Stable Throughput, Rate Control, and Delay in Multi-Access Channels

Beiyu Rong, Anthony Ephremides

Stable Throughput, Rate Control, and Delay in Multi-Access Channels

Beiyu Rong Anthony Ephremides
Department of Electrical and Computer Engineering
University of Maryland, College Park, MD 20742
byrong@umd.edu *etony@umd.edu*

***Abstract*—In this paper, we investigate the stability and delay issues of a two-user multi-access channel at the bit level. The two users, have the option to transmit at a higher rate (measured in bits/slot) separately, or to transmit simultaneously but at a lower rate because of the interference caused by concurrent transmissions. Source burstiness is considered by modeling random arrivals at the users, and the stability region in terms of bits/slot is derived. Further, we determine the condition under which the maximum stability region is achieved when both users transmit with probability 1 if they are backlogged; in this case, the stability region is shown to be coordinate convex. Then, under such condition for a convex stability region, we study the minimum delivery time problem where each user is given at the beginning some amount of data for the destination. For any initial queue size vector, we explicitly characterize the optimal rate allocation policy that empties the two users' queues within the shortest time.**

I. Introduction

The role of medium access control has significant impact in wireless networks, as the communication over wireless channels is affected by a lot of factors, such as the transmitting power, distance, target bit error rate, as well as interference from other transmitters. As such, the studies of wireless multi-access system, in which multiple users contend for the common channel, become interesting and crucial components for the basic understanding of the behavior of wireless networks.

The classical networking-theoretic approach in multi-access channels rests on the packet level, which handles the communication unit as *packet*, regardless of the content inside a packet. To do this, the physical layer properties which include the effects of fading, attenuation and interference are abstracted into a simple packet-erasure channel model: the transmission is considered to be successful with a certain probability, and unsuccessful otherwise. Then the performance measures of interest are the packet throughput, packet delay, etc. However, packets cannot be treated separately from the internal bits. If a packet contains more bits and hence is transmitted at a higher data rate, the successful reception probability of that packet is relatively low; on the other hand, higher packet throughput can be achieved by lowering the data rate. Therefore, it is more appropriate and accurate to consider the problem at the *bit* level, and evaluate the performance metrics such as bit throughput and bit delay. There have been several papers in the literature which investigate the problem of rate allocation in multi-access channels, with the objective to maximize the throughput or minimize the average delay [1]- [4]. However, those works are subject to a couple of limitations. First, their channel model is restricted to additive white Gaussian noise multiple access channel only, whose capacity region has a polymatroidal structure which is essential for their proof of optimality. Second, in those works, the transmission rates assigned to the users are selected from the Shannon capacity region, and the optimal rate policy is shown to be operated on the boundary of the capacity region. However, The Gaussian capacity region is obtained when we assume the availability of infinite number of source bits at the transmitters, which may not hold under the assumption of bursty arrivals. In addition, the rates on the boundary are asymptotically achievable when the codeword length approaches infinity, thus, such rates cannot be achieved with finite-length codeword transmitted in a finite time slot. And the infinite-length codeword implies infinite delay, which makes the delay minimization problem questionable.

In this paper, we first investigate the stability issue of a multi-access channel consisting of two users and a common destination. In each slot, information bits arrive at the two users independently according to some stationary processes. Due to the difficulty and cost of global coordination in the wireless environment, here, each user, if it is backlogged, decides to transmit with some probability independently of any other event. Then, the transmission rates are selected from a finite, discrete set, such that the allocated rates are feasible according to the action of both users. The feasible rate set can be derived from any channel model, and is not restricted to the Gaussian channel only. Basically, if only one user decides to transmit, the channel is able to support higher transmission rate; otherwise, if both users decide to transmit together in a time slot, the supportable data rate from each user must be lowered in order to combat interference and ensure successful reception at the destination. By doing this, we require that the users exchange the information about whether they will transmit or remain idle; considering that such coordination requires only one bit from each user (active or idle), we justify the implementation of this partially distributed rate allocation scheme. Under these assumptions, we explicitly characterize the stability region (in bits/slot) of this multi-access system. The stability region is defined as the set of all arrival rates for which there exist transmission probabilities such that all queues in the network remain finite[1]. Consequently, the structure property of the stability region is discussed, and we determine the necessary and sufficient condition under which the stability region is coordinate convex. In this case, the optimal policy for attaining the maximum achievable stability region is shown to be the rate allocation policy with transmission probability one for both users. According to our model, the stability region is the union of all stabilizable arrival rates under some transmission probabilities. Hence, we *define* a policy to be optimal if it can stabilize all arrival rates that are stabilizable with some transmission policy; and the stability region of the optimal policy, which is indeed the entire stability region, is also referred to as the maximum stability region in this paper.

In the second part of this paper, we focus our attention on the delay performance. We look into the case where the users are initially given a certain amount of data to be delivered to the destination, and there are no more arrivals. We are interested in characterizing the optimal policy that drains the two users' queues using the minimum time. Under the same condition for the stability region being convex, the optimal policy is shown to have similar structure as the policy investigated in [6], [7] for the Gaussian broadcast channels, and is the one that empties the two queues with the same expected time.

The rest of this paper is organized as follows: Section II describes the network model. Section III characterizes the stability region of the investigated multi-access channel in terms of bits/slot, followed by the condition under which the stability region is convex. Section IV provides the policy that delivers all bits in the system within the shortest time. Finally section V concludes the paper.

[1]The strict definition of queueing stability is provided in [5].

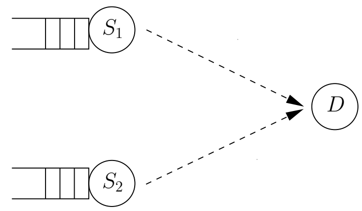

Fig. 1. The multi-access channel with two users S_1, S_2, and a common destination D.

II. Network Model

We consider throughout this paper the wireless multi-access channel as shown in Fig. 1. Two users, S_1 and S_2, unicast traffic to the common destination D. We assume slotted system with a normalized slot duration. Different from the packet-based approach in most work at the network layer, which assumes the transmission of one packet in a time slot, here, we take into account the bit-nature of the packet content. As such, we model that each user, if it transmits, will transmit a finite number of bits. Generally, the transmission rates can be chosen from a continuous set; but for simplifying the analysis, we assume that user S_i ($i \in \{1, 2\}$) is allowed to transmit with one of the two rates: R_i and r_i. The rates are chosen such that R_i is the maximum supportable rate S_i can transmit to D without the presence of $S_{\bar{i}}$'s transmission ($\bar{i}$ denotes the complementary of i); and r_i is the maximum rate S_i can transmit if both users transmit simultaneously. Due to the interference caused by concurrent transmissions, the supportable rate under single transmission must be greater than the supportable rate under simultaneous transmissions. That is, $R_i > r_i$, for $i = 1, 2$. These rates depend on various parameters such as the received power, tolerable bit error rate, as well as the coding and decoding mechanisms. For example, the rate can be defined as some increasing function of the received Signal-to-Interference-Noise Ratio (SINR) as in [8]. In some papers, the rate is represented by the Shannon capacity rate $\log_2(1 + \text{SINR})$. As we have commented in Section I, although this approach based on Shannon capacity is widely used for the rate control problem, its validity is questionable. However, the actual choices for the transmission rates won't affect our analysis.

We would like to remark that our model is very general since we do not restrict it to any specific channel model (e.g., additive Gaussian multiple-access channel);

all rates chosen following the principle described above are included in our model. In addition, we incorporate the physical layer properties into the network-layer studies by considering more realistic physical channel model than the packet model.

In each time slot, bits arrive at the two users according to some stationary random processes, and are queued for transmission. The arrival processes are independent from user to user, and are i.i.d. over slots. Denote by $q_1[n]$ and $q_2[n]$ the number of bits in the queues at S_1 and S_2 respectively, at the beginning of time slot n. Instead of a centralized scheduling like TDMA or a fully distributed random access scheme, here, we consider a class of transmission policy that is partially distributed with some coordination. The transmission policy is distributed as each user decides to transmit with some probability if it is backlogged, independent of the other user's action. Then, users communicate over a control sub-channel exchanging the information whether they decide to transmit or not; according to the action of the other user, each user selects the maximum supportable rate it can transmit.

At each time slot n, the transmission policy is described as follows:

- If $q_i[n] > 0$, user S_i decides to transmit with probability p_i; if $q_i[n] = 0$, S_i has to be idle with probability 1.
- If both users are backlogged and decide to transmit, by exchanging the information they know that they both will transmit, so each user S_i ($i \in \{1, 2\}$) selects the rate r_i to transmit, as r_i is the maximum supportable rate when they are transmitting simultaneously.
- If only one of the users, say user S_i, decides to transmit, it transmits with rate R_i.

We note that the control information shared by the two users requires only one bit from each side (transmit or idle), and hence, the overhead incurred by this type of coordination is negligible. Defined in this way, we see that in each time slot, there are three possible allocation rate pairs: $(R_1, 0)$, (r_1, r_2) and $(0, R_2)$, according to the queue state and the action chosen by the two users.

III. Stability Region

As defined by the transmission policy, user S_i decides to transmit with probability p_i if it is backlogged, and the feasible transmission rate will depend on both users' actions. That is to say, the departure process from one user's queue depends on the action of the other user. Further, the action of the user depends on the queue state: if a user is empty, it will not transmit; but if it is non-empty, it will transmit with some probability. Therefore, the queues at S_1 and S_2 interact in a complicated manner which makes the stability analysis extremely difficult. Indeed, the stability region for a slotted random multiple access channel has been solved for the two-user case only; the closure of the stability region remains an open problem for the system with more than two users.

For the reason to be able to derive analytical result, we restrict our analysis to the scenario with two users. As a means for bypassing the difficulty and characterizing the stability region, the stochastic dominance approach was introduced in [9] to decouple the users' queues; then for each single queue, Loynes' Theorem [5] can be applied to determine stability conditions. We will utilize these tools in the context of our bit-level multiple access channel model; the main result is provided in the following theorem.

Theorem 1: Denote by λ_1, λ_2 the average arrival rates to the users S_1 and S_2 respectively, measured in bits per slot. The stability region of the investigated multi-access channel is given by:

$$\Re = \bigcup_{(p_1,p_2)\in[0,1]^2} \Big(\Re_1(p_1,p_2) \bigcup \Re_2(p_1,p_2)\Big) \tag{1}$$

where

$$\Re_1(p_1,p_2) = \left\{ (\lambda_1,\lambda_2) : \begin{array}{l} \lambda_1 < p_1 R_1 - \frac{p_1(R_1-r_1)}{R_2-p_1(R_2-r_2)}\lambda_2 \\ \lambda_2 < p_2 R_2 - p_1 p_2 (R_2 - r_2) \end{array} \right\} \tag{2}$$

$$\Re_2(p_1,p_2) = \left\{ (\lambda_1,\lambda_2) : \begin{array}{l} \lambda_1 < p_1 R_1 - p_1 p_2 (R_1 - r_1) \\ \lambda_2 < p_2 R_2 - \frac{p_2(R_2-r_2)}{R_1-p_2(R_1-r_1)}\lambda_1 \end{array} \right\} \tag{3}$$

and $(p_1, p_2) \in [0, 1]^2$ is the transmission probability pair.

Proof: See Appendix A. ■

Lemma 1: By varying the transmission probabilities (p_1, p_2) over $[0, 1]^2$, the closed form of the stability region for a general channel model takes the following form:

$$\Re = \mathcal{L}_1 \bigcup \mathcal{L}_2 \bigcup \mathcal{L}_3 \tag{4}$$

where

$$\mathcal{L}_1 = \left\{ (\lambda_1,\lambda_2) : \lambda_2 < R_2 - \frac{(R_2-r_2)\lambda_1}{r_1}, \quad \text{for } \lambda_1 \in \left[0, \frac{r_1^2 \cdot R_2}{R_1(R_2-r_2)}\right) \right\} \tag{5}$$

$$\mathcal{L}_2 = \Bigg\{ (\lambda_1, \lambda_2) : \sqrt{\frac{(R_2 - r_2)\lambda_1}{R_1 R_2}} + \sqrt{\frac{(R_1 - r_1)\lambda_2}{R_1 R_2}} < 1, \quad \text{for } \lambda_1 \in \Big[\frac{r_1^2 \cdot R_2}{R_1(R_2 - r_2)}, \frac{R_1(R_2 - r_2)}{R_2}\Big) \Bigg\} \tag{6}$$

$$\mathcal{L}_3 = \Bigg\{ (\lambda_1, \lambda_2) : \lambda_2 < \frac{R_1 r_2}{R_1 - r_1} - \frac{r_2 \lambda_1}{R_1 - r_1}, \quad \text{for } \lambda_1 \in \Big[\frac{R_1(R_2 - r_2)}{R_2}, R_1\Big) \Bigg\} \tag{7}$$

Proof: See Appendix B. ■

In general, the stability region is bounded by straight lines in some part, and bounded by a strictly convex function in the remaining part. As a direct result from the proof of Lemma 1, a special case for which the stability region is coordinate convex is stated in the following lemma.

Lemma 2: If and only if

$$\frac{r_1}{R_1} + \frac{r_2}{R_2} \geq 1 \tag{8}$$

the optimal transmission probability pair for maximizing the stability region is $(p_1^*, p_2^*) = (1, 1)$. The resulting stability region is coordinate convex, and is given by:

$$\Re = \mathcal{I}_1 \bigcup \mathcal{I}_2 \tag{9}$$

where

$$\mathcal{I}_1 = \Bigg\{ (\lambda_1, \lambda_2) : \lambda_2 < R_2 - \frac{(R_2 - r_2)\lambda_1}{r_1}, \quad \text{for } \lambda_1 \in [0, r_1) \Bigg\} \tag{10}$$

$$\mathcal{I}_2 = \Bigg\{ (\lambda_1, \lambda_2) : \lambda_2 < \frac{R_1 r_2}{R_1 - r_1} - \frac{r_2 \lambda_1}{R_1 - r_1}, \quad \text{for } \lambda_1 \in [r_1, R_1) \Bigg\} \tag{11}$$

Proof: We observe from Eq. (27) in Appendix B that, if

$$\frac{r_1^2 \cdot R_2}{R_1(R_2 - r_2)} \geq \frac{R_1(R_2 - r_2)}{R_2} \iff \frac{r_1}{R_1} + \frac{r_2}{R_2} \geq 1 \tag{12}$$

the subregion represented by $\mathcal{L}_2$ does not exist. In this case, the stability region is bounded by the union of $\mathcal{L}_1$ and $\mathcal{L}_3$, as expressed in Eq. (5) and Eq. (7) respectively.

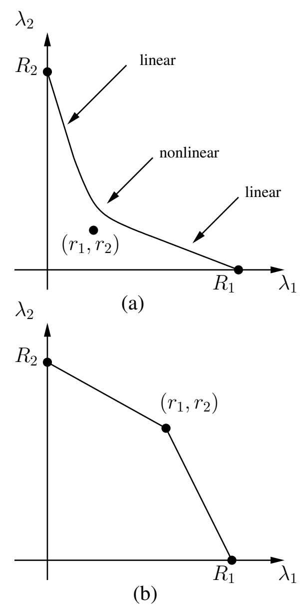

Fig. 2. The closure of the stability region. (a) Eq. (8) is not satisfied, $\Re$ is bounded by straight lines close to the axes and a convex curve in the interior. (b) Eq. (8) is satisfied, $\Re$ is coordinate convex.

Further, the optimal transmission probability (p_1^*, p_2^*) is (1,1), as a direct consequence by following the proof in Appendix B. The convexity of the stability region in this case is shown as follows. First, it can be easily calculated that, the two straight lines that bound the subregions $\mathcal{L}_1$ and $\mathcal{L}_3$ intersect at the point (r_1, r_2). Then, If we form a straight line by connecting the two points $(R_1, 0)$ and $(0, R_2)$ at the x-axis and y-axis, and denote this line by L; the intersection point (r_1, r_2) lies above line L if and only if the condition in Eq. (8) holds. Therefore, the stability region characterized by Eq. (9) in this special case is coordinate convex. ■

This result implies, when the inequality in Eq. (8) is satisfied, each user should transmit whenever it is backlogged: if both users are non-empty, the transmission rates (r_1, r_2) are assigned to the two users; otherwise, if only user S_i is non-empty, S_i transmits with rate R_i. This policy will stabilize any arrival rate vectors that are stabilizable under some rate control policy we investigate. The resulting stability region is coordinate convex, and is bounded by two straight lines that intersect at (r_1, r_2). In Fig. 2, we display some diagrams to illustrate the properties of the stability region in different cases. In Fig. 2(a), the condition in Eq. (8) is not satisfied, and the stability region is bounded by two straight lines from the axes and a convex function in the interior; whereas, in Fig. 2(b), the condition in Eq. (8) is satisfied, and the stability region becomes convex.

In the following part of this paper, we restrict our attention to the case when Eq. (8) is satisfied. Under such channel condition, we obtain some useful result regarding the delay issue.

IV. Minimum Delivery Time Policy

In the previous section, we establish the optimal rate allocation policy that yields the maximum stability region under the condition expressed in Eq. (8). However, a policy that is throughput-optimal does not guarantee optimal delay performance. For example, the backpressure algorithm introduced in [11] was proved to achieve the maximum stability region in a relatively general network model, but it can have poor delay performance. In this section, we focus our attention on the scenario where initially there are source data queued at the users but there are no new arrivals after slot $n = 0$; the objective is to minimize the total time to empty both users' queues. We will show that the minimum delivery time policy in our problem has a similar property as the Queue Proportional Scheduling (QPS) policy studied in [6], [7]. In those works, the authors worked on the Gaussian broadcast channels, and the control space for all feasible transmission rates is the Shannon capacity region. In [6], the QPS policy was shown to achieve the maximum stability region; in [7], the QPS policy was shown to minimize the total delivery time. In the following theorem, we provide the minimum delivery time policy in our multi-access channel.

Theorem 2: Denote by $\mathbf{q_0} = (q_1, q_2)$ the initial queue size vector at users S_1 and S_2. When the condition in Eq. (8) is satisfied, the minimum delivery time policy is described as follows:

1) if
$$\frac{q_2}{q_1} \geq \frac{r_2}{r_1} \tag{13}$$
the optimal policy is to allocate the rates (r_1, r_2) to the users with probability $p = \frac{q_1 R_2}{q_1 R_2 + q_2 r_1 - q_1 r_2}$, and allocate the rates $(0, R_2)$ with probability $1 - p = \frac{q_2 r_1 - q_1 r_2}{q_1 R_2 + q_2 r_1 - q_1 r_2}$ until one queue is emptied, after which the policy serves the backlogged user S_i with rate R_i;
2) otherwise, that is, if
$$\frac{q_2}{q_1} < \frac{r_2}{r_1} \tag{14}$$
the optimal policy is to allocate the rates (r_1, r_2) to the users with probability $p' = \frac{q_2 R_1}{q_2 R_1 + q_1 r_2 - q_2 r_1}$, and allocate the rates $(R_1, 0)$ with probability $1 - p' = \frac{q_1 r_2 - q_2 r_1}{q_2 R_1 + q_1 r_2 - q_2 r_1}$ until one queue is emptied, after which the policy serves the backlogged user S_i with rate R_i.

Proof: First, we note that to compare the total delivery time of different policies, it is equivalent to compare the average allocated rates to the two users. Suppose that the total delivery time of a policy is T, then the average allocated rate vector is given by $\mathbf{R}_{\text{Avg}} = \mathbf{q_0}/T$. As we have stated above, the set of all possible allocation rates in each time slot is $\{(R_1, 0), (r_1, r_2), (0, R_2)\}$. When the condition in Eq. (8) is satisfied, the union of all average allocated rates coincides with the maximum achievable stability region expressed by Eq. (9), which is coordinate convex. Hence, the average allocated rate vector must be inside the stability region, that is, $\mathbf{R}_{\text{Avg}} \in \Re$. Then, we show that the average allocated rate vector of the optimal policy must lie on the boundary of $\Re$. The argument is simple: suppose that the average allocated rate vector (d_1, d_2) of the optimal policy is not on the boundary but in the interior, we can always find rate vector (d_1', d_2') on the boundary with $d_1' > d_1$ and $d_2' > d_2$, then a different policy with average allocation rates (d_1', d_2') must have better performance. Finally we can show that the optimal policy must empty the two queues using the same expected time. This can be proved as follows: without loss of generality, we select a different policy that empties S_1's queue first, then the average rate allocated to user S_1 must be increased, at the expense of decreasing the rate allocated to user S_2. As a result, the queue at S_2 needs more time slots to empty, and the total delivery time is determined by the queue that is emptied last. Therefore, under the optimal policy, the two queues are emptied with the same expected time; and a policy with such property yields the average allocation rate vector proportional to the initial queue size vector.

Consequently, the explicit characterization of the optimal policy depends on the initial queue sizes. If $q_2/q_1 \geq r_2/r_1$, the initial queue size vector (q_1, q_2) falls inside the semi-infinite triangle (I) as illustrated by Fig. 3(a), and the optimal policy has the average allocation rates (l_1, l_2) as indicated in Fig. 3(a). To achieve this rate pair, the policy selects the rates (r_1, r_2) with probability p, and selects the rates $(0, R_2)$ with probability $1-p$; after simple calculation the probability p is as given in Theorem 2. Otherwise, if $q_2/q_1 < r_2/r_1$ and the initial queue size vector falls inside the region (II) as shown by Fig. 3(b), a parallel statement can be made. In case a queue is emptied first, we allocate the maximum supportable rate to the remaining queue until it is emptied. ■

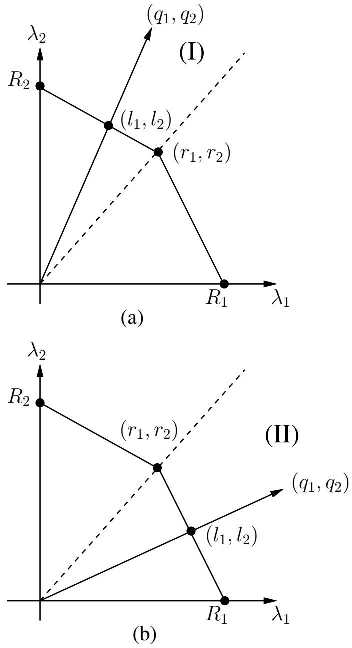

Fig. 3. The average allocation rates of the optimal policy to minimize the total delivery time. (a) Condition (13) holds, (q_1, q_2) is inside region (I). (b) Condition (14) holds, (q_1, q_2) is inside region (II).

Although it has been proved that, when the condition $\frac{r_1}{R_1} + \frac{r_2}{R_2} \geq 1$ holds, the transmission policy with transmission probability 1 for both users is optimal for the stable throughput, in the sense that it can stabilize any arrival rate vector that is inside the stability region. Here we show that, under the same condition, such policy does not drain the queues within the shortest time among all policies; and the minimum delivery time policy is the one described in Theorem 2 which takes into account the initial traffic demand in both queues. It will be more interesting to address the minimum average delay policy, which minimizes the average total time experienced by all bits in the system when there are random arrivals to the users. The minimum average delay problem appears to be challenging even in our two-user multiple access channel model, where both users are allowed to transmit simultaneously, though at a lower rate. This investigation is left to the future work.

V. Conclusion

We studied the rate allocation problem in a multi-access channel. Instead of using the Shannon capacity rate at the network layer, the transmission rates are selected from a finite, discrete set in which the rates are achievable in a finite time slot. Under a class of partially distributed transmission policies, we first characterized the stability region in terms of bits/slot. Then, the optimal policy for achieving the maximum stability region under a certain channel condition was determined, and the resulting stability region was shown to be coordinate convex. Finally, the minimum delivery time problem was investigated in our channel model. Under the same condition for a convex stability region, we explicitly characterized the optimal policy that empties the users' queues in minimum time. The minimum delivery time policy was shown to drain the queues with the same expected time. A related problem to investigate after this work is the minimum average delay policy with random arrivals. There are many other possible directions for the future work, for example, the consideration of the rate allocation policies when there is no coordination between the users; it is also possible to generalize the channel model to the non-stationary time-varying channel.

Appendix A
A proof of Theorem 1

We first find the stability region $\Re(p_1, p_2)$ for a fixed transmission probability pair (p_1, p_2). Tow parallel dominant systems $\mathcal{M}^i$ ($i \in \{1, 2\}$) are constructed in the following way:

- In the dominant system $\mathcal{M}^1$, user S_1 transmits dummy bits with probability p_1 if it is empty, while user S_2 acts the same as in the original system.
- In the dominant system $\mathcal{M}^2$, user S_2 transmits dummy bits with probability p_2 if it is empty, while user S_1 acts the same as in the original system.

The arrival processes, the channel model, the transmission and reception processes and all other assumptions remain unaltered in the dominant systems. As already established in [9], the transmission of dummy bits has no contribution to the throughput but causes unnecessary interference; hence, the dominant system stochastically dominates the original system in the sense that the queue sizes in the dominant system are not smaller than the corresponding queue sizes in the original system. It thus implies that the stability of the dominant systems is sufficient for the stability of the original system.

Then we analyze the stability conditions with respect to each dominant system. In system $\mathcal{M}^1$, user S_1 always attempts to transmit with probability p_1, regardless of its queue state. Therefore, user S_2 experiences an average service rate of $p_2 p_1 r_2 + p_2(1 - p_1)R_2$. This is because: if S_1 decides to transmit, which occurs with probability p_1, S_2 is able to transmit with rate r_2; otherwise, if S_1 decides not to transmit with probability $1 - p_1$, S_2 is able to transmit alone with rate R_2 to the destination. And if S_2 is backlogged, it will attempt to transmit with probability p_2. Hence, the service rate for S_2 in system $\mathcal{M}^1$ takes an average of $p_2p_1r_2 + p_2(1-p_1)R_2$. For each single queue, according to Loynes' Theorem, if the arrival and service processes of a queue are jointly stationary, the queue is stable if and only if the average arrival rate is less than the average service rate. Therefore, the queue at S_2 is stable if and only if:

$$\begin{aligned} \lambda_2 &< p_2p_1r_2 + p_2(1-p_1)R_2 \\ &= p_2R_2 - p_1p_2(R_2 - r_2) \end{aligned} \tag{15}$$

Now we analyze the stability condition of the queue at S_1. The service process of S_1 depends on the state of S_2's queue. When S_2 is not empty, S_2 will decide to transmit with probability p_2, in this case, S_1 is only able to deliver r_1 bits to the destination; otherwise, if S_2 decides not to transmit, S_1 is able to transmit with rate R_1. On the other hand, when S_2 is empty, S_1 can always transmit with rate R_1. As a result, the average service rate of S_1 should be calculated as

$$\begin{aligned} \mu_1 = p_1 \Big\{ &\mathbf{P}\left[q_2 \neq 0\right](p_2r_1 + (1-p_2)R_1) \\ &+ (1 - \mathbf{P}\left[q_2 \neq 0\right])R_1 \Big\} \end{aligned} \tag{16}$$

Since the queue at S_2 is a discrete-time M/M/1 queue, the stationary probability that the queue is not empty is $\mathbf{P}\left[q_2 \neq 0\right] = \frac{\lambda_2}{\mu_2} = \frac{\lambda_2}{p_2R_2 - p_1p_2(R_2-r_2)}$; by substituting this into Eq. (16), we can obtain the average service rate of S_1, with which we can determine the stability condition of S_1 by Loynes' Theorem. The queue at S_1 is stable if and only if:

$$\lambda_1 < p_1R_1 - \frac{p_1(R_1 - r_1)}{R_2 - p_1(R_2 - r_2)}\lambda_2 \tag{17}$$

Hence, the dominant system $\mathcal{M}^1$ is stable if and only if the queues at both S_1 and S_2 are stable; and its stability region, denoted by $\Re_1(p_1, p_2)$, is as shown in Eq. (2).

By following a parallel argument for system $\mathcal{M}^2$ in which user S_2 transmits dummy bits, the stability region of $\mathcal{M}^2$, denoted by $\Re_2(p_1, p_2)$, is characterized by Eq. (3) in Theorem 1.

We know from the stochastic dominance argument that, the stability of the dominant systems implies the stability of the original system. That is, the stability region of the original system $\Re(p_1, p_2)$ is lower-bounded by $\bigcup_{i=1,2} \Re_i(p_1, p_2)$, so we have $\Re(p_1, p_2) \supseteq \bigcup_{i=1,2} \Re_i(p_1, p_2)$. Then we proceed by arguing that the boundary of the original system indeed *coincides* with the boundary of the dominant systems, which is $\Re(p_1, p_2) = \bigcup_{i=1,2} \Re_i(p_1, p_2)$. Consider system $\mathcal{M}^1$ where S_1 continues to transmit dummy bits when it is empty, given that $\lambda_2 < p_2R_2 - p_1p_2(R_2 - r_2)$, for those λ_1 that S_1's queue is stable in the dominant system, the corresponding queue is also stable in the original system; conversely, for those λ_1 which makes the queue at S_1 unstable in the dominant system, S_1 never empties and hence S_1 always attempts to transmit source bits with probability p_1. We observe that as long as S_1 never empties, the dominant system $\mathcal{M}^1$ and the original system behave the same. Therefore, we see that at saturation, the two systems are "indistinguishable". A similar "indistinguishability" argument can be applied to system $\mathcal{M}^2$. Thus we conclude that $\Re(p_1, p_2)$ is equal to $\bigcup_{i=1,2} \Re_i(p_1, p_2)$. Once the stability region for a fixed transmission probability pair (p_1, p_2) is found, the stability region for the system is the union of all such regions as (p_1, p_2) varies over $[0, 1]^2$. This concludes the proof of Theorem 1. ■

Appendix B
A proof of Lemma 1

Since we have obtained the stability region for a fixed probability pair (p_1, p_2) in Theorem 1, we can utilize the constrained optimization technique as in [10] to derive the closure of the stability region. To do this, for a fixed λ_1, we maximize λ_2 as (p_1, p_2) varies over $[0, 1]^2$ where λ_1 and λ_2 are constrained by the conditions in Eq. (2) and Eq. (3).

By replacing λ_1 with x and λ_2 with y, the boundaries of the stability region given by Eq. (2) and Eq. (3) can be written as:

$$x = p_1R_1 - \frac{p_1(R_1 - r_1)y}{R_2 - p_1(R_2 - r_2)} \tag{18}$$

$$\text{for} \quad 0 \leq y < p_2R_2 - p_1p_2(R_2 - r_2) \tag{19}$$

and

$$y = p_2R_2 - \frac{p_2(R_2 - r_2)x}{R_1 - p_2(R_1 - r_1)} \tag{20}$$

$$\text{for} \quad 0 \leq x < p_1R_1 - p_1p_2(R_1 - r_1) \tag{21}$$

First we consider the constrained optimization problem suggested by Eq. (20) as:

$$\max_{p_2 \in [0,1]} y = \max_{p_2 \in [0,1]} \left(p_2R_2 - \frac{p_2(R_2 - r_2)x}{R_1 - p_2(R_1 - r_1)} \right) \tag{22}$$

Differentiating with respect to p_2 gives:

$$\frac{dy}{dp_2} = R_2 - \frac{R_1(R_2 - r_2)x}{(R_1 - p_2(R_1 - r_1))^2} \tag{23}$$

Setting Eq. (23) to zero yields

$$p_2^* = \frac{R_1}{R_1 - r_1} - \frac{1}{R_1 - r_1}\sqrt{\frac{R_1(R_2 - r_2)x}{R_2}} \tag{24}$$

A simple calculation shows that the second derivative at p_2^* is negative; so p_2^* must be the unique maximizer if it can be reached. But caution is needed here. Since p_2^* is a probability and it must satisfy the condition that $0 \le p_2^* \le 1$, it follows from Eq. (24) that

$$\frac{r_1^2 \cdot R_2}{R_1(R_2 - r_2)} \le x \le \frac{R_1 R_2}{R_2 - r_2} \tag{25}$$

The constraint in Eq. (21) is valid only for $p_2 < \frac{R_1 - x}{R_1 - r_1}$, then for p_2^* given by Eq. (24) to be reachable, x must satisfy

$$0 \le x < \frac{R_1(R_2 - r_2)}{R_2} \tag{26}$$

Combining the conditions in Eq. (25) and Eq. (26), for x in the range given by

$$\frac{r_1^2 \cdot R_2}{R_1(R_2 - r_2)} \le x < \frac{R_1(R_2 - r_2)}{R_2} \tag{27}$$

p_2^* in Eq. (24) is the real maximizer; by substituting p_2^* into Eq. (20), the resulting maximum y is

$$y_{\text{max}} = \frac{\left(\sqrt{R_1 R_2} - \sqrt{(R_2 - r_2)x}\right)^2}{R_1 - r_1} \tag{28}$$

Now we consider those x such that

$$0 \le x < \frac{r_1^2 \cdot R_2}{R_1(R_2 - r_2)} \tag{29}$$

we calculate that

$$\frac{dy}{dp_2} > 0, \qquad \text{for} \quad \forall p_2 \in [0, 1] \tag{30}$$

Therefore, y is a strictly increasing function of p_2 for $0 \le x < \frac{r_1^2 \cdot R_2}{R_1(R_2 - r_2)}$, and the optimal p_2^* is equal to one. The corresponding y_{max} is

$$y_{\text{max}} = R_2 - \frac{(R_2 - r_2)x}{r_1} \tag{31}$$

Then we check for x such that $x \ge \frac{R_1(R_2 - r_2)}{R_2}$. From Eq. (21), we know that x should be no greater than R_1. Hence, we investigate those x which fall into the range given by

$$\frac{R_1(R_2 - r_2)}{R_2} \le x < R_1 \tag{32}$$

We notice that Eq. (21) implies p_2 must satisfy $p_2 < \frac{R_1 - x}{R_1 - r_1}$. Under these conditions, we can easily check that $\frac{dy}{dp_2}$ is strictly positive in this case. So y strictly increases with p_2; but since p_2 is no greater than $\frac{R_1 - x}{R_1 - r_1}$, the optimal p_2^* that maximizes y is $\frac{R_1 - x}{R_1 - r_1}$. By inserting it into Eq. (20), the maximum y is

$$y_{\text{max}} = \frac{R_1 r_2}{R_1 - r_1} - \frac{r_2 x}{R_1 - r_1} \tag{33}$$

Thus, the closure for the stability region provided by Eq. (20) is solved; its form is the same as $\mathcal{L}_1 \bigcup \mathcal{L}_2 \bigcup \mathcal{L}_3$ expressed in Lemma 1. By applying the same procedure on the constrained optimization problem suggested by Eq. (18), we obtain the same closure of the stability region. Therefore, we conclude the proof of Lemma 1. ■

REFERENCES

[1] E. M. Yeh and A. S. Cohen, "Throughput and delay optimal resource allocation in multiaccess fading channels," *IEEE Intl. Symposium on Information Theory (ISIT)*, Yokohama, Japan, June/July 2003, pp. 245.

[2] E. M. Yeh, "Minimum delay multiaccess communication for general packet length distributions," *the Allerton Conference on Communication, Control, and Computing*, Monticello, IL, Sep/Oct. 2004, pp. 1536–1545.

[3] N. Ehsan and T. Javidi, "Delay optimal transmission policy in a wireless multiaccess channel," *IEEE Trans. Inform. Theory*, vol. 54, no. 8, pp. 3745–3751, Aug. 2008.

[4] J. Yang and S. Ulukus, "Delay minimization in multiple access channels," *IEEE Intl. Symposium on Information Theory (ISIT)*, Seoul, Korea, June/July 2009, pp. 2366–2370.

[5] R. Loynes, "The stability of a queue with non-independent inter-arrival and service times," *Proc. Camb. Philos. Soc.*, vol. 58, pp. 497– 520, 1962.

[6] A. Eryilmaz, R. Srikant, and J. R. Perkins, "Throughput-optimal scheduling for broadcast channels," *Modelling and Design of Wireless Networks, Proceedings of SPIE*, vol. 4531, July 2001, pp. 70–78.

[7] K. Seong, R. Narasimhan, and J. M. Cioffi, "Queue proportional scheduling via geometric programming in fading broadcast channels," *IEEE Journal on Selected Areas in Communications*, vol. 24, no. 8, pp. 1593–1602, Aug. 2006.

[8] J. Luo and A. Ephremides, "On rate control of packet transmission over fading channels," *IEEE Communications Letters*, vol. 11, no. 12, pp. 982–984, Dec. 2007.

[9] R. R. Rao and A. Ephremides, "On the stability of interacting queues in a multiple-access system," *IEEE Trans. Inform. Theory*, vol. 34, no. 5, pp. 918–930, Sep. 1988.

[10] V. Naware, G. Mergen, and L. Tong, "Stability and delay of finite-user slotted ALOHA with multipacket reception," *IEEE Trans. Inform. Theory*, vol. 51, no. 7, pp. 2636–2656, July 2005.

[11] L. Tassiulas and A. Ephremides, "Stability properties of constrained queueing systems and scheduling policies for maximum throughput in multihop radio networks," *IEEE Trans. Automat. Contr.*, vol. 37, no. 12, pp. 1936–1948, Dec. 1992.