Adaptive MAC Protocols Using Memory for Networks With Critical Traffic

Jaeok Park and Mihaela van der Schaar, *Fellow, IEEE*

***Abstract*—We consider a multiaccess communication network where network users are subject to critical events such as emergencies and crises. If a critical event occurs to a user, the user needs to send critical traffic as early as possible. However, most existing medium access control (MAC) protocols are not adequate to meet the urgent need for data transmission by users with critical traffic. In this paper, we devise a class of distributed MAC protocols that achieve coordination using the finite-length memory of users containing their own observations and traffic types. We formulate a protocol design problem and find protocols that solve the problem. The proposed protocols enable a user with critical traffic to transmit its critical traffic without interruption from other users after a short delay while allowing users to share the channel efficiently when there is no critical traffic. Moreover, the proposed protocols require low communications and computational overhead.**

***Index Terms*—Adaptive protocols with memory, distributed medium access control protocols, networks with critical traffic, slotted multiaccess communication.**

I. Introduction

THIS paper considers a medium access control (MAC) protocol design problem for a slotted multiaccess communication network [1] where network users sharing a common resource (channel bandwidth) may face critical events such as emergencies and crises. Examples of critical events include a fire in a building, a natural disaster in a region, a heart attack of a patient, and a military attack by an enemy. When a network user detects a critical event, the user needs to send information about the event as early as possible so that necessary measures can be taken to mitigate the risk or help affected parties recover. Thus, a desirable MAC protocol should allocate channel bandwidth to a user with critical traffic in case of a critical event. On the other hand, a MAC protocol needs to yield high throughput and fairness when there is no critical traffic in the network. In other words, two kinds of coordination need to be achieved: i) coordination between a user with critical traffic and other users in case of a critical event and ii) coordination among users when there is no critical traffic.

The two kinds of coordination can be easily achieved when message passing is used. In case of a critical event, the user with critical traffic can be given priority by broadcasting its traffic type to induce other users to wait while critical traffic is transmitted. Also, when there is no critical traffic, users can share the channel in a contention-free manner by using coordination messages from a central controller as in time-division multiple access (TDMA). However, explicit message passing consumes considerable energy and is often impractical in a distributed network environment. As an alternative to explicit message passing, we use *memory* to achieve coordination in a distributed way, based on the idea of [2]. The idea of using memory to achieve coordination can be found in the existing literature, for instance, utilizing the Gur game [3] and the minority game [4] in the context of sensor networks.

Our prior work [2] considers a stationary setting where there is a single traffic type. Thus, MAC protocols in [2] need to achieve only the second kind of coordination. With a protocol with memory, a user determines its transmission probabilities depending on the finite-length history of its own observations (transmission actions and feedback information). As users take transmission actions in a probabilistic manner, the histories of users evolve differently across users as time passes. Reference [2] shows that, using the variations in the histories of users as a coordination device, we can obtain some degree of coordination without relying on explicit message passing.

The setting considered in this paper is stochastic in that the traffic types of users vary over time depending on the arrivals of exogenous events. Thus, MAC protocols need to achieve the first kind of coordination as well. In order to give priority to a user with critical traffic, users with different traffic types need to be treated in a different way. To achieve this, we extend protocols with memory, formulated in [2], so that transmission probabilities adjust not only to the history of observations but also to the history of traffic types. The proposed protocols are *adaptive* because users can change the modes of operation depending on their traffic types. Adaptive protocols with memory proposed in this paper have the following desirable properties.

1) Coordination in a critical phase—When a critical event occurs, the user with critical traffic transmits its packets successfully after a small delay while other users wait until critical traffic is completely transmitted. Furthermore, a delay constraint can be imposed to guarantee the average delay below a desired level.
2) Coordination in a normal phase—When there is no critical traffic, success periods alternate with contention periods. A success period contains consecutive successes by a single user while a contention period selects a successful user for the following success period. The average duration

Manuscript received May 05, 2010; revised September 22, 2010; accepted November 11, 2010. Date of publication November 29, 2010; date of current version February 09, 2011. The associate editor coordinating the review of this manuscript and approving it for publication was Prof. Amir Leshem.

The authors are with the Electrical Engineering Department, University of California, Los Angeles (UCLA), Los Angeles, CA 90095-1594 USA (e-mail: jaeok@ee.ucla.edu; mihaela@ee.ucla.edu).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TSP.2010.2094613

of a success period can be made arbitrarily large (at the expense of reduced short-term fairness) without affecting the average duration of a contention period.

3) Low communications overhead—The proposed protocols can be implemented without explicit message passing between users or between a central controller and a user, assuming that waiting users can sense whether the channel is accessed or not and that transmitting users can learn whether their transmissions are successful or not.
4) Low computational overhead—The proposed protocols utilize finite memory of a short length, thus exhibiting low computational complexity.

The proposed adaptive protocols have advantages over existing MAC protocols in dealing with critical traffic. Distributed coordination function (DCF), which is widely deployed in the IEEE 802.11a/b/g wireless local area network (WLAN) [5], does not differentiate users, and thus it is unable to give priority to a user with critical traffic. Slotted Aloha [6] has the same limitation. Users can be given different priorities depending on their access categories in enhanced distributed channel access (EDCA), which is deployed in IEEE 802.11e [7]. EDCA specifies different contention window sizes and arbitration interframe spaces (AIFS) to different access categories, yielding a smaller medium access delay and more bandwidth for the higher-priority traffic categories [8]. However, EDCA is designed to support applications requiring quality-of-service, and a user having highest-priority data shares the channel with other users. Thus, EDCA is not directly applicable to networks with critical traffic, where it is desirable to allocate the entire resource to a user with critical traffic. P-MAC [9] also differentiates users with different traffic classes by specifying different contention window sizes. However, P-MAC does not use AIFS, which creates a problem when applied to a network with critical traffic because even a user with the highest priority has a positive probability of collision at each transmission attempt.

The rest of this paper is organized as follows. In Section II, we describe the system model and adaptive protocols. In Section III, we define two performance metrics and provide a method to compute the metrics using Markov chains. In Section IV, we formulate a protocol design problem and solve it numerically. In Section V, we discuss how adaptive protocols can be enhanced by utilizing longer memory. In Section VI, we provide simulation results. In Section VII, we conclude the paper.

II. System Model and Protocol Description

A. System Model

We consider a communication channel shared by N contending users, or transmitter-receiver pairs. We assume that the number of users is fixed over time and known to users.[1] Time is divided into slots of equal length, and users maintain synchronized time slots. A user always has packets to transmit and can attempt to transmit one packet in each slot. As in [1, Ch. 4], only one user can transmit successfully in a slot, and simultaneous transmission by more than one user results in a collision.[2] After a user makes a transmission attempt, it learns whether the transmission is successful or not using an acknowledgement (ACK) response. We assume that there is no error in sending and receiving ACK signals. While a user waits, it senses the channel to learn whether the channel is accessed or not. Given this feedback structure, the set of the observations of a user in a slot can be defined as $Y = \{\text{idle}, \text{busy}, \text{success}, \text{failure}\}$, as in [11]. The observation of user i, denoted by y_i, is idle if no user transmits, busy if user i does not transmit but at least one other user transmits, success if user i transmits and succeeds, and failure if user i transmits but fails.

Users are subject to critical events such as emergencies and crises. If a critical event occurs to a user, the user is required to send critical traffic such as a rescue message describing the critical event. We assume that the length of critical traffic, measured by the number of packets needed to transmit it, is determined randomly. We say that a user's traffic is normal if its traffic is not critical. We use critical and normal users to refer to users with critical and normal traffic, respectively. We denote the type of user i's traffic by z_i and the set of possible types by Z so that $Z = \{\text{normal}, \text{critical}\}$. We assume that the observation and traffic type of a user are its local information. That is, users cannot observe the observations and the types of other users directly. Lastly, we assume that critical events occur infrequently so that there is at most one critical user at a time in the system.[3]

B. Protocol Description

We restrict our attention to distributed protocols that use no control or coordination message exchanges between a central controller and a user or between users. We label slots by $t = 1, 2, \ldots$ and use superscript t to denote variables pertinent to slot t. The history of user i in slot t contains all the information it has obtained before making a transmission decision in slot t and can be written as

$$h_i^t = \left(z_i^1, y_i^1, \ldots, z_i^{t-1}, y_i^{t-1}, z_i^t\right),$$

for $t = 2, 3, \ldots$, and $h_i^1 = z_i^1$. Let H_t be the set of all possible histories for a user in slot t. Then the set of all possible histories for a user can be defined by $H \triangleq \cup_{t=1}^{\infty} H_t$. A decision rule for a user can be formally represented by a mapping from H to [0, 1], prescribing a transmission probability following each possible history. A protocol is defined to be a collection of decision rules, one for each user.

We consider a simple class of protocols with the following two properties. First, we require that protocols be symmetric in the sense that it assigns the same decision rule to every user. The symmetry requirement can be justified by noting that symmetric protocols are easy to implement and that users in our model are *ex ante* identical. Moreover, it simplifies our analysis significantly. Second, we require that protocols use only the most recent observation and the current traffic type in a

[1] We investigate the case of the unknown number of users in Section IV-E.

[2] It will be interesting to consider multipacket reception [10], and we leave it for future research.

[3] An example is an attack by a resource-constrained enemy who needs some time to develop the capability to mount an attack. We consider the possibility of having two critical users at the same time in Section V-C.

stationary way (i.e., independent of slot label t). This requirement is motivated by a presumption that a protocol using short memory is easy to program and validate. We call a protocol satisfying the above two requirements an adaptive MAC protocol with one-slot memory, or more simply, an adaptive protocol. It can be represented by a mapping $f : Y \times Z \to [0,1]$, which determines the transmission probability of user i in slot t as

$$p_i^t = f\left(y_i^{t-1}, z_i^t\right),$$

for $t = 1, 2, \ldots$. We set $y_i^0 = idle$ as initialization. Note that adaptive protocols can be regarded as an extension of protocols with one-slot memory [2] in that adaptive protocols allow transmission parameters to adjust to an exogenous state variable (the traffic type in this paper). In other words, with an adaptive protocol a user can change its modes of operation depending on its state.

We define a critical phase as a period that begins with an occurrence of a critical event and ends with the completion of the transmission of critical traffic associated with the critical event. We define a normal phase as a period without critical traffic, between two critical phases. Given an adaptive protocol f, in a normal phase a successful user has another success in the next slot with probability $f(\text{success}, \text{normal})(1 - f(\text{busy}, \text{normal}))^{N-1}$. The average number of consecutive successes by a user starting from an initial success in a normal phase is denoted by T_s and is given by

$$T_s = \frac{1}{1 - f(\text{success}, \text{normal})\,(1 - f(\text{busy}, \text{normal}))^{N-1}}.$$

A protocol yielding a large value of T_s can be considered as unfair in the short term because it suppresses the transmission opportunities of other waiting users once a success occurs. Hence, we define the fairness level of an adaptive protocol, denoted by F_{norm}, as the inverse of the average number of consecutive successes in a normal phase,

$$\begin{aligned} F_{\text{norm}} &= \frac{1}{T_s} \\ &= 1 - f(\text{success}, \text{normal})(1 - f(\text{busy}, \text{normal}))^{N-1}. \end{aligned}$$

We say that an adaptive protocol is θ-fair if $F_{\text{norm}} = \theta$, for $\theta \in (0,1]$. Also, we say that an adaptive protocol f is nonintrusive if $f(y, \text{critical}) = 1$ for all $y \in Y$ and $f(\text{busy}, \text{normal}) = 0$. A nonintrusive protocol guarantees that once a critical user has a successful transmission, its transmission is not interrupted by other users (with normal traffic) until it completes the transmission of its critical traffic.

In this paper, we restrict our attention to the class of θ-fair nonintrusive adaptive protocols. Setting $F_{\text{norm}} = \theta$ together with $f(\text{busy}, \text{normal}) = 0$ implies $f(\text{success}, \text{normal}) = 1 - \theta$. Thus, a θ-fair nonintrusive adaptive protocol can be expressed as

$$\begin{aligned} f(y, \text{critical}) &= 1 \text{ for all } y \in Y, \\ f(\text{idle}, \text{normal}) &= q, f(\text{busy}, \text{normal}) = 0, \\ f(\text{success}, \text{normal}) &= 1 - \theta, f(\text{failure}, \text{normal}) = r, \end{aligned}$$

for some $q, r \in [0,1]$.

Remark 1: Reference [12] defines that a protocol is M-short-term fair if $T_s \leq M$. Reference [2] captures short-term fairness by considering average delay, which is defined as the average waiting time of a user until its next success starting from an arbitrary point of time. Although the concept of average delay is more comprehensive than the average number of consecutive successes (as delay can be created by reasons other than consecutive successes), we use the latter in this paper because it is much simpler to compute and is a good proxy for the former in the case of protocols with one-slot memory.

Remark 2: The operation of a θ-fair nonintrusive adaptive protocol in a normal phase is analogous to p-persistent CSMA [1]. Users wait when the channel is sensed busy and transmit with probability q and r following an idle slot and a collision, respectively. Since we consider saturated arrivals where each user always has packets to transmit, we introduce θ as a stopping probability in order to prevent a single user from using the channel exclusively.

III. Performance Metrics

A. Definitions

1) Channel Utilization Rate in a Normal Phase: The channel utilization rate of users in a normal phase is defined as the proportion of time slots in which a successful transmission occurs during a normal phase and is given by

$$C_{\text{norm}} = \frac{\text{Number of successes in a normal phase}}{\text{Length of a normal phase}}.$$

2) Delay in a Critical Phase: Let L be the average length of critical traffic, measured in slots. A protocol determines the average number of slots that a critical phase lasts, denoted by T_{crit}. The delay in a critical phase is defined as the average number of nonsuccess slots for a critical user during a critical phase

$$D_{\text{crit}} = T_{\text{crit}} - L.$$

B. Computation of the Channel Utilization Rate in a Normal Phase

Consider a slot t in which a normal phase begins. Since slot $t-1$ is the last slot of a critical phase, there exists a user i that completed the transmission of its critical traffic in slot $t-1$. Since user i had a successful transmission in slot $t-1$, we have $y_i^{t-1} = \text{success}$ and $y_j^{t-1} = \text{busy}$ for all $j \neq i$, and thus in slot t user i transmits with probability $1-\theta$ while other users wait. Hence, a normal phase begins with a success by the user that had critical traffic in the previous critical phase with probability $1-\theta$ and with an idle slot with probability θ.[4]

To compute the channel utilization rate in a normal phase C_{norm}, we construct a Markov chain whose state space is $\{0, 1, \ldots, N\}$, where state k represents transmission outcomes in which exactly k users transmit. The transition probability

[4] If we extend adaptive protocols as $p_i^t = f(z_i^{t-1}, y_i^{t-1}, z_i^t)$, we can set $p_i^t = 0$ if $z_i^{t-1} = \text{critical}$ and $z_i^t = \text{normal}$. Then all users including the user that had critical traffic wait in the first slot of a normal phase, which makes users contend with an equal transmission probability in the second slot.

from state k to state k' in a normal phase, $P_{\rm norm}(k'|k)$, under a θ-fair nonintrusive adaptive protocol is given by

$$P_{\rm norm}(k'|0) = \binom{N}{k'} q^{k'}(1-q)^{N-k'} \quad \text{for } k'=0,\dots,N, \tag{1}$$

$$P_{\rm norm}(k'|1) = \begin{cases} \theta & \text{for } k'=0 \\ 1-\theta & \text{for } k'=1 \\ 0 & \text{for } k'=2,\dots,N, \end{cases}$$

$$P_{\rm norm}(k'|k) = \begin{cases} \binom{k}{k'} r^{k'}(1-r)^{k-k'} & \text{for } k'=0,\dots,k \\ 0 & \text{for } k'=k+1,\dots,N, \end{cases} \quad \text{for } k=2,\dots,N. \tag{2}$$

The transition matrix of the Markov chain can be written in the form of

$$\mathbf{P}_{\rm norm} = \begin{array}{c} \\ 0 \\ 2 \\ \vdots \\ N-1 \\ N \\ 1 \end{array} \begin{array}{c} \begin{array}{cccccc} 0 & 2 & \dots & N-1 & N & \quad 1 \end{array} \\ \left(\begin{array}{ccccc|c} * & * & \cdots & * & * & * \\ * & * & \cdots & 0 & 0 & * \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ * & * & \cdots & * & 0 & * \\ * & * & \cdots & * & * & * \\ \hline \theta & 0 & \cdots & 0 & 0 & 1-\theta \end{array}\right) \end{array}$$

where the entries marked with an asterisk can be found in (1) and (2).

The average number of consecutive successes in a normal phase, T_s, is determined by the fairness level θ, where the relationship is given by $T_s = 1/\theta$. A series of consecutive successes by a user ends with an idle slot, when the successful user waits. Let T_c be the average number of nonsuccess slots until the first success starting from an idle slot during a normal phase. A normal phase can be considered as the alternation of a success period and a contention period, which is continued until a critical event occurs. A success period is characterized by consecutive successes by a user, whereas a contention period begins with an idle slot and lasts until a user succeeds. Since all users transmit with the same transmission probability following an idle slot, each user has an equal chance of becoming a successful user for the following success period at the point when a contention period starts. In other words, a contention period selects the next successful user in a nondiscriminatory way.

Let $\mathbf{Q}_{\rm norm}$ be the N-by-N matrix in the upper-left corner of $\mathbf{P}_{\rm norm}$. Suppose that $0 < q,\ r < 1$ so that all the entries of $\mathbf{P}_{\rm norm}$ marked with an asterisk are nonzero. Then $(\mathbf{I} - \mathbf{Q}_{\rm norm})^{-1}$ exists and is called the fundamental matrix for $\mathbf{P}_{\rm norm}$, when state 1 is absorbing (i.e., $\theta = 0$) [13]. The average number of slots in state $k \neq 1$ starting from state 0 (an idle slot) is given by the $(1,k)$-entry of $(\mathbf{I}-\mathbf{Q}_{\rm norm})^{-1}$. Hence, the average number of slots to hit state 1 (a success slot) for the first time starting from an idle slot is given by the first entry of $(\mathbf{I}-\mathbf{Q}_{\rm norm})^{-1}\mathbf{e}$, where $\mathbf{e}$ is a column vector of length N all of whose entries are 1. Hence, we obtain $T_c = [(\mathbf{I}-\mathbf{Q}_{\rm norm})^{-1}\mathbf{e}]_1$. Note that T_c is independent of θ. That is, the average duration of a contention period is not affected by the average duration of a success period. The channel utilization rate of users in a normal phase can be computed by

$$C_{\rm norm} = \frac{T_s}{T_c + T_s} = \frac{1}{\theta\left[(\mathbf{I}-\mathbf{Q}_{\rm norm})^{-1}\mathbf{e}\right]_1 + 1}, \tag{3}$$

for $(q,r) \in (0,1)^2$.

An alternative method to compute the channel utilization rate in a normal phase is to use a stationary distribution. Since $\theta \in (0,1]$, all states communicate with each other under the transition matrix $\mathbf{P}_{\rm norm}$ for all $(q,r) \in (0,1)^2$. Hence, the Markov chain is irreducible, and there exists a unique stationary distribution $\mathbf{w}_{\rm norm}$, which satisfies

$$\mathbf{w}_{\rm norm} = \mathbf{w}_{\rm norm}\mathbf{P}_{\rm norm} \quad \text{and} \quad \mathbf{w}_{\rm norm}\mathbf{e} = 1. \tag{4}$$

Let $w_{\rm norm}(k)$ be the entry of $\mathbf{w}_{\rm norm}$ corresponding to state k, for $k = 0,1,\dots,N$. Then $w_{\rm norm}(k)$ gives the probability of state k during a normal phase. In particular, the channel utilization in a normal phase is given by $w_{\rm norm}(1)$. Since success and contention periods alternate from the beginning of a normal phase, the stationary distribution yields the probabilities of states for any duration of a normal phase (assuming that a normal phase lasts sufficiently longer than $T_s + T_c$), not just the limiting probabilities as a normal phase lasts infinitely long. By manipulating (4), we can derive that $w_{\rm norm}(1) = C_{\rm norm}$, whose expression is given in (3).

C. Computation of the Delay in a Critical Phase

Consider a slot t in which a critical phase begins. Since a critical user always transmits under a nonintrusive protocol, we consider a Markov chain whose state space is $\{0,1,\dots,N-1\}$, where state k represents transmission outcomes in which exactly k normal users transmit. The transition probability from state k to state k' in a critical phase, $P_{\rm crit}(k'|k)$, under a θ-fair nonintrusive adaptive protocol is given by

$$P_{\rm crit}(k'|k) = \begin{cases} \binom{k}{k'} r^{k'}(1-r)^{k-k'} & \text{for } k'=0,\dots,k \\ 0 & \text{for } k'=k+1,\dots,N-1, \end{cases} \quad \text{for } k=0,\dots,N-1. \tag{5}$$

The transition matrix of the Markov chain can be written in the form of

$$\mathbf{P}_{\rm crit} = \begin{array}{c} \\ 1 \\ 2 \\ \vdots \\ N-2 \\ N-1 \\ 0 \end{array} \begin{array}{c} \begin{array}{cccccc} 1 & 2 & \dots & N-2 & N-1 & \quad 0 \end{array} \\ \left(\begin{array}{ccccc|c} * & 0 & \cdots & 0 & 0 & * \\ * & * & \cdots & 0 & 0 & * \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ * & * & \cdots & * & 0 & * \\ * & * & \cdots & * & * & * \\ \hline 0 & 0 & \cdots & 0 & 0 & 1 \end{array}\right) \end{array}$$

where the entries marked with an asterisk can be found in (5). Note that state 0, which corresponds to a success by the critical user, is absorbing because once the critical user has a successful transmission, its transmissions in the following slots are not interrupted by other users with normal traffic. Hence, the delay in a critical phase under a θ-fair nonintrusive adaptive protocol is independent of the length of critical traffic and is measured by the average number of collisions that the critical user experiences before obtaining a successful transmission. Let $\mathbf{Q}_{\text{crit}}$ be the $(N-1)$-by-$(N-1)$ matrix in the upper-left corner of $\mathbf{P}_{\text{crit}}$. For $r \in [0,1)$, the matrix $\mathbf{I}-\mathbf{Q}_{\text{crit}}$ is invertible, and the average number of slots until the first success starting from state k is given by the kth entry of $(\mathbf{I}-\mathbf{Q}_{\text{crit}})^{-1}\mathbf{e}$, for $k=1,\ldots,N-1$.

The number of collisions that a critical user experiences in a critical phase depends on the transmission outcome in slot $t-1$, the last slot of the preceding normal phase. We represent the transmission outcome of slot $t-1$ by a pair (l,a), where l is the number of transmissions by users other than the user that becomes a critical user in the following critical phase and a is the transmission action of the user. We write $a=T$ if the user transmits and $a=W$ if it waits. Suppose that the transmission outcome of slot $t-1$ is represented by (l,a) with $l \geq 2$. Then the Markov chain starts from state l in slot $t-1$, regardless of a. Since the critical phase starts in slot t, the number of collisions in the critical phase does not include the collision in slot $t-1$. Hence, the average number of collisions until the first success in a critical phase when the preceding normal phase ended with l transmissions by users other than the critical user is given by

$$d(l,a) = \left[(\mathbf{I}-\mathbf{Q}_{\text{crit}})^{-1}\mathbf{e}\right]_l - 1,$$

for $l=2,\ldots,N-1$ and $a=T,W$.

Suppose that the transmission outcome of slot $t-1$ is represented by $(1,a)$. If $a=T$, then the Markov chain starts from state 1 in slot $t-1$, and the average number of collisions until the first success in a critical phase when the preceding normal phase ended with two transmissions including one by the critical user is given by

$$d(1,T) = \left[(\mathbf{I}-\mathbf{Q}_{\text{crit}})^{-1}\mathbf{e}\right]_1 - 1.$$

If $a=W$, then there is a successful user, different from the critical user in the following critical phase, in slot $t-1$. The successful user transmits with probability $1-\theta$ while all the other normal users wait in slot t. Thus, with probability θ, the critical user succeeds in slot t, and with probability $1-\theta$, state 1 occurs in slot t, from which it takes $[(\mathbf{I}-\mathbf{Q}_{\text{crit}})^{-1}\mathbf{e}]_1$ collisions on average to reach a success by the critical user. Therefore, the average number of collisions until the first success in a critical phase when the preceding normal phase ended with a success by a user other than the critical user is given by

$$\begin{aligned} d(1,W) &= \theta \cdot 0 + (1-\theta)\left[(\mathbf{I}-\mathbf{Q}_{\text{crit}})^{-1}\mathbf{e}\right]_1 \\ &= (1-\theta)\left[(\mathbf{I}-\mathbf{Q}_{\text{crit}})^{-1}\mathbf{e}\right]_1. \end{aligned}$$

Suppose that the transmission outcome of slot $t-1$ is represented by $(0,a)$. If $a=T$, then the critical user in the following critical phase has a success in slot $t-1$. Since $f(busy, normal)=0$, all normal users wait in slot t. Thus, the critical user has another success in slot t, which leads to zero delay in a critical phase when the preceding normal phase ended with a success by the critical user, i.e.,

$$d(0,T)=0.$$

If $a=W$, then all $(N-1)$ normal users transmit with probability q in slot t. Then with probability $\binom{N-1}{k}q^k(1-q)^{N-1-k}$, slot t contains transmission by k normal users, for $k=0,\ldots,N-1$. With probability $(1-q)^{N-1}$ the critical user experiences no collision while with probability $\binom{N-1}{k}q^k(1-q)^{N-1-k}$ the critical phase begins with state k, for $k=1,\ldots,N-1$. Therefore, the average number of collisions until the first success in a critical phase when the preceding normal phase ended with an idle slot is given by

$$\begin{aligned} d(0,W) =& (1-q)^{N-1}\cdot 0 + \sum_{k=1}^{N-1}\binom{N-1}{k}q^k(1-q)^{N-1-k} \\ &\times \left[(\mathbf{I}-\mathbf{Q}_{\text{crit}})^{-1}\mathbf{e}\right]_k \\ =& \sum_{k=1}^{N-1}\binom{N-1}{k}q^k(1-q)^{N-1-k}\left[(\mathbf{I}-\mathbf{Q}_{\text{crit}})^{-1}\mathbf{e}\right]_k. \end{aligned}$$

As discussed in Section III-B, the probability that the last slot of a normal phase has k transmissions is given by $w_{\text{norm}}(k)$, for $k=0,1,\ldots,N$. Since we consider symmetric protocols, the probability that a particular user is one of k transmitting users is given by k/N. Thus, the probability that the transmission outcome of the last slot of a normal phase is represented by (l,a), denoted by $v(l,a)$, is given by

$$v(l,T) = \frac{l+1}{N}w_{\text{norm}}(l+1) \text{ and } v(l,W) = \frac{N-l}{N}w_{\text{norm}}(l),$$

for $l=0,\ldots,N-1$. Then the delay in a critical phase can be computed as

$$D_{\text{crit}} = \sum_{l\in\{0,\ldots,N-1\}, a\in\{T,W\}} v(l,a)d(l,a).$$

IV. Protocol Design Problem and Optimal Protocols

A. Formulation of the Protocol Design Problem

We formulate a problem solved by the protocol designer based on the following assumptions. First, we assume that the protocol designer has the most preferred fairness level and considers the class of θ-fair nonintrusive adaptive protocols, where θ is chosen as the most preferred fairness level. Second, the protocol designer prefers a higher channel utilization rate in a normal phase. Third, the protocol designer has a threshold level $\eta > 0$ that he desires the delay in a critical phase does not exceed. Finally, the protocol designer considers q and r on a restricted domain, $(q,r) \in [\epsilon, 1-\epsilon]^2$ for a small $\epsilon > 0$. Then the protocol design problem can be formally expressed as

$$\max_{(q,r)\in[\epsilon,1-\epsilon]^2} C_{\text{norm}} \text{ subject to } D_{\text{crit}} \leq \eta. \tag{6}$$

Note that the results in Section III imply that, for a given fairness level $\theta \in (0,1]$, C_{norm} and D_{crit} are continuous functions of (q,r) on the interior of $[0,1]^2$. Thus, a solution to the protocol

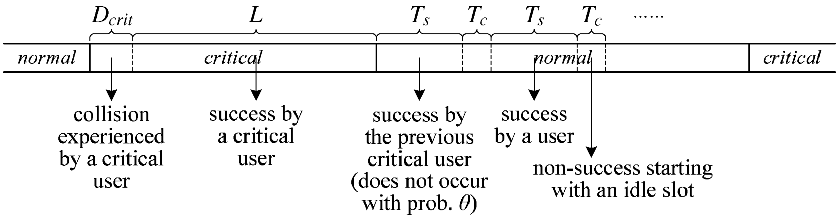

D_{crit}
L
T_s
T_c
T_s
T_c
......
normal
critical
normal
critical
collision experienced by a critical user
success by a critical user
success by the previous critical user (does not occur with prob. θ)
success by a user
non-success starting with an idle slot

Fig. 1. Operation of the system under a θ-fair nonintrusive adaptive protocol.

design problem (6) always exists, and we say that a protocol is optimal if it solves (6).[5]

Fig. 1 summarizes the operation of the system under a θ-fair nonintrusive adaptive protocol. Note that L is exogenously given while D_{crit}, T_s, and T_c are determined by the protocol specification. Since T_s is determined completely by the fairness level, the protocol design problem can be restated as to minimize T_c while keeping D_{crit} below a certain threshold level. Finally, we note that we can achieve $D_{\text{crit}} = 0$ and $T_c = 0$ (and thus $C_{\text{norm}} = 1$) for any fairness level if we allow coordination messages. Hence, the gaps between D_{crit} and 0 and between C_{norm} and 1 that arise when we are restricted to use distributed protocols without message passing can be considered as a performance loss due to the lack of explicit message passing.

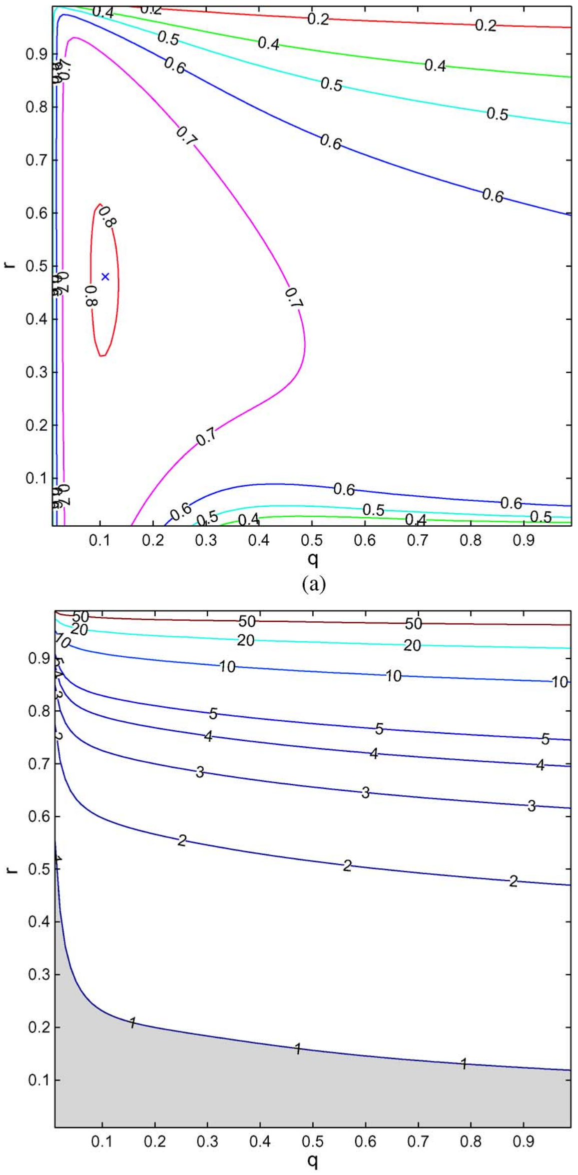
B. Graphical Illustration of the Protocol Design Problem

In the remainder of this section, we illustrate the solution to the protocol design problem using numerical examples. Throughout this section, we set $\epsilon = 0.01$. In Fig. 2, we show the dependence of the performance metrics, C_{norm} and D_{crit}, on (q, r). To obtain the results, we consider ten users, i.e., $N = 10$, and fix $\theta = 0.1$ so that $T_s = 10$. Fig. 2(a) plots the contour curves of C_{norm}. Let $(q^*, r^*) = \arg\max_{(q,r)\in[\epsilon,1-\epsilon]^2} C_{\text{norm}}$. That is, (q^*, r^*) represents the θ-fair nonintrusive adaptive protocol that maximizes the channel utilization rate in a normal phase when no constraint is imposed on the delay in a critical phase. With numerical methods, we find that (q^*, r^*) is unique with the value (0.105, 0.479) and achieves 0.804 as the maximum value of C_{norm}. By (3), C_{norm} and T_c are negatively related for a given fairness level θ, and the minimum value of T_c corresponding to the maximum value of C_{norm} is given by 2.44. That is, at (q^*, r^*), a contention period in a normal phase lasts for 2.44 slots on average, while the average duration of a success period is given by $1/\theta$ slots. The value of (q^*, r^*) can be explained as follows. Following an idle slot in a normal phase, every user transmits with probability q, and thus the probability of success is maximized when $q = 1/N$. Hence, q^* is chosen close to $1/N$. During a normal phase, a collision cannot follow a success, and following an idle slot, a collision involving two transmissions is most likely among all kinds of collisions when $q \approx 1/N$. Since noncolliding users do not transmit following a collision under a nonintrusive protocol, the probability of success between two contending users is maximized when $r = 1/2$. r^* is chosen slightly smaller than 1/2 because collisions involving more than two transmissions occur with small probability. Fig. 2(b) plots the contour curves of D_{crit}. As q and r are large, users transmit aggressively during a contention period in a normal phase, intensifying interference to a critical user before its first success. Thus, D_{crit} is increasing in both q and r. The set of (q, r) that satisfies the delay constraint $D_{\text{crit}} \leq \eta$ can be represented by the region below the contour curve of D_{crit} at level η. For example, the shaded area in Fig. 2(b) represents the constraint set corresponding to $D_{\text{crit}} \leq 1$.

Fig. 3 shows the contour curves of C_{norm} and D_{crit} in the same graph to illustrate the protocol design problem (6). The protocol design problem is to find the largest value of C_{norm} on the region of (q, r) that satisfies $D_{\text{crit}} \leq \eta$. Let η^* be the value of D_{crit} at (q^*, r^*). With $N = 10$ and $\theta = 0.1$, we obtain $\eta^* = 1.531$. We say that a constraint is binding if its removal results in a strict improvement in the objective value and nonbinding otherwise. Then the delay constraint is binding if $\eta < \eta^*$ and nonbinding if $\eta \geq \eta^*$. For example, if $\eta = 1$, the constraint is binding and the optimal protocol is given by the point on the contour curve of D_{crit} at level 1, marked with '+' in Fig. 3, where a contour curve of D_{crit} and that of C_{norm} are tangent to each other. In contrast, if $\eta = 2$, the constraint is nonbinding and the optimal protocol is given by the solution to the unconstrained problem, $(q^*, r^*) = (0.105, 0.479)$, marked with '×' in Fig. 3.

Fig. 4 shows the solutions to the protocol design problem for η between 0.1 and 2. Fig. 4(a) plots optimal protocols, denoted by (q^o, r^o), as η varies while Fig. 4(b) shows the values of D_{crit} and C_{norm} at the optimal protocols. We can divide the range of η into three regions: (0, 0.71], (0.71, 1.53), and $[1.53, \infty)$. For $\eta \leq 0.71$, the optimal protocol occurs at the corner with $r^o = \epsilon$. As η decreases in this region, q^o decreases to ϵ while r^o stays at ϵ, which makes C_{norm} decrease to 0. Smaller η means that higher priority is given to a critical user, and this can be achieved by inhibiting transmissions by users when they have normal traffic. For $\eta \in (0.71, 1.53)$, the solution to the protocol design problem is interior while the constraint $D_{\text{crit}} \leq \eta$ is still binding. The trade-off between D_{crit} and C_{norm} is less severe in this region than in (0, 0.71]. Reducing η from 1.53 to 0.71 results in a slight decrease in C_{norm} from 0.80 to 0.76. For $\eta \geq 1.53$, the constraint $D_{\text{crit}} \leq \eta$ is nonbinding, and thus (q^o, r^o) remains at $(q^*, r^*) = (0.105, 0.479)$ while C_{norm} remains at its unconstrained maximum level, 0.804. The rate of change in the maximum value of C_{norm} with respect to η suggests that keeping D_{crit} below 0.71 induces a large cost in terms of the

[5] The reason that we use a restricted domain for (q, r), $[\epsilon, 1-\epsilon]^2$, instead of $[0, 1]^2$ is to have the computation method for C_{norm} and D_{crit} in Section III valid and thus to make the protocol design problem more tractable. This restriction leads to a small performance loss when ϵ is small. Note that $q = 0$ and $r = 1$ cannot be optimal since $q = 0$ yields $C_{\text{norm}} = 0$ and $r = 1$ yields $D_{\text{crit}} = +\infty$. The computation method is valid for $(q, r) \in (0, 1] \times [0, 1)$, and thus the continuity property of C_{norm} and D_{crit} holds on the boundary where $q = 1$ or $r = 0$ as ϵ goes to zero.



(a)

(b)

Fig. 2. Contour curves of C_{norm} and D_{crit} as functions of (q, r) when $N = 10$ and $\theta = 0.1$. (a) C_{norm}; (b) D_{crit}.

reduced channel utilization rate in a normal phase, maintaining D_{crit} between 0.71 and 1.53 only a minor cost, and tolerating D_{crit} larger than 1.53 no cost. In other words, when the optimal solution to the protocol design problem is interior, the optimal dual variable on the constraint $D_{\text{crit}} \leq \eta$ is close to zero or is zero.

C. Varying the Number of Users

We examine how the optimal protocol changes as the number of users varies between 3 and 50. We fix $\theta = 0.1$ as before. We first solve the protocol design problem with a nonbinding constraint, assuming that η is sufficiently large. Fig. 5(a) shows optimal protocols (q^*, r^*) when the delay constraint is nonbinding. As N increases from 3 to 50, q^* decreases from 0.34 to 0.02 while r^* decreases from 0.49 to 0.48. Fig. 5(b) plots the values of D_{crit} and C_{norm} at (q^*, r^*). As N increases from 3 to 50, D_{crit} increases from 1.18 to 1.65 while C_{norm} decreases from 0.82 to 0.80. The results show that when the delay constraint is nonbinding, the delay in a critical phase increases at a diminishing rate as the number of users increases, while the channel utilization rate in a normal phase remains almost constant. Almost constant C_{norm} implies that the optimal protocols are capable of resolving contention among users efficiently in a normal phase even if there are many users sharing the channel. The values of D_{crit} at (q^*, r^*) can be interpreted as the minimum values of η that make the delay constraint nonbinding.

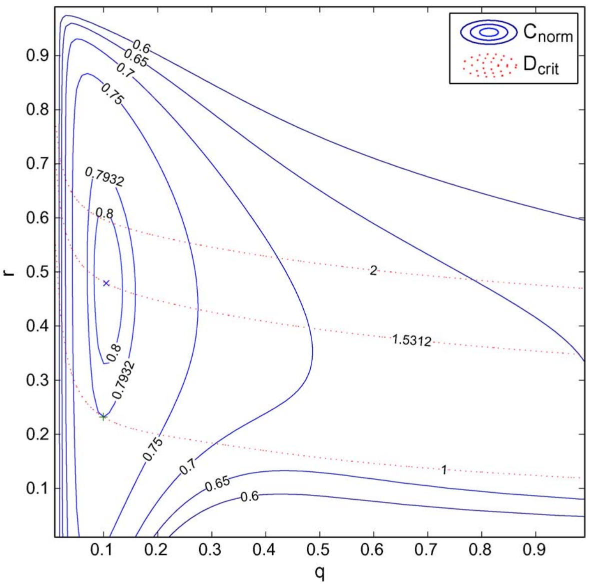


Fig. 3. Illustration of optimal protocols.

Now we set $\eta = 1$ so that the delay constraint is binding for all N between 3 and 50. Fig. 5(a) shows optimal protocols (q^o, r^o) when the delay constraint is given by $D_{\text{crit}} \leq 1$. As N increases from 3 to 50, q^o decreases from 0.34 to 0.02 while r^o decreases from 0.40 to 0.18. Imposing the constraint $D_{\text{crit}} \leq 1$ limits the values of q and r, but it impacts r more than q, i.e., $r^o < r^*$ and $q^o \approx q^*$, for given N, due to the shape of the contour curves of C_{norm} as illustrated in Fig. 3. Fig. 5(b) plots the values of D_{crit} and C_{norm} at (q^o, r^o). As N increases from 3 to 50, D_{crit} stays at 1, confirming that the constraint $D_{\text{crit}} \leq 1$ is binding, while C_{norm} decreases from 0.82 to 0.78. We can see that requiring $D_{\text{crit}} \leq 1$ decreases the maximum values of C_{norm} only slightly because the delay constraint with $\eta = 1$ is mild so that the optimal protocols remain interior. If we impose a sufficiently strong constraint, i.e., choose a small η, then we have the optimal protocol at the corner, $q^o < q^*$ and $r^o = \epsilon$, and C_{norm} is reduced significantly, as suggested in Fig. 4.

D. Varying the Fairness Level

We investigate the impact of the fairness level on optimal protocols and their performance. We first consider sufficiently large η so that the delay constraint is nonbinding. Fig. 6(a) shows optimal protocols (q^*, r^*) when the constraint is nonbinding. Since maximizing C_{norm} is equivalent to minimizing T_c, which is independent of θ, the optimal protocols do not depend on θ when the delay constraint is nonbinding. Fig. 6(b) plots the

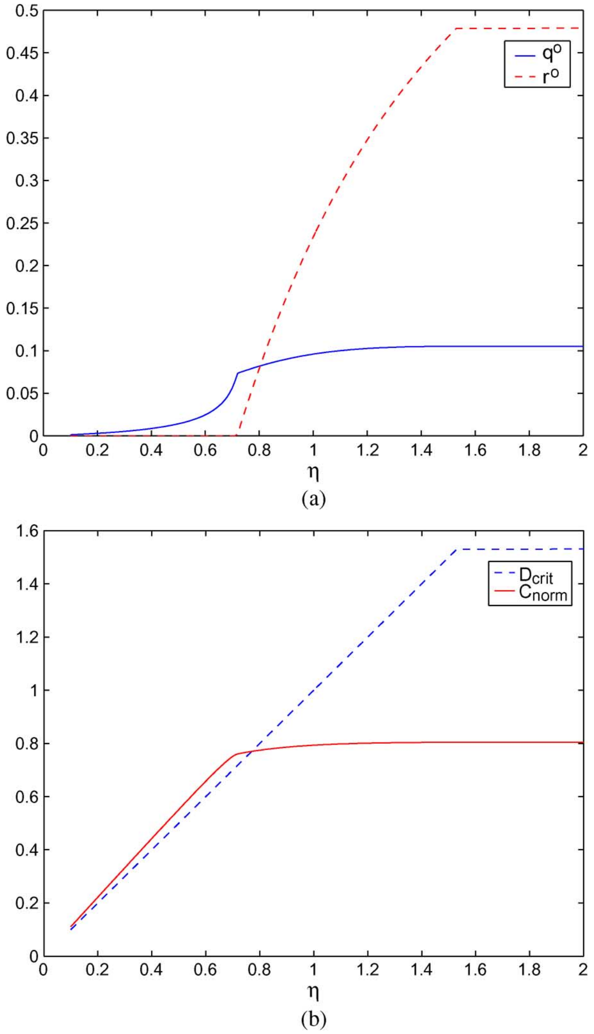


Fig. 4. Solution to the protocol design problem for η between 0.1 and 2 when $N = 10$ and $\theta = 0.1$: (a) optimal protocols, and (b) the values of D_{crit} and C_{norm} at the optimal protocols.

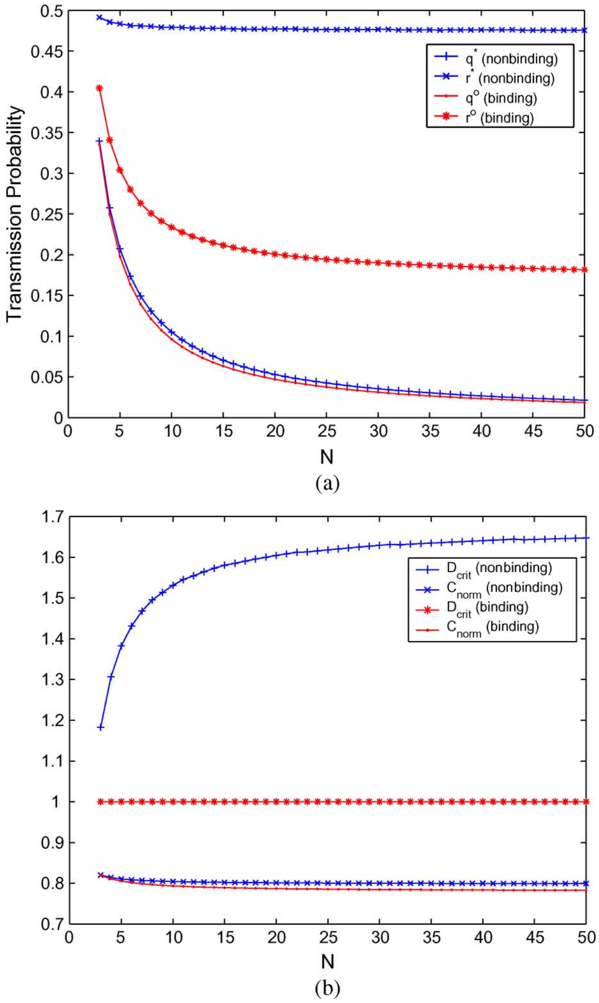


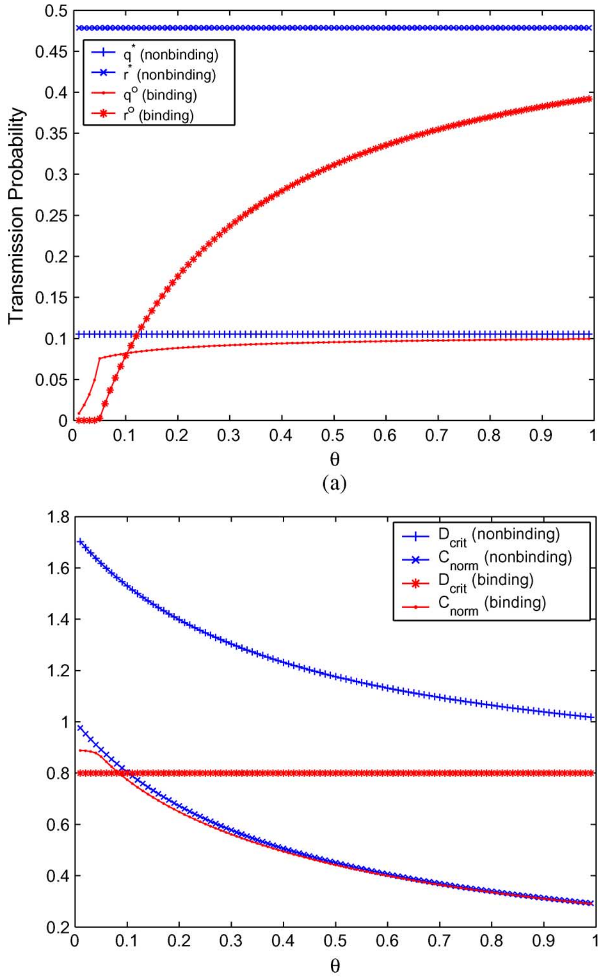
Fig. 5. Solution to the protocol design problem for N between 3 and 50 when $\theta = 0.1$: (a) optimal protocols, and (b) the values of D_{crit} and C_{norm} at the optimal protocols.

values of D_{crit} and C_{norm} at (q^*, r^*). From the expression in (3), we can see that C_{norm} is decreasing in θ. D_{crit} is also decreasing in θ because increasing θ induces idle slots to occur more frequently, from which the average number of collisions experienced by a critical user is small.

Since D_{crit} at (q^*, r^*) ranges between 1.02 and 1.70, we set $\eta = 0.8$ to analyze the protocol design problem with a binding delay constraint. Fig. 6(a) shows optimal protocols (q^o, r^o) with $\eta = 0.8$ while Fig. 6(b) plots the values of D_{crit} and C_{norm} at the optimal protocols. Note that the optimal protocols are at the corner with $r^o = \epsilon$ for $\theta \leq 0.04$. Imposing the constraint $D_{\mathrm{crit}} \leq 0.8$ limits the values of q and r. The decrease in q and r is larger when θ is smaller because requiring $D_{\mathrm{crit}} \leq 0.8$ imposes a stronger constraint for smaller θ, which can be seen by comparing the values of D_{crit} with binding and nonbinding delay constraints. However, the impact on C_{norm} is marginal as long as the optimal protocols are interior.

E. Estimated Number of Users

So far we have assumed that users know the exact number of users sharing the channel. We relax this assumption and consider a scenario where users follow optimal protocols computed based on their (possibly incorrect) estimates of the number of users. We investigate the consequence of using estimates instead of the exact number of users when computing optimal protocols. For simplicity, we assume that all users have the same estimate. We consider $N = 10$ and the estimated number of users, denoted by $\hat{N}$, between 5 and 15. In Fig. 7, we plot the values of D_{crit} and C_{norm} when N users follow the optimal protocol designed for $\hat{N}$ users. As before, we consider the two cases of nonbinding and binding delay constraints, with $\eta = 1$ for the binding constraint. In both cases, optimal q and r decrease with the estimated number of users in order to accommodate increased contention from more users, as shown in Fig. 5(a). Hence, D_{crit} decreases with $\hat{N}$ since interference from normal users is reduced as $\hat{N}$ increases. C_{norm} is not affected much by $\hat{N}$, reaching a peak when $\hat{N} = N$. This result suggests that the performance in a normal phase is robust to errors in the estimation of the number of users. Note that, in the case of the binding delay constraint, the constraint is violated when an underestimation occurs, i.e., $\hat{N} < N$. To have the binding delay constraint satisfied with equality, we can use an adjustment procedure for $\hat{N}$ that drives D_{crit} toward η. An estimation procedure can be



(a)

(b)

Fig. 6. Solution to the protocol design problem for θ between 0.01 and 0.99 when $N = 10$: (a) optimal protocols, and (b) the values of $D_{\rm crit}$ and $C_{\rm norm}$ at the optimal protocols.

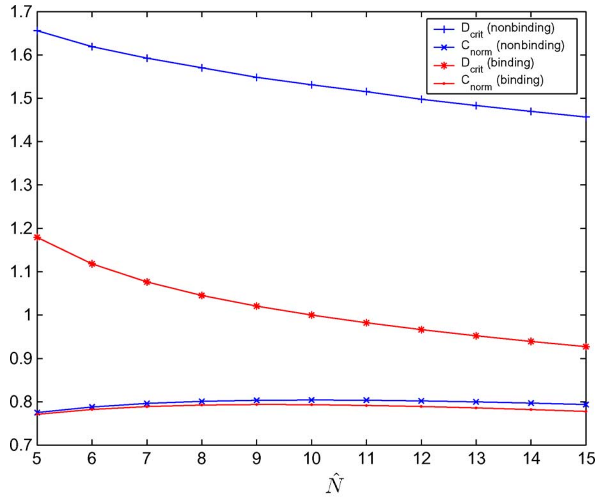


Fig. 7. Values of $D_{\rm crit}$ and $C_{\rm norm}$ for $\hat{N}$ between 5 and 15 when $N = 10$ and $\theta = 0.1$.

designed based on the approach of [14], whose details are left for future work.

V. Enhancement of Adaptive Protocols

In this section, we discuss improvements on adaptive protocols by utilizing longer memory. The main idea is the inference of the traffic types of other users based on the patterns of observations. Some patterns of observations reveal information about the types of other users, and users can adjust their transmission parameters to these patterns. As users maintain longer memory, there are more recognizable patterns, and exploiting them can yield performance improvement.

A. Reducing the Average Delay

Returning to the discussion in Section III-C, suppose that the transmission outcome of the last slot of a normal phase is represented by $(1, W)$ so that there is a successful user different from the critical user in the following critical phase. When users follow a θ-fair nonintrusive adaptive protocol, the successful user transmits with probability $1 - \theta$ in the first slot of the following critical phase. If the successful user transmits in the first slot, it collides with the critical user and then transmits with probability r in the next slot. However, a collision cannot follow a success when all users have normal traffic, and thus after a collision in the first slot of the critical phase, the successful user can infer the existence of a critical user. If the protocol is modified so that it requires normal users to wait after experiencing a pattern of success followed by failure, then the average number of collisions experienced by a critical user after a transmission outcome represented by $(1, W)$, $d(1, W)$, is reduced from $(1 - \theta)[(\mathbf{I} - \mathbf{Q}_{\rm crit})^{-1}\mathbf{e}]_1$ to $1 - \theta$. When θ is small, a success period lasts long in a normal phase, leading to a large weight on $(1, W)$, $v(1, W)$. Thus, requiring normal users to wait after $(\text{success}, \text{failure})$ reduces the delay in a critical phase significantly. For example, with $N = 10$, $\theta = 0.1$, and $(q, r) = (q^*, r^*) = (0.105, 0.479)$, we have $d(1, W)$ reduced from 1.73 to 0.9, which decreases $D_{\rm crit}$ from 1.53 to 0.93.

B. Bounding the Maximum Delay

In the range of parameter values considered in Section IV, the delay in a critical phase is reasonably small, not exceeding two slots. However, the realized number of collisions that a critical user experiences can be arbitrarily large with positive probability. That is, the worst-case delay in a critical phase is unbounded. We can bound the maximum delay by modifying the protocol so that it requires normal users to wait after experiencing B consecutive collisions. Since noncolliding normal users wait after a collision, colliding normal users must have the same number of consecutive collisions in any slot. Thus, normal users experiencing B consecutive collisions back off simultaneously, yielding a room for a critical user, if there is one. Therefore, a critical user cannot experience more than B collisions in a critical phase. When B is chosen moderately large, B consecutive collisions rarely occur in a normal phase, and thus the proposed modification has a negligible impact on the channel utilization rate in a normal phase, $C_{\rm norm}$. We summarize below the enhanced adaptive protocols including the feature discussed in footnote 4.

1) If $y_i^{t-2} = \text{success}$, $y_i^{t-1} = \text{failure}$, and $z_i^t = \text{normal}$, then $p_i^t = 0$.

2) If $y_i^{t-B} = \cdots = y_i^{t-1} = \text{failure}$ and $z_i^t = \text{normal}$, then $p_i^t = 0$.
3) If $z_i^{t-1} = \text{critical}$ and $z_i^t = \text{normal}$, then $p_i^t = 0$.
4) Otherwise, $p_i^t = f(y_i^{t-1}, z_i^t)$, where f is a θ-fair nonintrusive adaptive protocol.

C. Two Users With Critical Traffic

We now consider a scenario where the system can have up to two critical users at the same time. We describe how the enhanced adaptive protocols can be extended to accommodate two critical users. Depending on the timing of the two arrivals of critical traffic, we analyze three cases where two critical users coexist.

First, suppose that a second critical event occurs to user j in slot t while a critical user i is having successful transmissions in a critical phase. Since the traffic type of a user is its local information, user j does not know whether or not there is another critical user in slot t.[6] We propose a protocol with which a user transmits when its traffic type changes from normal to critical so that user j transmits in slot t. As in Section V-A, the transmission by user j informs user i that there exists another critical user in the system. If user i had normal traffic, it would respond by waiting in slot $t+1$ according to the enhanced adaptive protocol so that user j could capture the channel. However, since user i has critical traffic, we propose a protocol that makes user i respond by transmitting in slot $t+1$ to inform user j of its critical traffic. Then after the implicit information exchange in slots t and $t+1$, both user i and user j know that there are two critical users. From slot $t+2$ on, both users use the following decision rule g with initialization idle to share the channel between them.

$$g(\text{idle}) = g(\text{busy}) = 1, g(\text{success}) = 0, g(\text{failure}) = 1/2.$$

In slot $t+2$, they collide, and after a collision, they transmit with probability 1/2. Once one of the two users succeeds, they alternate between $(a_i, a_j) = (T, W)$ and $(a_i, a_j) = (W, T)$ until one of the users completes the transmission of its critical traffic. After one of the users completes the transmission of its critical traffic, an idle slot occurs, and the situation becomes the same as the one where a critical event arrives following an idle slot. We can decrease the delay by requiring the user that completed the transmission of its critical traffic earlier than the other to wait in the slot following the idle slot.

Suppose now that two critical events occur simultaneously. Two critical users, without knowing the existence of another critical user, transmit with probability 1. After experiencing $B+1$ consecutive collisions, they realize that another critical user exists because the maximum number of consecutive collisions is bounded by B when there is no or only one critical user given the enhancement discussed in Section V-B. Then after $B+1$ consecutive collisions, the two critical users switch to the decision rule g as above in order to share the channel between them.

Lastly, suppose that a second critical event occurs to user j in slot t while a critical user i is experiencing collisions in a critical phase. User i realizes the existence of another critical user after $B+1$ consecutive collisions, but at that point user j has experienced less than $B+1$ consecutive collisions. After experiencing $B+1$ consecutive collisions, user i switches to the decision rule g while user j still transmits with probability 1. From that point, there are only two possible transmission outcomes. Either user j succeeds, or users i and j collide. If user j experiences $B+1$ consecutive collisions before obtaining a success, it switches to g and the two users can share the channel from that point on. Suppose that user j obtains a success before experiencing $B+1$ consecutive collisions. Then it must be followed by a collision because user i uses g. Recognizing that the pattern $(\text{success}, \text{failure})$ cannot occur when all other users have normal traffic, user j also learns the existence of another critical user and switches to g.

To summarize, when two critical users coexist, they can infer the existence of another critical user within a finite number of slots by using patterns that can be realized only when there are two critical users. Once the inference is made, they switch to another mode of operation that enables them to share the channel equally.

VI. Simulation Results

We have run simulations in order to confirm the results obtained in Sections III and IV as well as the improvements from using an enhanced adaptive protocol introduced in Section V. We consider three values of N, $N = 3$, 10, 50, and three values of θ, $\theta = 0.1$, 0.2, 0.5. For each considered pair of N and θ, we have simulated 1000 rounds of a normal phase, which starts with an idle slot and lasts for 100 slots, followed by a critical phase (assuming only one critical user) while choosing (q, r) as the optimal protocol with a nonbinding delay constraint, (q^*, r^*). Table I summarizes the simulation results, showing the values of variables T_s, T_c, C_{norm}, D_{crit} averaged over 1000 rounds as well as the maximum value of D_{crit} among the values in 1000 rounds. The results show that the simulation results match closely the results from analysis in the case of adaptive protocols and that enhanced adaptive protocols achieve a smaller delay in both average and maximum senses without degrading the performance in a normal phase. For the considered values of N and θ, D_{crit} ranges from 0.68 to 1.02 in the case of enhanced adaptive protocols and from 0.93 to 1.66 in the case of adaptive protocols. Hence, even without imposing a delay constraint, we can achieve a reasonably small delay in a critical phase by using a protocol proposed in this paper. Also, T_c ranges from 2.19 to 2.55, which shows that contention among normal users is resolved effectively by a proposed protocol.

VII. Conclusion

We have explored the possibility of achieving coordination in a network with dynamically changing user types by using adaptive MAC protocols with memory. The general theme of this research agenda is to investigate the extent to which memory containing local information can substitute explicit message passing in achieving coordination. In this paper, we are able to obtain a satisfactory performance with protocols utilizing short memory because we have focused on a relatively simple setting where there are only two types and there can be at most one or two critical users. In a more complex setting

[6] This will be especially the case if the average length of critical traffic L is not too large compared to the average duration of a success period T_s.

TABLE I
SUMMARY OF SIMULATION RESULTS (AP: ADAPTIVE PROTOCOLS, EAP: ENHANCED ADAPTIVE PROTOCOLS WITH $B = 5$)

			T_s	T_c	C_{norm}	D_{crit}	$\max D_{crit}$
$N = 3$ $(q^* = 0.3397,$ $r^* = 0.4896)$	$\theta = 0.1$	AP (analysis)	10	2.1959	0.8199	1.1786	
		AP (simulation)	10.1727	2.1865	0.8146	1.1820	11
		EAP (simulation)	10.1761	2.1879	0.8145	0.6820	5
	$\theta = 0.2$	AP (analysis)	5	2.1959	0.6948	1.0899	
		AP (simulation)	4.9567	2.2122	0.6872	1.1340	11
		EAP (simulation)	4.9544	2.2133	0.6870	0.7330	5
	$\theta = 0.5$	AP (analysis)	2	2.1959	0.4767	0.9352	
		AP (simulation)	2.0105	2.1853	0.4788	0.9340	9
		EAP (simulation)	2.0106	2.1876	0.4786	0.7660	5
$N = 10$ $(q^* = 0.1051,$ $r^* = 0.4786)$	$\theta = 0.1$	AP (analysis)	10	2.4374	0.8040	1.5297	
		AP (simulation)	9.9352	2.4316	0.7953	1.4480	8
		EAP (simulation)	9.9365	2.4358	0.7952	0.9180	5
	$\theta = 0.2$	AP (analysis)	5	2.4374	0.6723	1.3978	
		AP (simulation)	5.0662	2.4562	0.6696	1.3910	11
		EAP (simulation)	5.0657	2.4585	0.6692	0.9380	5
	$\theta = 0.5$	AP (analysis)	2	2.4374	0.4507	1.1759	
		AP (simulation)	1.9926	2.4350	0.4501	1.1180	8
		EAP (simulation)	1.9928	2.4367	0.4499	0.8930	5
$N = 50$ $(q^* = 0.0213,$ $r^* = 0.4754)$	$\theta = 0.1$	AP (analysis)	10	2.5138	0.7991	1.6468	
		AP (simulation)	9.8745	2.5389	0.7872	1.6570	10
		EAP (simulation)	9.8807	2.5452	0.7870	1.0230	5
	$\theta = 0.2$	AP (analysis)	5	2.5138	0.6654	1.4995	
		AP (simulation)	5.0067	2.4998	0.6615	1.5090	10
		EAP (simulation)	5.0114	2.5011	0.6616	1.0050	5
	$\theta = 0.5$	AP (analysis)	2	2.5138	0.4431	1.2546	
		AP (simulation)	1.9945	2.5180	0.4418	1.2470	9
		EAP (simulation)	1.9950	2.5193	0.4417	1.0060	5

where there are more than two types or more possible distributions of types (e.g., three or more critical users can coexist), achieving coordination by using only local information will become more difficult and, if possible, require longer memory. In some complicated scenarios, using explicit message passing will be more beneficial than using memory. It is an interesting future research topic to investigate the performance of adaptive protocols with memory in a general setting of dynamically changing user types and to build a framework in which the benefit and cost of using memory can be compared with those of using explicit message passing.

REFERENCES

[1] D. Bertsekas and R. Gallager, *Data Networks*, 2nd ed. Saddle River, NJ: Prentice-Hall, 1992.

[2] J. Park and M. van der Schaar, “Medium access control protocols with memory,” *IEEE/ACM Trans. Netw.*, vol. 18, no. 6, pp. 1921–1934, Dec. 2010.

[3] R. Iyer and L. Kleinrock, “QoS control for sensor networks,” in *Proc. ICC*, 2003, pp. 517–521.

[4] A. Galstyan, B. Krishnamachari, and K. Lerman, “Resource allocation and emergent coordination in wireless sensor networks,” in *AAAI Workshop on Sensor Networks*, 2004.

[5] *Part 11: Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specifications*, IEEE 802.11, 1999.

[6] L. G. Roberts, “Aloha packet system with and without slots and capture,” *ACM SIGCOMM Comput. Commun. Rev.*, vol. 5, no. 2, pp. 28–42, Apr. 1975.

[7] *Draft Supplement to Part 11: Wireless Medium Access Control (MAC) and physical layer (PHY) specifications: Medium Access Control (MAC) Enhancements for Quality of Service (QoS)*, IEEE 802.11e/D5.0, 2003.

[8] D. Gu and J. Zhang, “QoS enhancement in IEEE 802.11 wireless local area network,” *IEEE Commun. Mag.*, vol. 41, no. 6, pp. 120–124, Jun. 2003.

[9] D. Qiao and K. G. Shin, “Achieving efficient channel utilization and weighted fairness for data communications in IEEE 802.11 WLAN under the DCF,” in *Proc. IWQoS*, 2002, pp. 227–236.

[10] A. B. MacKenzie and S. B. Wicker, “Stability of multipacket slotted Aloha with selfish users and perfect information,” in *Proc. INFOCOM*, 2003, pp. 1583–1590.

[11] A. H. Mohsenian-Rad, J. Huang, M. Chiang, and V. W. S. Wong, “Utility-optimal random access without message passing,” *IEEE Trans. Wireless Commun.*, vol. 8, no. 3, pp. 1073–1079, Mar. 2009.

[12] R. T. Ma, V. Misra, and D. Rubenstein, “An analysis of generalized slotted-Aloha protocols,” *IEEE/ACM Trans. Netw.*, vol. 17, no. 3, pp. 936–949, Jun. 2009.

[13] C. M. Grinstead and J. L. Snell*, Introduction to Probability*. Providence, RI: AMS, 1997.

[14] G. Bianchi and I. Tinnirello, “Kalman filter estimation of the number of competing terminals in an IEEE 802.11 network,” in *Proc. IEEE INFOCOM*, 2003, pp. 844–852.

Jaeok Park received the B.A. degree in economics from Yonsei University, Seoul, Korea, in 2003 and the M.A. and Ph.D. degrees in economics from the University of California, Los Angeles, in 2005 and 2009, respectively.

He is currently a Postdoctoral Scholar in the Electrical Engineering Department at the University of California, Los Angeles. From 2006 to 2008, he served in the Republic of Korea Army. His primary research interests include game theory, mechanism design, network economics, and wireless communication.

Mihaela van der Schaar (F’10) is a Professor in the Electrical Engineering Department at University of California, Los Angeles. Her research interests are in multimedia networking and communication, multimedia systems, multiuser communication networks, online learning, network economics and game theory.

Prof. van der Schaar received in 2004 an NSF CAREER Award, in 2005 the Best Paper Award from IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS FOR VIDEO TECHNOLOGY, in 2006 the Okawa Foundation Award, in 2005, 2007, and 2008 the IBM Faculty Award, and in 2006 the Most Cited Paper Award from *EURASIP: Image Communications* journal. She was an Associate Editor for IEEE TRANSACTIONS ON MULTIMEDIA, SIGNAL PROCESSING LETTERS, CIRCUITS AND SYSTEMS FOR VIDEO TECHNOLOGY, *Signal Processing Magazine*, etc. She also holds 33 granted U.S. patents and three ISO awards for her contributions to the MPEG video compression and streaming international standardization activities. Starting January 2011, she is the Editor-in-Chief of the IEEE TRANSACTIONS ON MULTIMEDIA.