can distinguish without error between empty and nonempty slots. Furthermore, if some packets are successfully demodulated at the end of a slot, the central controller can identify the source of these packets and inform their sources so that their buffers can be released. However, if at least one packet is successfully demodulated at the end of a slot, 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 Figure 1 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 that happened in the third and the forth slot are indistinguishable.

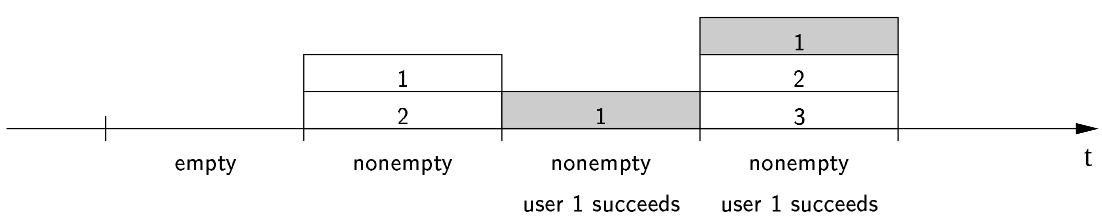


Fig. 1: Possible outcomes of a slot

figure

B. The Problem

Our goal here is to design, for general MPR channels, a MAC protocol to be carried out at the centrol controller. Specifically, without the state information of users, the centrol 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 $\mathcal{A}(t)$ and (only) users in $\mathcal{A}(t)$ access the channel if they have packets to transmit. At the end of slot t, the centrol controller observes the channel outcome $F(t)$ which contrains information on whether slot t is empty and whose packets are successfully recived in slot t (see Figure 1). The sources of successfully recieved packets are notified so that they can release their buffer and generate new packets. Users who transmitted but didn't receive acknowledgement assume their packets were lost and will retransmit the next time they are enabled. In this paper, we assume that the down link channel (from the centrol controller to the users) is error free and the time for acknowlegement and broadcasting $\mathcal{A}(t)$ is neglible. The problem here is given the traffic load p and the channel reception matrix $\mathbf{C}$, design a protocol for choosing the access set for each slot so that the channel reception capability is fully utilized and each user's delay requirement is satisfied.

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. In Section II-C, we give a necessary and sufficent condition for the existance of a MAC protocol that ensures a given set of delay requirements.

C. The Heterogeneous Delay Constraints

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 can not 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 constraints being achievable.

Proposition 1: Let M_l $(l = 1, \cdots, L)$ be the number of users who require their packet delay at $p = 1$ no larger than d_l. Then for the network model specified in Section II-A, 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 III 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 following relation between the throughput $T_l(p)$ and the delay $D_l(p)$ under the equilibrium condition:

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

A proof of (8) following [11] is provided in Appendix A. At $p = 1$, we have

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

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

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

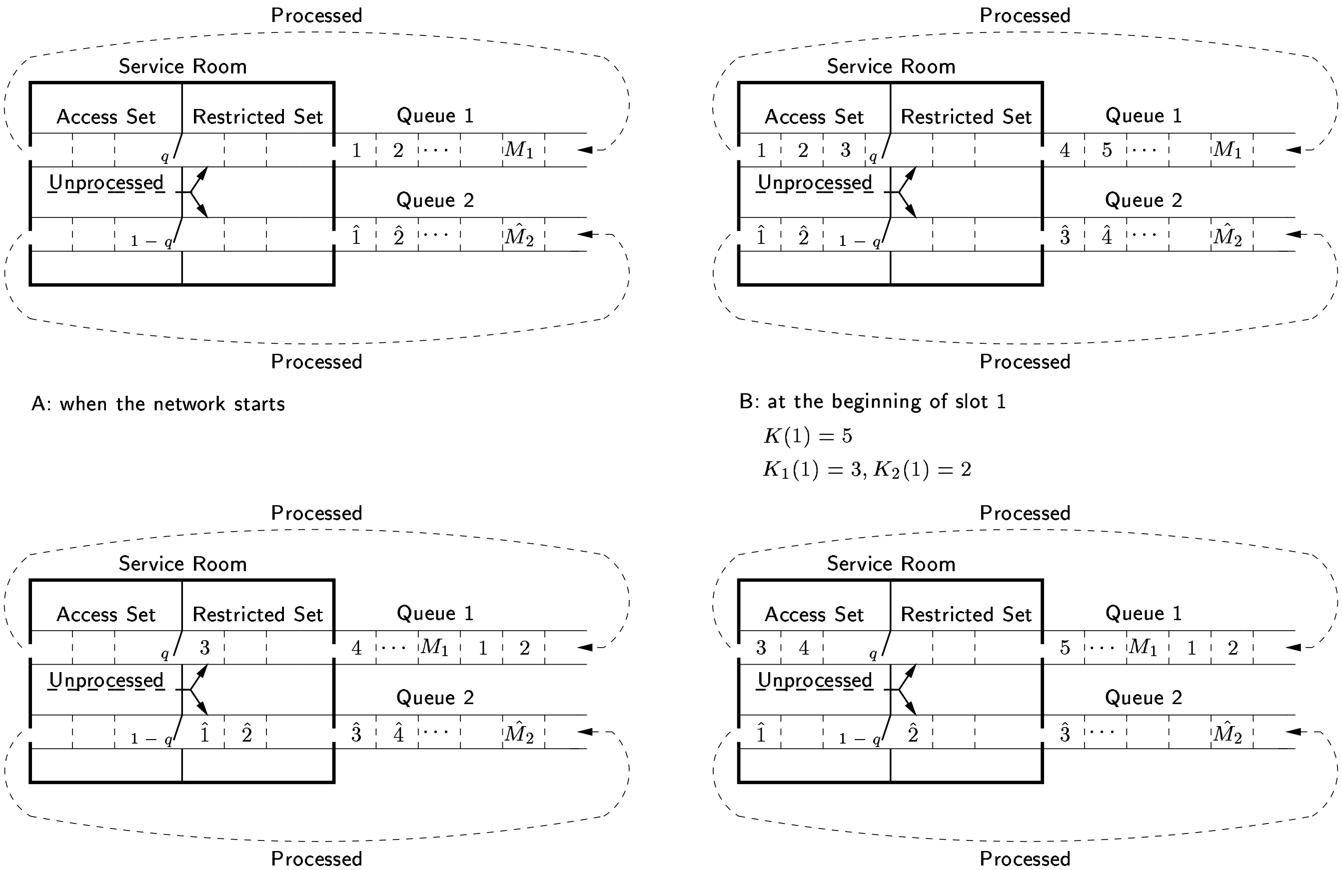
□□□

III. The Multi-Queue Service Room Protocol

A. The Basic Structure

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 $D_1(1) \leq d_1$ 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 Figure 2, where users from the first group are indexed by i $(i = 1, \cdots, M_1)$ and those from the second by $\hat{i}$ $(i = 1, \cdots, M_2)$. As shown in Figure 2-A, when the network starts, users of the two groups are waiting, respectively,



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. 2: The basic procedure of the multi-queue service room protocol.

figure

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 restricted set. Users in the access set transmit, in the current slot, packets generated before entering the service room while users in the restricted set can not 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 can not 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 the beginning of the restricted set. The central controller then chooses the access set for slot $t+1$ by specifying the size $K(t+1)$ ($1 \leq K(t+1) \leq M$) 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 restricted set or in the waiting 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

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

For $k = M$, we have

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

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 restricted set) 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 restricted set, 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 restricted set that are from the lth group enter the access set while the last $\alpha_l(t) - K_l(t+1)$ users remain in the restricted set.

We now consider the example in Figure 2 to illustrate the procedure of the MQSR protocol. Suppose that at the beginning of the first slot (Figure 2-B), the central controller decides that $K(1) = 5$ with $K_1(1) = 3$ and $K_2(1) = 2$. Hence user $1, 2, 3$, and $\hat{1}, \hat{2}$ form the access set and transmit their packets (if any) generated before entering the service room. At the end of this slot (Figure 2-C), the central controller successfully receives two packets from user 1 and 2, respectively. We thus have $\mathcal{P}(1) = \{1, 2\}$; these two users leave the service room and go to the end of their queue. The unprocessed users $3, \hat{1}, \hat{2}$ go to the restricted set. At the beginning of the second slot (Figure 2-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 $\hat{2}$ remain in the restricted set. 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 $\alpha_2(2) = 1$.

After specifying the basic structure of the MQSR protocol, we now consider parameters that remain to be designed. 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, $K(t)$, and $\mathcal{P}(t)$ is formulated in Section III-B with its solution detailed in Section III-C to III-E.

We point out that the optimal window protocol proposed in [11] has a similar structure as the MQSR protocol at $L = 1$. Relying on exhaustive search, however, the window protocol is only computationally feasible for networks with 2 or 3 users and no MPR. Furthermore, homogeneous QoS constraints are assumed in [11].

B. The Problem Formulation

We now attack the problem of estimating the state of users and choosing the optimal access set under the structure of the MQSR protocol specified in Section III-A.

B.1 The Optimal Estimate of User State

The state of each user at the beginning of each slot (after new packet generation) is the most crucial information for optimal channel accessing. Had this information been known to the central controller, perfect scheduling of transmission could be performed. Without the knowledge of every user's state, an optimal MAC protocol should be built on an optimal estimate of each user's state.

Let $X_i^{(l)}(t)$ be the state of the ith $(i = 1, \cdots, M_l)$ user of the lth $(l = 1, 2)$ group at the beginning of slot t (after new packet generation). Under the single buffer assumption, $X_i^{(l)}(t)$ is a random variable with possible values 0 and 1. We now consider the information available for estimating $X_i^{(l)}(t)$ $(l = 1, 2,\ i = 1, \cdots, M_l)$.

Let $\mathcal{A}(t)$ denote the access set for slot t. Let $F(t)$ be the channel outcome of slot t, which contains information on whether slot t is empty and whose packets are successfully received in slot t (see Figure 1). With $\mathbf{C}$ and p given, the information, denoted by $I_{[1,t-1]}$, available at the beginning of slot t is the initial condition of the network in the form of the distribution of $X_i^{(l)}(1)$ $(l = 1, 2,\ i = 1, \cdots, M_l)$, the access sets $\mathcal{A}(1), \cdots, \mathcal{A}(t-1)$, and the channel outcomes $F(1), \cdots, F(t-1)$. The distribution of $X_i^{(l)}(t)$ $(l = 1, 2,\ i = 1, \cdots, M_l)$ conditioned on $I_{[1,t-1]}$ could provide an estimate of each user's state. However, the marginal distribution of $X_i^{(l)}(t)$ $(l = 1, 2,\ i = 1, \cdots, M_l)$ conditioned on $I_{[1,t-1]}$ does not capture all information provided by $I_{[1,t-1]}$. After accessing the channel simultaneously in one slot, two users' states conditioned on the outcome of this slot become correlated despite the independence of their traffic generation. Thus, it is the joint distribution of $X_i^{(l)}(t)$ $(l = 1, 2,\ i = 1, \cdots, M_l)$ conditioned on $I_{[1,t-1]}$ that characterizes the state of the whole network at the beginning of slot t. Let $P_{I_{[1,t-1]}}\{X_i^{(l)}(t), l = 1, 2, i = 1, \cdots, M_l\}$ denote the joint distribution of $X_i^{(l)}(t)$ $(l = 1, 2,\ i = 1, \cdots, M_l)$ conditioned on $I_{[1,t-1]}$. We show in Section III that $P_{I_{[1,t-1]}}\{X_i^{(l)}(t), l = 1, 2, i = 1, \cdots, M_l\}$ can be obtained recursively from $P_{I_{[1,t-2]}}\{X_i^{(l)}(t-1), l = 1, 2, i = 1, \cdots, M_l\}$ by incorporating information, specifically, $\mathcal{A}(t-1)$ and $F(t-1)$, obtained from slot $t-1$.

For the convenience of presentation, we relabel users in each group at the beginning of each slot, starting from the service room to the end of the lth queue. Due to the time constraint imposed on a packet being eligible for transmission, the state $X_i^{(l)}(t)$ of the ith $(i = 1, \cdots, M_l)$ user in the lth $(l = 1, 2)$ group when it is inside the service room is redefined as the number of packets generated before its entering the service room. When it is waiting in the queue, $X_i^{(l)}(t)$ still denotes the number of packets in its buffer at the beginning of slot t.

To determine the processed set and the optimal access set, the conditional joint distribution $P_{I_{[1,t-1]}}\{X_i^{(l)}(t), l = 1, 2, i = 1, \cdots, M_l\}$ needs to be computed at the beginning of slot t. This computation is simplified by the time control imposed on a packet being eligible for transmission and a user being eligible for leaving the service room. Specifically, 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 $P_{I_{[1,t-1]}}\{X_i^{(l)}(t), l = 1, 2, i = 1, \cdots, M_l\}$ 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. Furthermore, 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 complexity of computing $P_{I_{[1,t-1]}}\{X_i^{(l)}(t), l = 1, 2, i = 1, \cdots, M_l\}$. We detail this computation in Section III-D and III-E.

B.2 The Optimal Access Set

At the beginning of slot t, the access set $\mathcal{A}(t)$ need to be determined based on the current estimate of each user's state. The criterion we use for determining access set 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 $I_{[1,t-1]}$ depends on $\mathcal{A}(t)$ and the channel MPR matrix $\mathbf{C}$. Let $\mathcal{Q}_M$ denote the set consisting of all nonempty subsets of the M users in the network. We then have

$$\mathcal{A}(t) = \arg\max_{\mathcal{A}\in\mathcal{Q}_M} E_{I_{[1,t-1]}}[S(t) \mid \mathcal{A}] \ \text{ subject to } D_l(1) \leq d_l \text{ for } l = 1, \cdots, L, \tag{13}$$

where $E_{I_{[1,t-1]}}[S(t)]$ is a shorthand for $E[S(t) \mid I_{[1,t-1]}]$. For a given access set $\mathcal{A}$, $E_{I_{[1,t-1]}}[S(t)]$ can be computed as

$$E_{I_{[1,t-1]}}[S(t) \mid \mathcal{A}] = \sum_{n=1}^{|\mathcal{A}|} \mathcal{C}_n P[Y_{\mathcal{A}} = n \mid I_{[1,t-1]}], \tag{14}$$

where $Y_{\mathcal{A}}$ is the total number of packets held by users in $\mathcal{A}$, and its distribution conditioned on $I_{[1,t-1]}$ can be obtained from $P_{I_{[1,t-1]}}\{X_i^{(l)}(t), l = 1, 2, i = 1, \cdots, M_l\}$.

To gain insights into the optimization problem given in (13), we temporarily drop the delay constraints and assume that every user has the same QoS requirement ($L = 1$). In this case, the optimal access set for slot t can be obtained by computing $E_{I_{[1,t-1]}}[S(t)]$ for all possible access sets and choose the one that gives the maximum. The difficulty with this solution is that the number of possible access sets (total $2^M - 1$ of them) increases exponentially with M. To reduce the computational complexity, we consider all M users are waiting in a queue for channel access and the access set for each slot contains consecutive users starting from the first one in the queue. With this queue structure, the number of possible access sets reduces to M. However, this structure produces unfair channel access; the first user in the queue is always granted access while the last one has the fewest chances to access the channel. To ensure the fairness of channel access among all users, we impose a time control on packets that are eligible for transmission. A user in the access set can only transmit its packet generated before a certain time instant. When its packet generated before this time instant has been successfully transmitted or the central controller detects that it does not have a packet generated before the specified time instant, this user leaves the access set and goes to the end of the queue. With this time control strategy and the circular movement of users in the queue, we ensure fair channel access and prevent the situation where a user who keeps generating new packets seizes the channel. Furthermore, as shown in Section III, by carefully specifying the time instant, this control strategy also simplifies the update of the joint distribution of users' states at the beginning of each slot.

Recall that $\mathcal{A}(t)$ denotes the access set for slot t. With the multi-queue structure of the protocol and the relabeling of users at the beginning of each slot, we have

$$\mathcal{A}(t) = \{1, \cdots, K_1(t)\} \bigcup \{\hat{1}, \cdots, \hat{K}_2(t)\}, \tag{15}$$

which is obtained by choosing $K(t)$ users from the two groups with a priority factor q. Hence, the problem of determining the optimal access set given in (13) reduces to

$$\{q, K(t)\} = \arg \max_{k=1,\cdots,M} E_{I_{[1,t-1]}}[S(t) \mid K(t) = k] \ \text{ subject to } D_1(1) = d_1. \tag{16}$$

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 $E_{I_{[1,t-1]}}[S(t)]$ is maximized. This decoupling is based on the fact that the maximization of $E_{I_{[1,t-1]}}[S(t)]$ 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 $E_{I_{[1,t-1]}}[S(t)]$ 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,

$$\begin{aligned} K(t) &= \arg \max_{K(t)=1,\cdots,M} E_{I_{[1,t-1]}}[S(t)] \\ &= \arg \max_{K(t)=1,\cdots,M} \mathcal{C}_{K(t)} \\ &= n_0, \end{aligned} \tag{17}$$

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

$$T(1) = \eta. \tag{18}$$

□□□

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 (16). As shown in Section III-C, 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 $p = 1$.

C. The 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

$$q = \frac{M_1}{d_1 \eta}. \tag{19}$$

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 (17)) 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 i.i.d. sequences. Thus, we have, at $p = 1$,

$$E[S(t)] = \eta, \quad E[S_1(t)] = T_1(1), \quad E[K_1(t)] = qn_0, \tag{20}$$

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,

$$E[S_1(t) \mid K_1(t) = u, S(t) = s] = us/n_0, \tag{21}$$

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. Average over all the realizations of $K_1(t)$ and $S(t)$, and consider the independence between $K_1(t)$ and $S(t)$, we get

$$\begin{aligned} E[S_1(t)] &= \frac{1}{n_0} E[K_1(t) S(t)] \\ &= \frac{1}{n_0} E[K_1(t)] E[S(t)] \\ &= q\eta, \end{aligned} \tag{22}$$

which, along with (20), leads to

$$T_1(1) = q\eta. \tag{23}$$

Combining with (9), we have

$$D_1(1) = \frac{M_1}{q\eta}. \tag{24}$$

To ensure $D_1(1) = d_1$, q should be determined by (19).

□□□

When the condition of $M_l \geq n_0$ $(l = 1, 2)$ is violated, $K_1(t)$ given $K(t) = n_0$ no long has a binomial distribution and the last equality in (20) does not hold. However, from the distribution given in (11,12), 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 the following equation:

$$E[K_1(t) \mid K(t) = n_0] = \frac{n_0 M_1}{d_1 \eta}. \tag{25}$$

D. The Determination of $\mathcal{P}(t)$

At the end of slot t, the central controller needs to determine the processed set $\mathcal{P}(t)$. By definition, $\mathcal{P}(t)$ is a subset of $\mathcal{A}(t)$. It consists of users whose packets are successfully transmitted in slot t and users who are detected by the central controller as having an empty buffer upon entering the service room. 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. Let $X_i^{(l)}(t^+)$ $(l = 1, 2, i = 1, \cdots, K_l(t))$ be the number of packets held by the ith user of the lth group at t^+, excluding packets generated during its current visit to the service room. We then have

$$\begin{aligned} \mathcal{P}(t) = \quad & \{i: \ 1 \leq i \leq K_1(t), \ E_{I_{[1,t]}}[X_i^{(1)}(t^+)] = 0\} \\ & \bigcup \{\hat{i}: \ 1 \leq i \leq K_2(t), \ E_{I_{[1,t]}}[X_i^{(2)}(t^+)] = 0\}. \end{aligned} \tag{26}$$

We now evaluate $E_{I_{[1,t]}}[X_i^{(l)}(t^+)]$. If slot t is empty, we readily have

$$E_{I_{[1,t]}}[X_i^{(l)}(t^+)] = 0, \quad l = 1, 2, \ i = 1, \cdots, K_l(t), \tag{27}$$

and

$$\mathcal{P}(t) = \mathcal{A}(t). \tag{28}$$

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, we have, by the single buffer assumption, $E_{I_{[1,t]}}[X_i^{(l)}(t^+)] = 0$ for $(i, l) \in \{(i, l) : s_i^{(l)} = 1\}$. However, $s_i^{(l)} = 1$ is a sufficient but not necessary condition for the ith user in the lth group being processed. Besides the outcome of slot t, the channel reception matrix $\mathbf{C}$ also provides information on $X_i^{(l)}(t^+)$. Consider, for example, the conventional channel. A successful transmission of one user implies that other users in the access set do not have packets. To identify all processed users, we compute the joint distribution of $\{X_i^{(l)}(t^+),\ l=1,2,\ i=1,\cdots,N_l(t)\}$ conditioned on $I_{[1,t]}$, 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 restricted set) in slot t. For $0 \le x_i^{(l)} \le 1-s_i^{(l)}$ $(l=1,2,\ i=1,\cdots,N_l(t))$, we have,

$$P[\{X_i^{(l)}(t^+)=x_i^{(l)},\ l=1,2,\ i=1,\cdots,N_l(t)\} \mid I_{[1,t]}]$$

$$= \frac{P[\{X_i^{(l)}(t^+)=x_i^{(l)},\ l=1,2,\ i=1,\cdots,N_l(t)\},\ F(t) \mid \mathcal{A}(t),\ I_{[1,t-1]}]}{P[F(t) \mid \mathcal{A}(t),\ I_{[1,t-1]}]}$$

$$= \frac{z!/(S!(z-S)!)C_{z,S}P[\{X_i^{(l)}(t)=x_i^{(l)}+s_i^{(l)},\ l=1,2,\ i=1,\cdots,N_l(t)\} \mid I_{[1,t-1]}]}{\sum_{\{0\le x_i^{(l)}\le 1-s_i^{(l)},l=1,2,\ i=1,\cdots,N_l(t)\}} z!/(S!(z-S)!)C_{z,S}P[\{X_i^{(l)}(t)=x_i^{(l)}+s_i^{(l)},\ l=1,2,\ i=1,\cdots,N_l(t)\} \mid I_{[1,t-1]}]}, \tag{29}$$

where $z=\sum_{l=1}^{2}\sum_{i=1}^{K_l(t)}(x_i^{(l)}+s_i^{(l)})$ and $S=\sum_{l=1}^{2}\sum_{i=1}^{K_l(t)} s_i^{(l)}$ is the total number of successfully transmitted packets in slot t.

In (29), the joint distribution of $\{X_i^{(l)}(t),\ l=1,2,\ i=1,\cdots,N_l(t)\}$ conditioned on $I_{[1,t-1]}$ is assumed known at the end of slot t. As shown in (34), this joint distribution is obtained recursively from the conditional joint distribution of $\{X_i^{(l)}(t-1)^+,\ l=1,2,\ i=1,\cdots,N_l(t-1)\}$ and available at the beginning of slot t. At the beginning of the first slot, the joint distribution of $\{X_i^{(l)}(1),\ l=1,2,\ i=1,\cdots,N_l(1)\}$ is obtained, based on the assumption of independent traffic generation, from the initial condition of the network given as the marginal distribution of $X_i^{(l)}(1)$ $(l=1,2,\ i=1,\cdots,M_l)$.

With the conditional joint distribution of $\{X_i^{(l)}(t^+),\ l=1,2,\ i=1,\cdots,N_l(t)\}$, we can then evaluate $E_{I_{[1,t]}}[X_i^{(l)}(t^+)]$ for $l=1,2$, $i=1,\cdots,K_l(t)$ and obtain $\mathcal{P}(t)$.

E. The Determination of $K(t+1)$

At the end of slot t, after determining the processed users and removing them from the service room, we choose the access set $\mathcal{A}(t+1)$ by specifying $K(t+1)$.

As shown in (16), $K(t+1)$ is obtained by maximizing $E_{I_{[1,t]}}[S(t+1)]$ with q determined by (25), *i.e.*,

$$K(t+1) = \arg\max_{k=1,\cdots,M} E_{I_{[1,t]}}[S(t+1) \mid K(t+1)=k], \tag{30}$$

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

$$E_{I_{[1,t]}}[S(t+1) \mid K(t+1)=k] = \sum_{k_1=\max(0,k-M_2)}^{\min(k,M_1)} P[K_1(t+1)=k_1 \mid K(t+1)=k] \quad (31)$$

$$E_{I_{[1,t]}}[S(t+1) \mid K_1(t+1)=k_1, K_2(t+1)=k-k_1] \quad (32)$$

with

$$\begin{aligned} & E_{I_{[1,t]}}[S(t+1) \mid K_1(t+1)=k_1, K_2(t+1)=k-k_1] \\ & = \textstyle\sum_{n=1}^{k} \mathcal{C}_n P[\sum_{i=1}^{k_1} X_i^{(1)}(t+1) + \sum_{i=1}^{k-k_1} X_i^{(2)}(t+1) = n \mid I_{[1,t]}]. \end{aligned} \quad (33)$$

To obtain $P[\sum_{i=1}^{k_1} X_i^{(1)}(t+1) + \sum_{i=1}^{k-k_1} X_i^{(2)}(t+1) = n \mid I_{[1,t]}]$ for all possible k and k_1, we need the conditional joint distribution of $\{X_i^{(l)}(t+1),\ l=1,2,\ i=1,\cdots,M_l\}$, which can be computed 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 $t+1$ can not be transmitted until the next time they enter the service room, we have $X_i^{(l)}(t+1) = X_i^{(l)}(t^+)$ for $l=1,2,\ i=1,\cdots,\alpha_l(t)$. Hence, the conditional joint distribution of $\{X_i^{(l)}(t+1),\ l=1,2,\ i=1,\cdots,\alpha_l(t)\}$ can be easily obtained from the conditional joint distribution of $\{X_i^{(l)}(t^+),\ l=1,2,\ i=1,\cdots,N_l(t)\}$ given by (29) by summing over all possible values taken by $X_i^{(l)}(t^+)$ ($l=1,2,\ i=\alpha_l(t)+1,\cdots,N_l(t)$), *i.e.*,

$$\begin{aligned} & P[\{X_i^{(l)}(t+1) = x_i^{(l)},\ l=1,2, i=1,\cdots,\alpha_l(t)\} \mid I_{[1,t]}] \\ & = \textstyle\sum_{\{x_i^{(1)},\ x_j^{(2)},\ i,\hat{j} \in \mathcal{P}(t)\}} P[\{X_i^{(l)}(t^+) = x_i^{(l)},\ l=1,2, i=1,\cdots,N_l(t)\} \mid I_{[1,t]}]. \end{aligned} \quad (34)$$

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

$$P[X_i^{(l)}(t+1)=x] = \begin{cases} (1-p)^{W_i^{(l)}(t+1)} & \text{if } x=0 \\ 1-(1-p)^{W_i^{(l)}(t+1)} & \text{if } x=1 \\ 0 & \text{otherwise} \end{cases}, \quad (35)$$

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 time entered the service room or the index of the slot in which this user last time successfully transmitted a packet, whichever is larger.

By the independence of traffic generation among all users, the conditional joint distribution of $\{X_i^{(l)}(t+1),\ l=1,2,\ i=1,\cdots,M_l\}$ can be obtained as the product of the conditional joint distribution of $\{X_i^{(l)}(t+1),\ l=1,2,\ i=1,\cdots,\alpha_l(t)\}$ given in (34) and the marginal distribution of $X_i^{(l)}(t+1)$ $(l=1,2,\ i=\alpha_l(t)+1,\cdots,M_l(t))$ given in (35). With this joint distribution, $E_{I_{[1,t]}}[S(t+1) \mid K(t+1)=k]$ can be computed 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. We summarize the basic procedure of the MQSR protocol in Figure 3.

IV. 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 [12] and the slotted ALOHA with optimal retransmission probability. As shown in [11], for a network model specified in Section II-A, 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,

The Multi-Queue Service Room Protocol

Initialization:

1. Choose q as given in (19).
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,
   - empty slot: $\mathcal{P}(t) = \mathcal{A}(t)$.
   - nonempty slot and $s_i^{(l)}$ packets from the ith user in the lth group are successfully received:
     (a) compute $P_{I_{[1,t]}}\{X_i^{(l)}(t^+), l = 1, 2, i = 1, \cdots, N_l(t)\}$ as given in (29).
     (b) obtain $\mathcal{P}(t)$ by evaluating $E_{I_{[1,t]}}[X_i^{(l)}(t^+)]$ $(l = 1, 2, i = 1, \cdots, K_l(t))$.
     (c) compute $P_{I_{[1,t]}}\{X_i^{(l)}(t+1), l = 1, 2, i = 1, \cdots, \alpha_l(t)\}$ as given in (34).
3. Compute the marginal of $X_i^{(l)}(t+1)$ $(l = 1, 2, i = \alpha_l(t) + 1, \cdots, M_l)$ as given in (35).
4. Obtain $K(t+1)$ by solving (30).
5. Determine $\mathcal{A}(t+1)$ by choosing $K_1(t+1)$ and $K_2(t+1)$.
6. Set $N_l(t+1) = \max(K_l(t+1), \alpha_l(t+1))$.
7. Obtain $P_{I_{[1,t]}}\{X_i^{(l)}(t+1), l = 1, 2, i = 1, \cdots, N_l(t+1)\}$.
8. Set $t = t + 1$.

Fig. 3: The multi-queue service room protocol.

figure

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 is obtained by simulations while that of the slotted ALOHA is 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 Figure 4, 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 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 that 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.

Figure 4 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 $L = 2,\ M_1 = M_2 = 5$, and users of the first group require their packet delay D_1 at $p = 1$ no larger than d_1. We considered different delay requirement of the first group, as illustrated by asterisks in Figure 5. The corresponding q was obtained by (19). The simulated delay of the first group was indicated by the solid line in Figure 5. The circles and dashed line indicate, respectively, the calculated delay and simulated delay of the second group for a given q. Figure 5 shows that the delay requirement of the first group was satisfied for the choice of q given in (19).

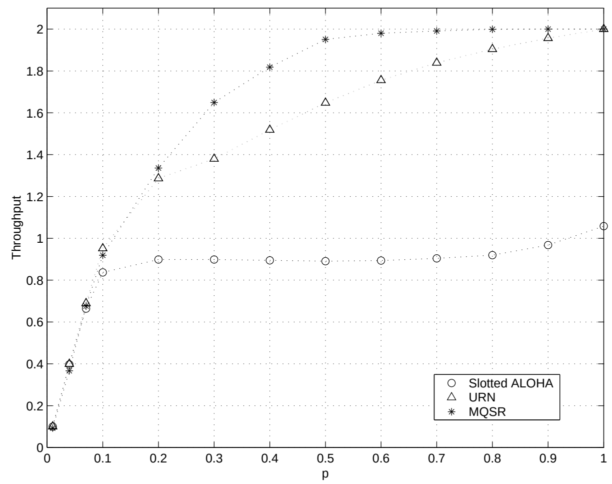


Fig. 4: Throughput comparison

figure

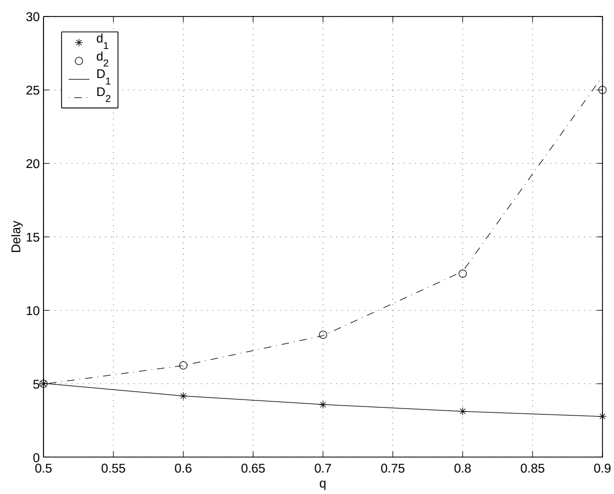


Fig. 5: Delay performance of the MQSR protocol at $p = 1$

figure

V. Conclusion

In this paper, we have proposed a multi-queue service room 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

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

$$N_l = B_l + T_l. \tag{36}$$

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,

$$T_l = p(M_l - B_l). \tag{37}$$

Solving for B_l from (37) and substituting into (36), we get

$$N_l = T_l + M_l - \frac{T_l}{p}. \tag{38}$$

¿From Little's Theorem, we also have

$$D_l = \frac{N_l}{T_l}. \tag{39}$$

(8) then follows by substituting (38) into (39).

□□□

Appendix B

An Analysis on 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 $n = 1, \cdots, M$ and $k = 0, \cdots, n$ is considered.

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

$$D(1) = \frac{M}{\eta}. \tag{40}$$

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,

$$D \leq ME[\tau]. \tag{41}$$

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) and it is idle (it enters the service room without a packet). Define

$$m_1 \triangleq E[\tau \mid \text{the UoI is active}], \quad m_2 \triangleq E[\tau \mid \text{the UoI is idle}]. \tag{42}$$

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 1 slot, we have

$$m_1 \leq ME[\tau_A]. \tag{43}$$

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

$$\begin{aligned} E[\tau_A] &= \sum_{n=1}^{\infty} P[\tau_A \geq n] \\ &\leq \sum_{n=1}^{\infty} P[\text{in each of (n-1) slots, not all transmitted packets are successfully received}] \end{aligned}$$

$$\begin{aligned} &\leq \sum_{n=1}^{\infty}(1 - \min_{l=1,\cdots,M} \mathcal{C}_{l,l})^{n-1} \\ &= \frac{1}{\min_{l=1,\cdots,M} \mathcal{C}_{l,l}}. \end{aligned} \quad (44)$$

Note that $\mathcal{C}_{l,l} > 0$ for all l, a consequence of the condition that $C_{n,k} > 0$ for $n = 1, \cdots, M$ and $k = 0, \cdots, n$. Thus, from (43) and (44), we have

$$m_1 \leq \frac{M}{\min_{l=1,\cdots,M} \mathcal{C}_{l,l}}. \quad (45)$$

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 $n = 1, \cdots, M$ and $k = 0, \cdots, n$, 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 $E[\tau_0^{(i)}]$ 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 $j \leq M-1$,

$$m_2 \leq (M-1)m_1 + \max_{i=1,\cdots,M} E[\tau_0^{(i)}]. \quad (46)$$

We now bound $E[\tau_0^{(i)}]$ for $i = 1, \cdots, M$. It is clear that $E[\tau_0^{(M)}] = 1$. 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 $E[\xi^{(i)}]$ 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

$$\begin{aligned} E[\tau_0^{(i)}] &\leq E[\xi^{(i)}] + E[\tau_0^{(i+1)}] \\ &\leq \sum_{r=i}^{M-1} E[\xi^{(r)}] + E[\tau_0^{(M)}] \\ &= \sum_{r=i}^{M-1} E[\xi^{(r)}] + 1. \end{aligned} \quad (47)$$

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

$$E[\tau_0^{(i)}] \leq i \sum_{r=i}^{M-1} E[\xi^{(r)}] + 1. \tag{48}$$

Now consider the user, denoted by 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, $E[\xi_n^{(i)}]$ denote the average number of slots till 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

$$E[\xi^{(i)}] = \sum_{n=1}^{\infty} E[\xi_n^{(i)}] p_n. \tag{49}$$

We now need to bound $E[\xi_n^{(i)}]$ 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

$$E[H] \leq (M-i)m_1, \tag{50}$$

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 $E[\xi_n^{(i)}]$ as follows.

$$E[\xi_n^{(i)}] \leq n(M-i)m_1. \tag{51}$$

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

$$p_n \leq (1-(1-p)^{(M-i)m_1})^{n-1}. \tag{52}$$

Since $(1-(1-p)^{(M-i)m_1}) < 1$, we have, from (49),

$$\begin{aligned} E[\xi^{(i)}] &\leq (M-i)m_1 \sum_{n=1}^{\infty} n(1-(1-p)^{(M-i)m_1})^{n-1} \\ &= \frac{(M-i)m_1}{(1-p)^{2(M-i)m_1}}. \end{aligned} \tag{53}$$

Thus, from (48,53), we have

$$E[\tau_0^{(i)}] \leq 1 + i \sum_{r=i}^{M-1} \frac{(M-r)m_1}{(1-p)^{2(M-r)m_1}}. \tag{54}$$

With (46), we then have

$$m_2 \leq 1 + (M-1)m_1 + \max_{i=1,\cdots,M-1} i \sum_{r=i}^{M-1} \frac{(M-r)m_1}{(1-p)^{2(M-r)m_1}}, \tag{55}$$

which, combining with (41) and (45), 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.*, page 37, 1970.

[2] N. Abramson. "The Throughput of Packet Broadcasting Channels". *IEEE Trans. Comm*, COM-25(1):117–128, January 1977.

[3] J. Q. Bao and L. Tong. "Performance Analysis of Slotted Aloha Ad Hoc Networks with Multiple Packet Reception". *Submitted to IEEE Trans. Communications.*

[4] J.I. Capetanakis. "Generalized TDMA: The Multi-Accessing Tree Protocol". *IEEE Trans. Communications*, 27(10):1476–1484, Oct. 1979.

[5] J.I. Capetanakis. "Tree Algorithms for Packet Broadcast Channels.". *IEEE Trans. Information Theory*, 25(5):505–515, Sept. 1979.

[6] G. del Angel and T.L. Fine. "Randomized Power Control Strategies for Optimization of Multiple Access Radio Systems". In *Proc. 38th Allerton Conference on Communication, Control and Computing*, October 2000.

[7] R. Gallager. "Conflict Resolution in Random Access Broadcast Networks". In *Proc. of AFOSR Workshop on Comm. Theory and Appl.*, pages 74–76, Sept. 1978.

[8] S. Ghez, S. Verdú, and S.C. Schwartz. "Stability Properties of Slotted Aloha with Multipacket Reception Capability" . *IEEE Trans. Automat. Contr.*, 33(7):640–649, July 1988.

[9] S. Ghez, S. Verdú, and S.C. Schwartz. "Optimal Decentralized Control in the Random Access Multipacket Channel". *IEEE Trans. Automat. Contr.*, 34(11):1153–1163, Nov. 1989.

[10] M.G. Hluchyj. "Multiple Access Window Protocol: Analysis for Large Finite Populations". In *Proc. IEEE Conf. on Decision and Control*, pages 589–595, New York, NY, 1982.

[11] M.G. Hluchyj and R.G. Gallager. "Multiaccess of A Slotted Channel by Finitely Many Users". In *Proc. Nat. Telecomm. Conf.*, pages D4.2.1–D4.2.7, New Orleans, LA., Aug. 1981.

[12] L. Kleinrock and Y. Yemini. "An Optimal Adaptive Scheme for Multiple Access Broadcast Communication". In *Proc. International Conference on Communications*, pages 7.2.1–7.2.5, June 1978.

[13] J.J. Metzner. "On Improving Utilization in ALOHA Networks". *IEEE Trans. Communications*, 24:447–448, Apr. 1976.

[14] L.G. Roberts. "Aloha Packet System with and without Slots and Capture". In *(ASS Note 8). Stanford, CA: Stanford Research Institute, Advanced Research Projects Agency, Network Information Center.*, 1972.

[15] M. Sidi and I. Cidon. "Splitting Protocols in Presence of Capture". *IEEE Trans. Info. Theory*, IT-31:295–301, March 1985.

[16] Q. Zhao and L. Tong. "Semi-Blind Collision Resolution in Random Access Wireless Ad Hoc Networks". *IEEE Trans. Signal Processing*, 48(10):2910–2920, Oct. 2000.

Contents

List of Figures