Cognitive Multiple Access Via Cooperation: Protocol Design and Performance Analysis

Ahmed K. Sadek, *Member, IEEE*, K. J. Ray Liu, *Fellow, IEEE*, and Anthony Ephremides, *Fellow, IEEE*

Abstract—In this paper, a novel cognitive multiple-access strategy in the presence of a cooperating relay is proposed. Exploiting an important phenomenon in wireless networks, source burstiness, the cognitive relay utilizes the periods of silence of the terminals to enable cooperation. Therefore, no extra channel resources are allocated for cooperation and the system encounters no bandwidth losses. Two protocols are developed to implement the proposed multiple-access strategy. The maximum stable throughput region and the delay performance of the proposed protocols are characterized. The results reveal that the proposed protocols provide significant performance gains over conventional relaying strategies such as selection and incremental relaying, specially at high spectral efficiency regimes. The rationale is that the lossless bandwidth property of the proposed protocols results in a graceful degradation in the maximum stable throughput with increasing the required rate of communication. On the other hand, conventional relaying strategies suffer from catastrophic performance degradation because of their inherent bandwidth inefficiency that results from allocating specific channel resources for cooperation at the relay. The analysis reveals that the throughput region of the proposed strategy is a subset of its maximum stable throughput region, which is different from random access, where both regions are conjectured to be identical.

Index Terms—Cooperative communications, delay analysis, diversity techniques, multiple access, queueing theory, relay channel, stability analysis, stability region.

I. INTRODUCTION

COOPERATIVE diversity is a recently emerged technique for wireless communications that has gained wide attention [1]. As a new communication paradigm for wireless networks, the impact of cooperation at different network layers needs to be studied and understood. Most of the work on cooperative communications has focused on the physical layer aspects of the problem. Cooperation protocols for the single-relay case were developed and analyzed in [1], and later in [2] and [4] distributed space–time codes were developed and analyzed for multiple-relay scenarios. In the previous works it was demonstrated that full diversity gain can be achieved by forming such

virtual multiple-input multiple-output (MIMO) systems. The diversity–multiplexing tradeoff of cooperative diversity has been also explored in [5] using an information-theoretic framework. Achievable capacity and coding strategies for wireless relay channels have been considered recently in [3].

Beside characterizing cooperation gains based upon an information-theoretic framework, as in the previously cited works, the symbol-error-rate performance for cooperative communications was analyzed in [6], [7], for a single-relay scenario, and later in [9] for the multiple-relay scenario. The impact of cooperation on extending coverage area in wireless networks has been investigated in [10] and [11] in which distributed and centralized relay assignment protocols have been developed, respectively.

Few works started to study the impact of cooperation at higher network layers. Cooperation in random-access networks has been considered in [12]–[14]. In [12], the authors proposed a distributed version of network diversity multiple-access (NDMA) [31] protocol and they provided pairwise error probability analysis to demonstrate the diversity gain. In [13] and [14], the authors presented the notion of utilizing the spatial separation between users in the network to assign cooperating pairs (also groups) to each other. In [14], spread-spectrum random-access protocols were considered in which nearby inactive users are utilized to gain diversity advantage via cooperation. Also [14] provided throughput analysis for a symmetrical setup where all terminals are statistically identical. The main focus of these works was still the achievable diversity gains of cooperative communications.

Despite the demonstrated promised gains of cooperative communications, the impact of cooperation on higher network levels is not completely understood yet. Most of the previous work on cooperation assumes the user has always a packet to transmit which is not generally true in a wireless network. For example, most of the sources in a network are bursty in nature, which results in periods of silence in which the users have no data to transmit. Such a phenomenon may affect important system parameters that are relevant to higher network layers, for example, buffer stability and packet delivery delay. In this work, we focus on the multiple-access layer trying to address some important questions, such as, can we design cooperation protocols taking higher layer network features into account? Can the gains promised by cooperation at the physical layer be leveraged to the multiple-access layer? More specifically, what is the impact of cooperation on important multiple-access performance metrics such as stable throughput region and packet delivery delay?

In this work, we try to address these important questions to demonstrate the possible gains of cooperation at the multiple-access layer. We consider a slotted time-division multiple-ac-

cess (TDMA) framework in which each time slot is assigned only to one terminal, i.e., orthogonal multiple access. If a terminal does not have a packet to transmit in its time slot, then this time slot is not utilized. These unutilized time slots are wasted channel resources that could be used to enhance the system performance. Recently, the concept of cognitive radio has been introduced to allow the utilization of unused channel resources by enabling the operation of a secondary system overlapping with the original system (see [33] and references therein). In our work, we propose a novel cognitive multiple-access strategy with the concept of cooperation. In the proposed protocol, the cognitive relay tries to "*smartly*" utilize the periods of source silence to cooperate with other terminals in the network in order to increase the reliability of communications against random channel fades. In particular, when the relay senses the channel for empty time slots, the slots are then used to help other terminals in the network by forwarding their packets lost in some previous transmissions. Thus, this new protocol has cooperative cognitive aspects in the sense that unused channel resources are being utilized by the relay to cooperate with other terminals in the network. It should be pointed out that the proposed cooperative protocol does not result in any bandwidth loss because there are no channel resources reserved for the relay to cooperate. We demonstrate later that this important feature of the proposed protocol results in significant gains especially in the high spectral efficiency regimes.

We develop two protocols to implement this new cooperative cognitive multiple-access (CCMA) strategy. The first protocol is CCMA within a single frame (CCMA-S), where the relay keeps a lost packet no more than one time frame and then drops the packet if it was not able to deliver it successfully to the destination. The dropped packet then has to be retransmitted by the originating terminal. It turns out that in this protocol the relay's queue is always bounded, and that the terminals queues are interacting. To analyze the stability of the system's queues we resort to a stochastic dominance approach. Analyzing the stability of interacting queues is a difficult problem that has been addressed for ALOHA systems initially in [20], and later in [21], where the dominant system approach was explicitly introduced. Many other works followed that to study the stability of ALOHA [15], [22], [27], and [28]. Characterizing the stable throughput region for interacting queues with $M > 3$ terminals is still an open problem. In CCMA-S, the interaction among the queues arises due to the role of the relay in enabling cooperation, which is different from the intrinsic cause of interaction in the ALOHA system. To analyze the stability of CCMA-S, we introduce a new dominant system to resolve this interaction, and we characterize the stability region for the two-user case, and the maximum stable throughput for the M-user symmetric case.

The second protocol that we propose, named CCMA-Multiple-enhanced (CCMA-Me), differs in the way the relay handles the lost packets. In CCMA-Me, if a packet is captured by the relay and not by the destination, then this packet is removed from the corresponding terminal's queue and it becomes the relay's responsibility to deliver it to the destination. The term enhanced refers to the design of the protocol, as the relay only helps the terminals with inferior channel gains reflected in the
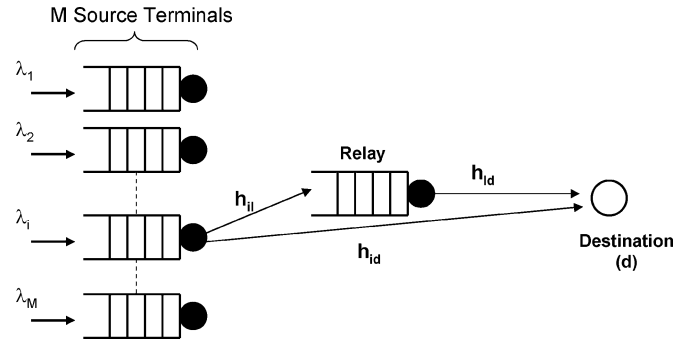

Fig. 1.   Network and channel model.

distances from the terminals to the destination. Different from CCMA-S, the size of the relay's queue can possibly be unbounded and so its stability must be studied.

In addition to characterizing the stable throughput region of our proposed protocols, we also analyze the queueing delay performance. Delay is an important performance measure and network parameter that may affect the tradeoff between rate and reliability of communication. Delay analysis for interacting queues is a notoriously hard problem that has been investigated in [25] and [26] for ALOHA. We consider a symmetric two-user scenario when analyzing the delay performance of the proposed protocols.

Our theoretical results demonstrate significant performance gains of the proposed protocols over TDMA without relaying. We also consider ALOHA, selection decode-and-forward [1], and incremental decode-and-froward [1] in the comparison. The analytical and numerical results show that CCMA-Me provides significantly better performance over these protocols. The rationale behind these gains is that in CCMA-Me the system does not incur any bandwidth loss as cooperation takes place only in the idle time slots, whereas in conventional cooperation protocols, designed mainly from a physical layer perspective, some channel resources are dedicated for the relay to enable cooperation. The resulting loss in the bandwidth efficiency causes large degradation in the system performance compared to our proposed protocols especially for high spectral efficiency regimes. The rest of the paper is organized as follows. In the next section, we describe the system model considered in this paper. In Section III, we present our new CCMA protocols. The stable throughput region of the proposed protocols is characterized in Section IV, and the saturated throughput is considered in Section V. Delay performance is analyzed in Section VI, and finally, conclusions are drawn in Section VII.

II. SYSTEM MODEL

We consider the uplink of a TDMA system. Our network consists of a finite number $M < \infty$ of source terminals numbered $1, 2, \ldots, M$, a relay node l,[1] and a destination node d, see Fig. 1. Let $\mathcal{T} = \{\mathcal{M}, l\}$ denote the set of transmitting nodes, where $\mathcal{M} = \{1, 2, \ldots, M\}$ is the set of source terminals, and $\mathcal{D} = \{l, d\}$ denotes the set of receiving nodes or possible destinations. For simplicity of presentation, in the following we use terminal to refer to a source terminal.

---

[1]We use l to denote the relay not to confuse with r that denotes distance

First, we describe the queueing model for the multiple-access channel. Each of the M terminals and the relay l has an infinite buffer for storing fixed-length packets. The channel is slotted, and a slot duration is equal to a packet duration. The arrival process at any terminal's queue is independent and identically distributed (i.i.d.) from one slot to another, and the arrival processes are independent from one terminal to another. Let λ_i denote the arrival process at the ith queue ($i \in \{1, 2, \ldots, M\}$). Terminals access the channel by dividing the channel resources, time in this case, among them, hence, each terminal is allocated a fraction of the time. Let $\Omega = [\omega_1, \omega_2, \ldots, \omega_M]$ denote a resource-sharing vector, where $\omega_i \geq 0$ is the fraction of the time allocated to terminal $i \in \mathcal{M}$, or it can represent the probability that terminal i is allocated the whole time slot [30]. In this work, we use the later notation that allows us to consider fixed duration time slots with continuous values of the resource-sharing vector. A time frame is defined as M consecutive time slots. The set of all feasible resource-sharing vectors is specified as follows:

$$F \triangleq \left\{ \Omega = [\omega_1, \omega_2, \ldots, \omega_M] \in \mathrm{R}_+^M : \sum_{i \in \mathcal{M}} \omega_i \leq 1 \right\}. \quad (1)$$

A fundamental performance measure of a communication network is the stability of its queues. Stability can be loosely defined as having a certain quantity of interest kept finite. In our case, we are interested in the queue size and the packet delivery delay to be finite. More rigorously, stability can be defined as follows. Denote the queue sizes of the transmitting nodes at any time t by the vector $\boldsymbol{Q}^t = [Q_i^t, i \in \mathcal{T}]$. We adopt the following definition of stability used in [15].

Definition 1: Queue $i \in \mathcal{T}$ of the system is stable, if

$$\lim_{t \to \infty} \mathrm{Pr}\left[Q_i^t < x\right] = F(x) \quad \text{and} \quad \lim_{x \to \infty} F(x) = 1. \quad (2)$$

If $\lim_{x \to \infty} \lim_{t \to \infty} \inf \mathrm{Pr}[Q_i^t < x] = 1$, the queue is called substable.

If the arrival and service processes of a queueing system are strictly stationary, then one can apply Loynes's theorem to check for stability conditions [16]. This theorem states that if the arrival process and the service process of a queueing system are strictly stationary, and the average arrival rate is less than the average service rate, then the queue is stable; if the average arrival rate is greater than the average service rate then the queue is unstable.

Next, we describe the physical channel model. The wireless channel between any two nodes in the network is modeled as a Rayleigh narrowband flat-fading channel with additive Gaussian noise. The transmitted signal also suffers from propagation path loss that causes the signal power to attenuate with distance. The signal received at a receiving node $j \in \mathcal{D}$ from a transmitting node $i \in \mathcal{T}$ at time t can be modeled as

$$y_{ij}^t = \sqrt{G r_{ij}^{-\gamma}} h_{ij}^t x_i^t + n_{ij}^t, \qquad i \in \mathcal{T}, j \in \mathcal{D}, i \neq j \quad (3)$$

where G is the transmitting power, assumed to be the same for all transmitting terminals, r_{ij} denotes the distance between the two nodes i, j, γ is the path loss exponent, and h_{ij}^t captures the channel fading coefficient at time t and is modeled as i.i.d. zero-mean, circularly symmetric complex Gaussian random process

with unit variance. The term x_i^t denotes the transmitted packet with average unit power at time t, and n_{ij}^t denotes i.i.d. additive white Gaussian noise with zero mean and variance N_o. Since the arrival, the channel gains, and the additive noise processes are assumed stationary, we can drop the index t without loss of generality. We consider the scenario in which the fading coefficients are known to the appropriate receivers, but are not known at the transmitters.

In this paper, we characterize the success and failure of packet reception by outage events and outage probability, which is defined as follows. For a target signal-to-noise (SNR) ratio β, if the received SNR as a function of the fading realization h is given by $\mathrm{SNR}(h)$, then the outage event O is the event that $\mathrm{SNR}(h) < \beta$, and $\mathrm{Pr}[\mathrm{SNR}(h) < \beta]$ denotes the outage probability. This definition is equivalent to the *capture model* in [19], [18]. The SNR threshold β is a function of different parameters in the communication system; it is a function of the application, the data rate, the signal processing applied at encoder/decoder sides, error-correction codes, and other factors.

For the channel model in (3), the received SNR of a signal transmitted between two terminals i and j can be specified as follows:

$$\mathrm{SNR}_{ij} = \frac{\mid h_{ij} \mid^2 r_{ij}^{-\gamma} G}{N_o} \quad (4)$$

where $\mid h_{ij} \mid^2$ is the magnitude channel gain square and has an exponential distribution with unit mean. The outage event for an SNR threshold β is equivalent to

$$O_{ij} = \{h_{i,j} : \mathrm{SNR}_{ij} < \beta\} = \{h_{i,j} : \mid h_{ij} \mid^2 < \frac{\beta N_o r_{i,j}^{\gamma}}{G}\}. \quad (5)$$

Accordingly, the probability of outage is given by

$$\mathrm{Pr}[O_{ij}] = \mathrm{Pr}\left[\mid h_{ij} \mid^2 < \frac{\beta N_o r_{i,j}^{\gamma}}{G}\right] = 1 - \exp\left(-\frac{\beta N_o r_{i,j}^{\gamma}}{G}\right) \quad (6)$$

where the above follows from the exponential distribution of the received SNR. Since we will use the above expression frequently in our subsequent analysis, and for compactness of representation, we shall use the following convention to denote the success probability (no outage) at SNR threshold β:

$$f_{ij} = \exp\left(-\frac{\beta N_o r_{i,j}^{\gamma}}{G}\right). \quad (7)$$

III. COOPERATIVE COGNITIVE MULTIPLE-ACCESS (CCMA) PROTOCOLS

In a TDMA system without relays, if a terminal does not have a packet to transmit, its time slot remains idle, i.e., wasted channel resources. In this work, we investigate the possibility of utilizing these wasted channel resources by employing a relay. In this section, we introduce our proposed cognitive multiple-access strategy based on employing relays in the wireless network. Furthermore, we develop two protocols to implement this new approach. We assume that the relay can sense the communication channel to detect empty time slots. This assumption is reasonable for the orthogonal multiple-access scheme used,

as there is no interference, and the relay can employ coherent or feature detectors that have high detection probability. In the presence of interference, knowledge of the interference structure can help in the detection, however, this is out of the scope of the paper. The second assumption we make is that the errors and delay in packet acknowledgement feedback is negligible, which is reasonable for short length ACK/NACK packets as low rate codes can be employed in the feedback channel.

First, we describe the new multiple-access strategy. Due to the broadcast nature of the wireless medium, the relay can listen to the packets transmitted by the terminals to the destination. If the packet is not received correctly by the destination, the relay stores this packet in its queue, given that it was able to decode this packet correctly. Thus, the relay's queue contains packets that have not been transmitted successfully by the terminals. At the beginning of each time slot, the relay listens to the channel to check whether the time slot is empty (not utilized for packet transmission) or not. If the time slot is empty, the relay will retransmit the packet at the head of its queue, hence utilizing this channel resource that was previously wasted in a TDMA system without a relay. Moreover, this introduces spatial diversity in the network as the channel fades between different nodes in the network are independent. In the following, we develop two different protocols to implement the proposed cognitive multiple-access approach . The proposed protocol is cognitive in the sense that it introduces a relay in the network that tries detecting unutilized channel resources and use them to help other terminals by forwarding packets lost in previous transmissions.

A. CCMA-Single Frame (CCMA-S)

The first protocol that we propose is CCMA within a single frame duration or (CCMA-S). The characteristic feature of CCMA-S is that any terminal keeps its lost packet in its queue until it is captured successfully at the destination. CCMA-S operates according to the following rules.

- Each terminal transmits the packet at the head of its queue in its assigned slot, if the terminal's queue is empty the slot is free.
- If the destination receives a packet successfully, it sends an ACK which can be heard by both the terminal and the relay. If the destination does not succeed in receiving the packet correctly but the relay does, then the relay stores this packet at the end of its queue. The corresponding terminal still keeps the lost packet at the head of its queue
- The relay senses the channel, and at each empty time slot the relay transmits the packet at the head of its queue, if its queue is nonempty. If the transmitted packet is received correctly by the destination it sends an ACK and the corresponding terminal removes this packet from its queue.
- If the relay does not succeed in delivering the packet in the following $M - 1$ time slots, then the relay drops this packet from its queue. In this case, the corresponding terminal becomes responsible for delivering the packet to the destination.

Following are some important remarks on the above protocol. According to the above description of CCMA-S, the relay's queue has always a finite number of packets (at most has M backlogged packets). This follows because according to the protocol, the relay can have at most one packet from each terminal. Thus, the stability of the system is only determined by the stability of the terminals' queues. Second, successful service of a packet in a frame depends on whether the other terminals have idle time slots or not. Therefore, individual terminals' queues are *interacting*.

B. CCMA-Multiple Frames (CCMA-M)

The main difference between protocols CCMA-S and CCMA-M is in the role of the relay and the behavior of the terminals' regarding their backlogged packets. More specifically, a terminal removes a packet from its queue if it is received successfully by either the destination or the relay. CCMA-M operates according to the following rules.

- Each terminal transmits the packet at the head of its queue in its assigned time slot. If the queue is empty the time slot is free.
- If a packet is received successfully by either the destination or the relay, the packet is removed from the terminal's queue (the relay needs to send an ACK if one is not heard by the destination in this case).
- If a packet is not received successfully by both the relay and the destination, the corresponding terminal retransmits this packet in its next assigned time slot.
- At each sensed empty time slot, the relay retransmits the packet at the head of its queue.

One can now point out the differences between the queues in systems CCMA-S and CCMA-M: i) The size of the relay's queue can possibly grow in CCMA-M as it can have more than one packet from each terminal, however, it can not exceed size M in CCMA-S. ii) The terminal's queues in CCMA-M are not interacting as in CCMA-S. This is because the terminal removes the packets which were received correctly by the relay or the destination from its queue. In other words, servicing the queue of any terminal depends only on the channel conditions from that terminal to the destination and relay, and does not depend on the status of the other terminals' queues. iii) The stable throughput region of CCMA-M requires studying the stability of both the terminals' queues and the relay's queue.

IV. STABILITY ANALYSIS

The aim of this section is to characterize the stable throughput region of the proposed cooperation protocols.

A. Stability Analysis of CCMA-S

1) The Two-Terminal Case: In CCMA-S, we observed in the previous section that the relay's queue size is always finite, hence it is always stable. For the two-terminal case, the queues evolve as a two-dimensional Markov chain in the first quadrant. From the protocol description in the previous section, one can observe that the system of queues in CCMA-S is interacting. In other words, the transition probabilities differ according to whether the queues are empty or not. For example, if one of the two terminals' queues was empty for a long time, then the relay serves the lost packets from the other terminal more often. On the other hand, if one of the two terminals queues never empties, then the other terminal will never get served by the

relay. Studying stability conditions for interacting queues is a difficult problem that has been addressed for ALOHA systems [20], [21], and [22].

To analyze the stability of CCMA-S, we develop a dominant system to decouple the interaction of the terminals originating from the role of the relay in cooperation.. Using our developed dominant system we are able to characterize the stability region of two-user CCMA-S for a fixed resource sharing vector, and hence the whole stability region. The following lemma states the stability region of CCMA-S for a fixed resource sharing vector $[\omega_1, \omega_2]$.

Lemma 1: The stability region of two-user CCMA-S for a fixed resource sharing vector $[\omega_1, \omega_2]$ is given by $\mathcal{R}(S_1) \bigcup \mathcal{R}(S_2)$ where

$$\mathcal{R}(S_1) = \{[\lambda_1, \lambda_2] \in \mathrm{R}_+^2 : \lambda_2$$
$$< h(\lambda_1; w_1, w_2, f_{1d}, f_{2d}, f_{2l}, f_{ld}), \text{ for } \lambda_1 < \omega_1 f_{1d}.\} \quad (8)$$

and

$$\mathcal{R}(S_2) = \{[\lambda_1, \lambda_2] \in \mathrm{R}_+^2 : \lambda_1$$
$$< h(\lambda_2; w_2, w_1, f_{2d}, f_{1d}, f_{1l}, f_{ld}), \text{ for } \lambda_2 < \omega_2 f_{2d}.\} \quad (9)$$

where

$$h(x; \alpha_1, \alpha_2, \alpha_3, \alpha_4, \alpha_5, \alpha_6)$$
$$= \alpha_2 \alpha_4 + \alpha_1 \alpha_2 \left(1 - \frac{x}{\alpha_1 \alpha_3}\right)(1 - \alpha_3)\alpha_5 \alpha_6.$$

Proof of Lemma 1: The proof depends on constructing a dominant system that decouples the interaction between the queues and thus renders the analysis tractable. By dominance, we mean that the queues in the dominant system stochastically dominate the queues in the original CCMA-S system, i.e., with the same initial conditions for queue sizes in both the original and dominant systems, the queue sizes in the dominant system are not smaller than those in the original system [21].

We define the dominant system for CCMA-S as follows. For $j \in \{1, 2\}$, define S_j as

- S_j and CCMA-S are identical, except that, the packets successfully transmitted by the relay for user j are not removed from user's j queue in S_j.

The above definition of the dominant system implies that queue j evolves exactly as in a TDMA system without a relay. If both the dominant system and the original CCMA-S started with the same initial queue sizes, then the queues in S_j are always not shorter than those in CCMA-S. This follows because a packet successfully transmitted for queue j in S_j is always successfully transmitted from the corresponding queue in CCMA-S. However, the relay can succeed in forwarding some packets from queue j in CCMA-S in the empty time slots of the other terminal. This implies that queue j empties more frequently in CCMA-S and therefore the other terminal is better served in CCMA-S compared to S_j. Consequently, stability conditions for the dominant system S_j ($j \in \{1, 2\}$) are sufficient for the stability of the original CCMA-S system. In the following, we first derive the sufficient conditions for stability of CCMA-S.

Consider system S_1 in which the relay only helps terminal 2 and terminal 1 acts exactly as in a TDMA system. In order to apply Loynes' theorem, we require the arrival and service processes for each queue to be stationary. The queue size for

terminal $i \in \{1, 2\}$ in system S_1 at time t, denoted by $Q_i^t(S_1)$, evolves as follows

$$Q_i^{t+1}(S_1) = \left(Q_i^t(S_1) - Y_i^t(S_1)\right)^+ + X_i^t(S_1), \quad (10)$$

where $X_i^t(S_1)$ represents the number of arrivals in slot t and is a stationary process by assumption with finite mean $E[X_i^t(S_1)] = \lambda_i$. The function $()^+$ is defined as $(x)^+ = \max(x, 0)$. $Y_i^t(S_1)$ denotes the possible (*virtual*) departures from queue i at time t; by virtual we mean that $Y_i^t(S_1)$ can be equal to 1 even if $Q_i^t(S_1 = 0)$. We assume that departures occur before arrivals, and the queue size is measured at the beginning of the slot [21]. For terminal $i = 1$, the service process can be modeled as

$$Y_1^t(S_1) = \mathbf{1}\left[A_1^t \cap \overline{O_{1,d}^t}\right], \quad (11)$$

where $\mathbf{1}[\cdot]$ is the indicator function, A_1^t denotes the event that slot t is assigned to terminal 1, and $\overline{O_{1,d}^t}$ denotes the complement of the outage event between terminal 1 and the destination d at time t.[2] Due to the stationarity assumption of the channel gain process $\{h_{i,d}^t\}$, and using the outage expression in (6), the probability of this event is given by $\Pr\left[\overline{O_{1,d}^t}\right] = f_{1d}$. From the above, it is clear that the service process $Y_1^t(S_1)$ is stationary and has a finite mean given by $E[Y_1^t(S_1)] = \omega_1 f_{1d}$, where $E[\cdot]$ denotes statistical expectation. According to Loynes, stability of queue 1 in the dominant system S_1 is achieved if the following condition holds

$$\lambda_1 < \omega_1 f_{1d}. \quad (12)$$

Consider now queue 2 in system S_1. The difference between the evolution of this queue and queue 1 is in the definition of the service process $Y_2^t(S_1)$. A packet from queue 2 can be served in a time slot in either one of the two following events: 1) If the time slot belongs to queue 2 and the associated channel $h_{2,d}^t$ is not in outage; or 2) the time slot belongs to queue 1, queue 1 is empty, in the previous time slot the channel to relay from terminal 2 was not in outage, and the relay-destination channel is not in outage. This can be modeled as

$$Y_2^t(S_1) = \mathbf{1}\left[A_2^t \bigcap \overline{O_{2,d}^t}\right]$$
$$+ \mathbf{1}\left[A_1^t \bigcap \{Q_1^t(S_1) = 0\} \bigcap A_2^{t-1} \bigcap \overline{O_{2,l}^{t-1}} \bigcap O_{2,d}^{t-1} \bigcap \overline{O_{l,d}^t}\right] \quad (13)$$

where $\{Q_1^t(S_1) = 0\}$ denotes the event that terminal's 1 queue is empty in time slot t. The two indicator functions in the right hand side of (13) are mutually exclusive, hence, the average rate of the service process is given by (14) at the bottom of the next page, where $\Pr[O_{ij}]$ is the probability of outage between nodes i and j. Under (12), the probability of queue 1 empty is given by

$$\Pr\left[\{Q_1^t(S_1^1) = 0\}\right] = 1 - \frac{\lambda_1}{\omega_1 f_{1d}}. \quad (15)$$

Using the expression of the outage probability in (6) and Loynes conditions for stability [16], and given (12), the stability condition for queue 2 in the dominant system S_1 is given by

$$\lambda_2 < \omega_2 f_{2d} + \omega_1 \omega_2 \left(1 - \frac{\lambda_1}{\omega_1 f_{1d}}\right)(1 - f_{2d}) f_{2l} f_{ld}. \quad (16)$$

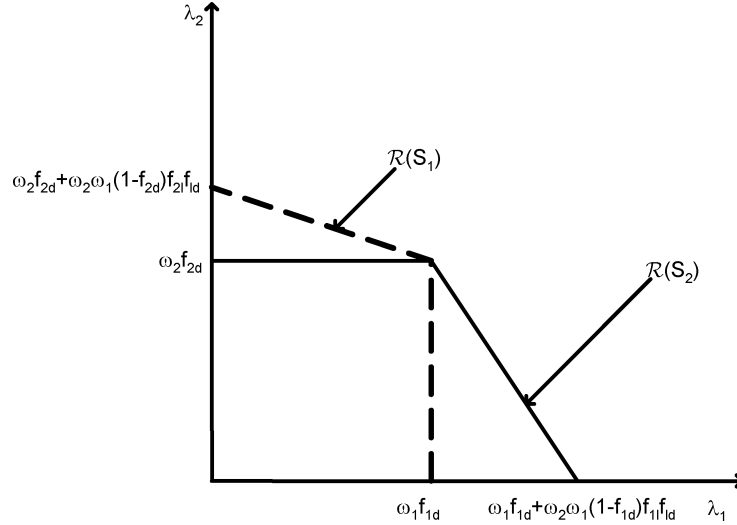[2]$\overline{(\cdot)}$ denotes the complement of the event.

Fig. 2. Stable throughput region for system CCMA-S for a fixed resource-sharing vector (ω_1, ω_2) given by $\mathcal{R}(S_1)\bigcup\mathcal{R}(S_2)$.

Both conditions (12) and (16) represent sufficient conditions for the stability of system S_1 for a specific resource-sharing vector (ω_1, ω_2) pair. Denote the region (12) and (16) as $\mathcal{R}(S_1)$. Using parallel arguments for the dominant system S_2, we can characterize a similar region $\mathcal{R}(S_2)$ for this system by the following pair of inequalities:

$$\lambda_2 < \omega_2 f_{2d}, \quad \lambda_1 < \omega_1 f_{1d} + \omega_2 \omega_1 \left(1 - \frac{\lambda_2}{\omega_2 f_{2d}}\right)(1 - f_{1d})\, f_{1l} f_{ld}. \tag{17}$$

Since stability conditions for a dominant system is sufficient for the stability of CCMA-S, any point inside the regions $\mathcal{R}(S_1)$ and $\mathcal{R}(S_2)$ can be achieved by the original system CCMA-S, hence $\mathcal{R}(S_1)\bigcup\mathcal{R}(S_2)$ is a subset from the stability region of CCMA-S for a fixed resource-sharing pair (ω_1, ω_2). This region is depicted in Fig. 2.

Up to this point we only proved the sufficient conditions for the stability of CCMA-S in the lemma. To prove the necessary conditions, we follow a similar argument that was used by [21] and [27] for ALOHA systems to prove the *indistinguishability* of the dominant and original systems at saturation. The argument is as follows. Consider the dominant system S_1 whose stability region is characterized by the pair of inequalities (12), (16). Note that if queue 2 does not empty, packets of user 1 are never relayed in both S_1 and CCMA-S, and both systems become identical. In S_1, for $\lambda_1 < \omega_1 f_{1d}$, if

$$\lambda_2 > \omega_2 f_{2d} + \omega_1 \omega_2 \left(1 - \frac{\lambda_1}{\omega_1 f_{1d}}\right)(1 - f_{2d})\, f_{2l} f_{ld}$$

then using similar arguments as before Q_1^t is stable and Q_2^t is unstable by Loynes, i.e., $\lim_{t\to\infty} Q_2^t \to \infty$ almost surely. If Q_2^t tends to infinity almost surely, i.e., does not empty, then S_1 and CCMA-S are identical, and if both systems are started from the same initial conditions, then on a set of sample paths of positive

probability Q_2^t in CCMA-S never returns to zero for $t \geq 0$. Hence Q_2^t in CCMA-S tends to infinity with positive probability, i.e., CCMA-S is also unstable. Using parallel arguments for system S_2, if queue 1 never empties then queue 2 can be never served by the relay and both the dominant and the original systems are unstable. This means that the boundary for the stability region of the dominant system is also a boundary for the stability region of the original CCMA-S system. Thus, conditions for stability of the dominant system are sufficient and necessary for the stability of the original system. This completes the proof of Lemma 1 $\qquad\square$

The whole stability region for system CCMA-S can be determined by taking the union over all feasible resource-sharing vectors as follows:

$$\mathcal{R}(\text{CCMA-S}) = \bigcup_{\Omega \in F} \left\{ \mathcal{R}_1(S^1)\bigcup\mathcal{R}_2(S^1) \right\}. \tag{18}$$

We give a complete characterization of the stability region of CCMA-S in the following theorem.

Theorem 1: The stability region for a two-user CCMA-S system is given by

$$\mathcal{R}(\text{CCMA-S}) = \left\{ [\lambda_1, \lambda_2] \in \mathbb{R}_+^2 : \lambda_2 < \max\left[g_1(\lambda_1), g_2(\lambda_1)\right] \right\} \tag{19}$$

where the functions $g_1(\cdot)$ and $g_2(\cdot)$ are defined in (20) at the top of the following page, and the function $g_2(\cdot)$ is specified as

$$g_2(\lambda_1)$$
$$= \begin{cases} f_{2d} - \dfrac{f_{2d}}{f_{1d}}\lambda_1, & 0 \leq \lambda_1 < \dfrac{f_{1d}^2}{(1-f_{1d})f_{1l}f_{ld}} \\ \dfrac{f_{1d}f_{2d}}{K_1} + f_{2d} - 2f_{2d}\sqrt{\dfrac{\lambda_1}{K_1}}, & \dfrac{f_{1d}^2}{(1-f_{1d})f_{1l}f_{ld}} \leq \lambda_1 \leq \lambda_1^*. \end{cases} \tag{21}$$

$$E\left[Y_2^t(S_1^1)\right] = \omega_2 f_{2d} + \omega_1 \Pr\left[\{Q_1^t(S_1^1) = 0\}\right]\omega_2\left(1 - \Pr[O_{2l}]\right)\Pr[O_{2d}]\left(1 - \Pr[O_{ld}]\right) \tag{14}$$

$$g_1(\lambda_1) = \begin{cases} K_2 \left(\frac{\lambda_1 + f_{1d}}{2f_{1d}} - \frac{f_{2d}}{2K_2} \right)^2 - \frac{K_2 \lambda_1}{f_{1d}} + f_{2d}, & 0 \leq \lambda_1 \leq f_{1d} - \frac{f_{1d} f_{2d}}{K_2} \\ f_{2d} - \frac{f_{2d}}{f_{1d}} \lambda_1, & f_{1d} - \frac{f_{1d} f_{2d}}{K_2} < \lambda_1 \leq f_{1d} \end{cases} \quad (20)$$

where

$$\lambda_1^* = \begin{cases} f_{1d}, & (1 - f_{1d}) f_{1l} f_{1d} < f_{1d} \\ \frac{1}{4K_1} (f_{1d} + K_1)^2, & f_{1d} \leq (1 - f_{1d}) f_{1l} f_{1d}. \end{cases} \quad (22)$$

and $K_i = (1 - f_{id}) f_{il} f_{ld}$, $i \in \{1, 2\}$.

Proof of Theorem 1: See Appendix A. □

An interesting observation that we make from the above theorem is that both functions $g_1(\lambda_1)$ and $g_2(\lambda_1)$ are linear on some part of their domain and strictly convex on the other part. It is not obvious, however, whether both functions are strictly convex over their domain of definition, and hence, whether the boundary of the stability region of CCMA-S given by $\max\{g_1, g_2\}$ is convex or not. In the following lemma, we prove this property.

Lemma 2: The boundary of the stability region of two-user CCMA-S given by $\max\{g_1, g_2\}$ is a convex function.

Proof of Lemma 2: See Appendix B. □

Lemma 2 will prove useful in characterizing the relation among the stability regions of the different multiple-access protocols considered in this paper. The first relation that we state is that between the stability regions of TDMA and CCMA-S.

Lemma 3: The stability region of two-user TDMA is contained inside that of two-user CCMA-S. In other words

$$\mathcal{R}(\text{TDMA}) \subseteq \mathcal{R}(\text{CCMA-S}). \quad (23)$$

The two regions are identical if the following two conditions are satisfied simultaneously:

$$(1 - f_{2d}) f_{2l} f_{1d} < f_{2d}, \qquad (1 - f_{1d}) f_{1l} f_{1d} < f_{1d}. \quad (24)$$

Proof of Lemma 3: We use Theorem 1 and Lemma 2 in the proof of this lemma. The stability region of TDMA is determined according to the following parametric inequalities:

$$\lambda_1 \leq \omega_1 f_{1d}, \qquad \lambda_2 \leq \omega_2 f_{2d} \quad (25)$$

or equivalently

$$\frac{\lambda_1}{f_{1d}} + \frac{\lambda_2}{f_{2d}} \leq 1. \quad (26)$$

From Lemma 2, both functions g_1 and g_2 that determines the boundary of the stability region of CCMA-S are convex. From the proof of the convexity, we note that the straight line $\lambda_2 = f_{2d} - \frac{f_{2d}}{f_{1d}} \lambda_1$ is a tangent for both functions, hence, it lies below both functions. Since this straight line is itself the boundary for the TDMA stability region, then the stability region of TDMA is a subset of that of CCMA-S. To prove the second part of the lemma, we use the definitions of the functions g_1, g_2 in Theorem 1. From (83), observe that if $(1 - f_{1d}) f_{1l} f_{1d} < f_{1d}$

then the maximum λ_1 is determined by f_{1d}. If simultaneously $(1 - f_{2d}) f_{2l} f_{2d} < f_{2d}$, then by substituting both conditions in the domain definitions of the functions g_1, g_2, it can be seen that both functions reduce to

$$g_1(\lambda_1) = g_2(\lambda_1) = f_{2d} - \frac{f_{2d}}{f_{1d}} \lambda_1, \qquad 0 \leq \lambda_1 \leq f_{1d} \quad (27)$$

which is the boundary for the stability region of TDMA. Hence, if both conditions in (24) are satisfied, CCMA-S and TDMA have the same stable throughput regions. This completes the proof of the lemma. □

2) The Symmetric M-Terminals Case: Stability analysis for the general M-terminal case is very complicated. For the ALOHA case, only bounds on the stability region have been derived [22], [27]. In this paper, we only focus on the symmetric scenario. We define the dominant system for M-terminal CCMA-S, S_M^M, as follows. S_M^M and the original system are identical, except that, the relay does not help any of the users in S_M^M. It is clear that the queues in this system are never shorter than those in the original system CCMA-S. For S_M^M, the success probability of transmitting a packet is equal for all terminals and is given by

$$P_s(S_M^M) = \Pr[\text{SNR} \geq \beta] = f_{1d}. \quad (28)$$

The service rate per terminal is thus given by $\mu(S_M^M) = \frac{f_{1d}}{M}$, due to the symmetry of the problem. Since system S_M^M acts as a TDMA system without a relay, the queues are decoupled and hence the arrival process and departure process of each of them is strictly stationary. Applying Loynes theorem, the stability condition for S_M^M is given by

$$\lambda < f_{1d} \quad (29)$$

where λ is the aggregate arrival rate for the M terminals.

Next, let us consider the stability of symmetric CCMA-S. Since S_M^M as described before dominates CCMA-S, if S_M^M is stable then CCMA-S is also stable. Therefore, for $\lambda < f_{1d}$, system CCMA-S is stable. On the other hand, if all the queues in S_M^M are unstable, then none of these queues ever empty, hence, the relay loses its role and both systems CCMA-S and S_M^M are *indistinguishable* if both started with the same initial conditions. Therefore, if we have $\lambda > f_{1d}$, then all the queues in S_M^M are unstable and accordingly system CCMA-S is unstable as well. Therefore, the maximum stable throughput for system CCMA-S can be summarized in the following theorem.

Theorem 2: The maximum stable throughput $\lambda_{MST}(\text{CCMA-S})$ for system CCMA-S is equal to that of a TDMA system without a relay and is given by

$$\lambda_{MST}(\text{CCMA-S}) = f_{1d}. \quad (30)$$

However, we conjecture that for the general asymmetric M-terminal scenario, the whole stability region of TDMA will be contained inside that of CCMA-S. Another important issue to point out is that although CCMA-S and TDMA have the same maximum stable throughput for the symmetrical case, the two systems do not have the same delay performance as will be discussed later.

B. Stability Analysis of CCMA With Multiple Frames (CCMA-M)

In CCMA-M, the relay's queue can possibly grow and hence should be taken into account when studying the system stability. This means that for stability we require both the M terminals' queues and the relay's queue to be stable. The stability region of the whole system is the intersection of the stability regions of the M terminals and that of the relay. First, we consider the $M = 2$-terminal case. According to the operation of system CCMA-M, a terminal succeeds in transmitting a packet if either the destination or the relay receives this packet correctly. The success probability of terminal $i \in \{1, 2\}$ in CCMA-M can thus be specified as follows:

$$P_i = f_{id} + f_{il} - f_{id}f_{il}. \tag{31}$$

We first consider the stability region for the system determined just by the terminals' queues. Since for each queue $i \in \mathcal{M}$, the queue behaves exactly as in a TDMA system with the success probability determined by (31), the stability region $\mathcal{R}_{\mathcal{M}}(\text{CCMA-M})$ for the set of queues in $\mathcal{M}$ is given by

$$\mathcal{R}_{\mathcal{M}}(\text{CCMA-M}) = \{[\lambda_1, \lambda_2, \ldots, \lambda_M] \in R_+^M : \lambda_i < \omega_i P_i,$$
$$\forall i \in \mathcal{M} \text{ and } [\omega_1, \omega_2, \ldots, \omega_M] \in F\}. \tag{32}$$

Next we study the stability of the relay's queue l. The evolution of the relay's queue can be modeled as

$$Q_l^t = (Q_l^t - Y_l^t)^+ + X_l^t \tag{33}$$

where X_l^t denotes the number of arrivals at time slot t and Y_l^t denotes the possibility of serving a packet at this time slot from the relay's queue ($Y_l^t(G)$ takes values in $\{0, 1\}$). If the terminals' queues are stable, then by definition the departure processes from both terminals are stationary. A packet departing from a terminal queue is stored in the relay's queue (i.e., counted as an arrival) if simultaneously the following two events happen: the terminal-destination channel is in outage and the terminal-relay channel is not in outage. Hence, the arrival process to the queue can be modeled as follows:

$$X_l^t = \sum_{i \in \mathcal{M}} \mathbf{1} \left[A_i^t \bigcap \{Q_i^t \neq 0\} \bigcap O_{id}^t \bigcap \overline{O_{il}^t} \right]. \tag{34}$$

In (34), $\{Q_i^t \neq 0\}$ denotes the event that terminal's i queue is not empty, which under (32) has probability $\lambda_i/(\omega_i P_i)$, where P_i is defined in (31). The random processes involved in the above expressions are all stationary, hence, the arrival process to

the relay is stationary. The expected value of the arrival process can be computed as follows:

$$\lambda_l = \sum_{i \in \mathcal{M}} \lambda_i \frac{(1 - f_{id})f_{il}}{P_i}. \tag{35}$$

Similarly, we establish the stationarity of the service process from the relay's queue. The service process of the relay's queue depends by definition on the empty slots available from the terminals, and the channel from the relay to the destination being not in outage. By assuming the terminals' queues to be stable, they offer stationary empty slots (stationary service process) to the relay. Also the channel statistics is stationary, hence, the relay's service process is stationary. The service process of the relay's queue can be modeled as

$$Y_l^t = \sum_{i \in \mathcal{M}} \mathbf{1} \left[A_i^t \bigcap \{Q_i^t = 0\} \bigcap \overline{O_{ld}^t} \right] \tag{36}$$

and the average service rate of the relay can be determined from the following equation:

$$E[Y_l^t] = \sum_{i \in \mathcal{M}} \omega_i \left(1 - \frac{\lambda_i}{\omega_i P_i}\right) f_{ld}. \tag{37}$$

Using Loynes theorem and (35) and (37), the stability region for the relay $\mathcal{R}_l(\text{CCMA-M})$ is determined by the condition $E[X_l^t] < E[Y_l^t]$. The total stability region for system CCMA-M is given by the intersection of two regions $\mathcal{R}_{\mathcal{M}}(\text{CCMA-M}) \bigcap \mathcal{R}_l(\text{CCMA-M})$ which is easily shown to be equal to $\mathcal{R}_l(\text{CCMA-M})$. The stability region for CCMA-M with two terminals is thus characterized as follows:

$$\mathcal{R}(\text{CCMA-M}) = \left\{ [\lambda_1, \lambda_2] \in R^{+2} : \frac{\lambda_1}{P_1}\left((1 - f_{1d})f_{1l} + f_{ld}\right) \right.$$
$$\left. + \frac{\lambda_2}{P_2}\left((1 - f_{2d})f_{2l} + f_{ld}\right) < f_{ld} \right\}. \tag{38}$$

For $M = 2$, this reveals that the stability region of CCMA-M is bounded by a straight line. Since the stability region for TDMA is also determined by a straight line, when comparing both stability regions it is enough to compare the intersection of these lines with the axes. These intersections for CCMA-M are equal to

$$\lambda_1^*(\text{CCMA-M}) = \frac{f_{ld}P_1}{f_{ld} + (1 - f_{1d})f_{1l}}, \quad \lambda_2^*(\text{CCMA-M})$$
$$= \frac{f_{ld}P_2}{f_{ld} + (1 - f_{2d})f_{2l}} \tag{39}$$

while the corresponding values for TDMA are given by

$$\lambda_1^*(\text{TDMA}) = f_{1d}, \qquad \lambda_2^*(\text{TDMA}) = f_{2d}. \tag{40}$$

It is clear that the stability region for TDMA is completely contained inside the stability region of CCMA-M if $\lambda_1^*(\text{CCMA-M}) > \lambda_1^*(\text{TDMA})$ and $\lambda_2^*(\text{CCMA-M}) > \lambda_2^*(\text{TDMA})$. Using (39) and (40), these two conditions are equivalent to

$$f_{ld} > f_{1d}, \qquad f_{ld} > f_{2d}. \tag{41}$$

These conditions have the following intuitive explanation: if the relay-destination channel is better than the terminal-destination channel then it is better to have the relay help the terminal transmit its packets. Note that (41) implies that TDMA

can offer better performance for the terminal whose success probability does not satisfy (41). This possible degradation in the performance does not appear in CCMA-S because the design of CCMA-S does not allow the relay to store the packets it received for ever. Hence, the performance of protocol CCMA-S cannot be less than that of TDMA. This calls for the development of an enhanced version of protocol CCMA-M that takes this into account.

1) Enhanced Protocol CCMA-Me: In this enhanced strategy, the relay only helps the terminals which, on average, have worst channel condition than the relay itself. In other words, the relay helps the terminal whose outage probability to the destination satisfy $f_{ld} > f_{id}$ for $i \in \mathcal{M}$, which can be determined only by the distance between the terminal and the destination. Other terminals that do not satisfy this inequality operate as in TDMA, i.e., the relay does not help them.

Next we calculate the stability region for the enhanced system and consider $M = 2$ terminals for illustration. Assume that the relay only helps terminal 1. Similar to our calculations for the arrival and service processes for the relay in CCMA-M, we can show that the average arrival rate to the relay in CCMA-Me is given by

$$E[X_l^t(\text{CCMA-Me})] = \frac{\lambda_1}{P_1}(1 - f_{1d})f_{1l} \qquad (42)$$

and the average service rate to the relay is given by

$$\begin{aligned} &E[Y_l^t(\text{CCMA-Me})] \\ &= \left(\omega_1 \left(1 - \frac{\lambda_1}{\omega_1 P_1} \right) + \omega_2 \left(1 - \frac{\lambda_2}{\omega_2 f_{2d}} \right) \right) f_{ld}. \quad (43) \end{aligned}$$

Using Loynes theorem [16] and (42) and (43), the stability region $\mathcal{R}(\text{CCMA-Me})$ is given by

$$\begin{aligned} &\mathcal{R}(\text{CCMA-Me}) \\ &= \left\{ [\lambda_1, \lambda_2] \in R_+^2 : \frac{\lambda_1}{P_1}\big((1 - f_{1d})f_{1l} + f_{ld}\big) + \lambda_2\frac{f_{ld}}{f_{2d}} < f_{ld} \right\}. \end{aligned}$$
$$(44)$$

The stability region for the enhanced protocol CCMA-Me is not smaller than the stability region of TDMA $\mathcal{R}(\text{TDMA}) \subseteq \mathcal{R}(\text{CCMA-Me})$, and the proof simply follows from the construction of the enhanced protocol CCMA-Me.

For a general M-terminal case, the analysis is the same and the stability region for CCMA-Me can be fully characterized as follows.

Theorem 3: The stability region for M-terminals CCMA-Me is specified as

$$\begin{aligned} &\mathcal{R}(\text{CCMA-Me}) \\ &= \Bigg\{ [\lambda_1, \lambda_2, \ldots, \lambda_M] \in R_+^M : \sum_{i \in \mathcal{M}_1} \frac{\lambda_i}{P_i}\big((1 - f_{id})f_{il} + f_{ld}\big) \\ &\quad + \sum_{j \in \mathcal{M}_2} \lambda_j \frac{f_{ld}}{f_{jd}} < f_{ld} \Bigg\} \end{aligned}$$
$$(45)$$

where $\mathcal{M}_1 = \{i \in \mathcal{M} : f_{ld} > f_{id}\}$, or the set of terminals that the relay helps, and $\mathcal{M}_2 = \{i \in \mathcal{M} : f_{ld} < f_{id}\}$ is the complement set.

It remains to specify the relation between the stability regions of CCMA-S and CCMA-Me, which is characterized in the following theorem.

Theorem 4: The stability region of two-user CCMA-Me contains that of two-user CCMA-S. In other words,

$$\mathcal{R}(\text{TDMA}) \subseteq \mathcal{R}(\text{CCMA-S}) \subseteq \mathcal{R}(\text{CCMA-Me}). \qquad (46)$$

Proof of Theorem 4: See Appendix C. $\qquad \square$

C. Existing Cooperation Protocols: Stability Analysis

In this subsection, we discuss stability results for some existing decode-and-forward cooperation protocols. In particular, we consider the family of adaptive relaying proposed in [1], which comprises selection and incremental relaying. In the following, we discuss stability results for these two protocols and compare them to our proposed CCMA protocol.

1) Stability Region for Selection Decode-and-Forward: In this subsection, we characterize the stability region of selection decode-and-forward (SDF) described as follows. The cooperation is done in two phases. In the first phase, the source transmits and both the relay and the destination listen. In the second phase, if the relay is able to decode the signal correctly, then it is going to forward the received packet to the destination, otherwise, the source retransmits the packet. Accordingly, there is always a specified channel resource dedicated for the relay to help the source, which is different from the opportunistic nature of cooperation in our proposed algorithms.

In the following, we analyze the outage probability for SDF under two scenarios. In the first scenario, the structure of the packets arrivals at the terminal is not allowed to be altered. Thus, each packet is transmitted in two consecutive time slots with the original spectral efficiency (for example, using the same modulation scheme). This, however, results in SDF having half the bandwidth efficiency of TDMA and CCMA because each packet requires two time slots for transmission. In the second scenario, the bandwidth efficiency is preserved among all protocols. This can be done by allowing the terminal and the relay to change the structure of the incoming packets so that each of them transmits at twice the incoming rate (twice the spectral efficiency). Hence, each packet is now transmitted in one time slot again, and the average spectral efficiency for SDF under this scenario is equal to those of TDMA and CCMA [1].

For the first scenario, an outage occurs if both the source–destination link and the source–relay–destination link are in outage. This can be specified as follows:

$$\begin{aligned} &\text{Pr}_{\text{SDF}}(O) \\ &= \text{Pr} \Bigg[\Big(\{\text{SNR}_{i,d} < 2\beta\} \bigcap \{\text{SNR}_{i,l} < 2\beta\} \bigcap \{\text{SNR}_{i,d} < 2\beta\} \Big) \\ &\quad \bigcup \Big(\{\text{SNR}_{i,d} < 2\beta\} \bigcap \{\text{SNR}_{i,l} > 2\beta\} \\ &\quad \bigcap \{\text{SNR}_{l,d} < 2\beta\} \Big) \Bigg], \qquad i \in \mathcal{M}. \end{aligned}$$
$$(47)$$

The factor 2 in front of the SNR threshold β is to account to the fact that the transmitted power is divided by 2 in the first scenario to have the same energy per bit. (Note that the bandwidth efficiency of SDF in the first scenario is half that of CCMA, hence, we need to reduce the transmit power by half.) The first term on the right-hand side of (47) corresponds to the event that both the source–destination and the source–relay links were in outage in the first time slot, and the source–destination link remained in outage in the second time slot. The second term in (47) corresponds to the event that the source–destination link was in outage and the source–relay link was not in outage in the first time slot, but the relay–destination link was in outage in the second time slot. The probability in (47) can be expressed as

$$\Pr_{i,\text{SDF}}(O) = \left(1 - f_{id}^2\right)^2 \left(1 - f_{il}^2\right) + \left(1 - f_{id}^2\right) f_{il}^2 \left(1 - f_{ld}^2\right) \tag{48}$$

where f_{ij} is defined in (7). Since a single packet is transmitted in two time slots, one can think of this protocol as a modified TDMA system with the cooperation time slot has twice the length of the time slot in TDMA. The average arrival rate per cooperation time slot is $2\lambda_i$ for $i \in \mathcal{M}$. Loynes condition for stability is given by

$$\frac{\lambda_1}{1 - \Pr_{1,\text{SDF}}(O)} + \frac{\lambda_2}{1 - \Pr_{2,\text{SDF}}(O)} < \frac{1}{2} \tag{49}$$

where $1 - \Pr_{\text{SDF}}(O)$ is the success probability for terminal i.

For the second scenario, we need to calculate the SNR threshold β' corresponding to transmitting at twice the rate. It is in general very difficult to find an explicit relation between the SNR threshold β and the transmission rate, and thus we render to a special case to capture the insights of this scenario. Let the outage be defined as the event that the mutual information I between two terminals is less than some specific rate R [29]. If the transmitted signals are Gaussian, then according to our channel model, the mutual information between terminal $i \in \mathcal{T}$ and terminal $j \in \mathcal{D}$ is given by $I = log(1 + \text{SNR}_{ij})$. The outage event for this case is defined as

$$O_I \triangleq \{h_{ij} : I < R\}. \tag{50}$$

The above equation implies that if the outage is defined in terms of the mutual information and the transmitted signals are Gaussian, then the SNR threshold β and the spectral efficiency R are related as $\beta = 2^R - 1$, i.e., they exhibit an exponential relation. Hence, for protocol SDF when transmitting at twice the rate the corresponding SNR threshold β' is given by $\beta' = 2^{2R} - 1$, and given β one can find β' through the previous equation. Note that we do not reduce the power in this second scenario because both SDF and CCMA will have the same spectral efficiency.[3]

2) Stability Region for Incremental Decode-and-Forward: The second relaying strategy that we are considering in our comparison is incremental relaying. In such a strategy, feedback from the destination in the form of ACK or NACK is utilized

at the relay node to decide whether to transmit or not. In our comparison, we consider a modified version of the incremental relaying strategy proposed in [1]. In particular, we consider a decode-and-forward incremental relaying with selection capability at the relay (SIDF). In SIDF, the first phase is exactly as amplify-and-forward incremental relaying [1]. In the second phase, if the destination does not receive correctly then the relay, if it was able to decode the source signal correctly, forwards the re-encoded signal to the destination, otherwise, the source retransmits again. One can think of this protocol as combining the benefits of selection and incremental relaying.

Next we analyze the outage probability of SIDF. As we did when studying SDF, we are also going to consider two scenarios for SIDF, namely, when the packet structure is not allowed to be changed and the scenario of equal spectral efficiency. The spectral efficiency of SIDF in the first scenario is less that TDMA or CCMA because the relay is occasionally allocated some channel resources for transmission with positive probability. Since both SDF and SIDF have the same mechanism for the outage event, it is readily seen that the outage event for SIDF is also given by (47) with the difference that we only use β in this case without the term 2 because SIDF will use the same transmit power.[4] The outage event is thus given by

$$\Pr_{\text{SIDF}}(O) = \left(1 - f_{id}\right)^2 \left(1 - f_{il}\right) + \left(1 - f_{id}\right) f_{il} \left(1 - f_{ld}\right). \tag{51}$$

The above expression represents the success probability of transmitting a packet in one or two consecutive time slots. Terminal i uses one time slot with probability f_{id} and two time slots with probability $1 - f_{id}$. The average number of time slots used by terminal i during a frame in SIDF is thus given by $2 - f_{id}$. The set of queues are not interacting in this case and the stability region is simply given by

$$\sum_{i \in \mathcal{M}} \frac{\lambda_i (2 - f_{id})}{1 - \Pr_{\text{SIDF}}(O)} < 1. \tag{52}$$

Next, we consider the second scenario of SIDF, where the spectral efficiency is preserved for SIDF as for TDMA or CCMA. In this scenario, both the terminals and the relay will be transmitting at a higher rate $\tilde{R}$ such that the average spectral efficiency is equal to the spectral efficiency R of TDMA or CCMA. The average spectral efficiency $R(\text{SIDF})$ of SIDF when transmitting at a spectral efficiency $\tilde{R}$ is given by

$$R(\text{SIDF}) = \tilde{R} \tilde{f}_{id} + \frac{\tilde{R}}{2} \left(1 - \tilde{f}_{id}\right) \triangleq v(\tilde{R}) \tag{53}$$

where $\tilde{f}_{id}$ is the success probability for terminal's i-destination link when operating at spectral efficiency $\tilde{R}$, and we denote the whole function in the above expression by $v(\cdot)$. For the sake of comparison $R(\text{SIDF}) = v(\tilde{R})$ should be equal to R. Thus, for a given R one should solve for $\tilde{R} = v^{-1}(R)$. This function can lead to many solutions for $\tilde{R}$, and we are going to choose the minimum $\tilde{R}$ [1]. The stability region is thus given by

$$\sum_{i \in \mathcal{M}} \frac{\lambda_i}{1 - \Pr_{\text{SIDF}}(\tilde{O})} < 1 \tag{54}$$

---

[3]Intuitively, under a fixed modulation scheme and fixed average power constraint, one can think of the SNR threshold as being proportional to the minimum distance between the constellation points, which in turn depends on the number of constellation points for fixed average power, and the latter has an exponential relation to the number of bits per symbol that determines the spectral efficiency R.

[4]This is in favor of SIDF because the transmit power should be reduced to account for the reduction in the average spectral efficiency.
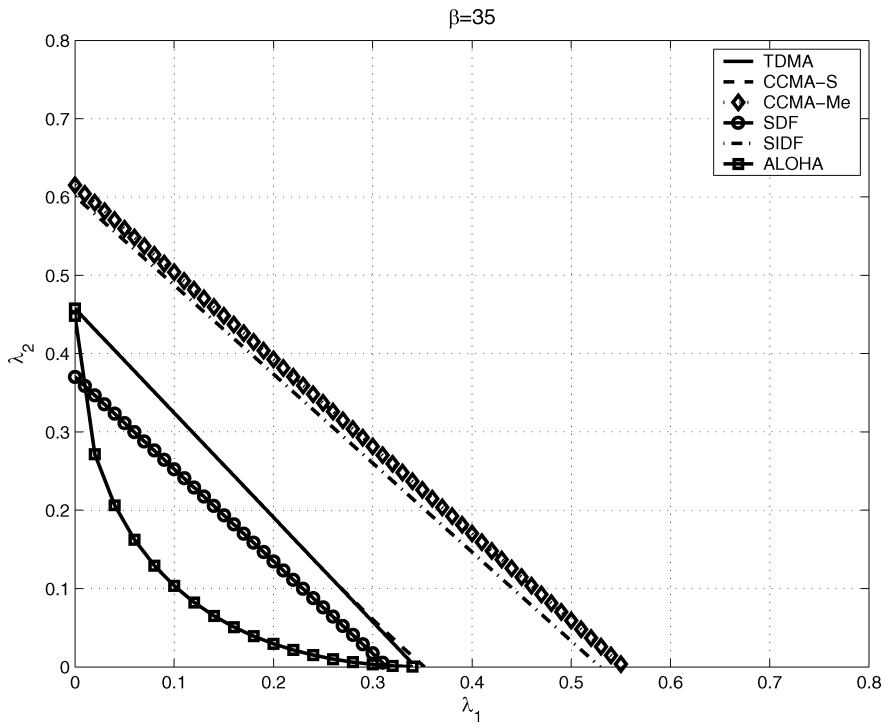
Fig. 3. Stability regions for the different considered protocols at an SNR threshold equal to $\beta = 35$. For this value of β CCMA-S is equivalent to TDMA as depicted. CCMA-Me has the largest throughput region.
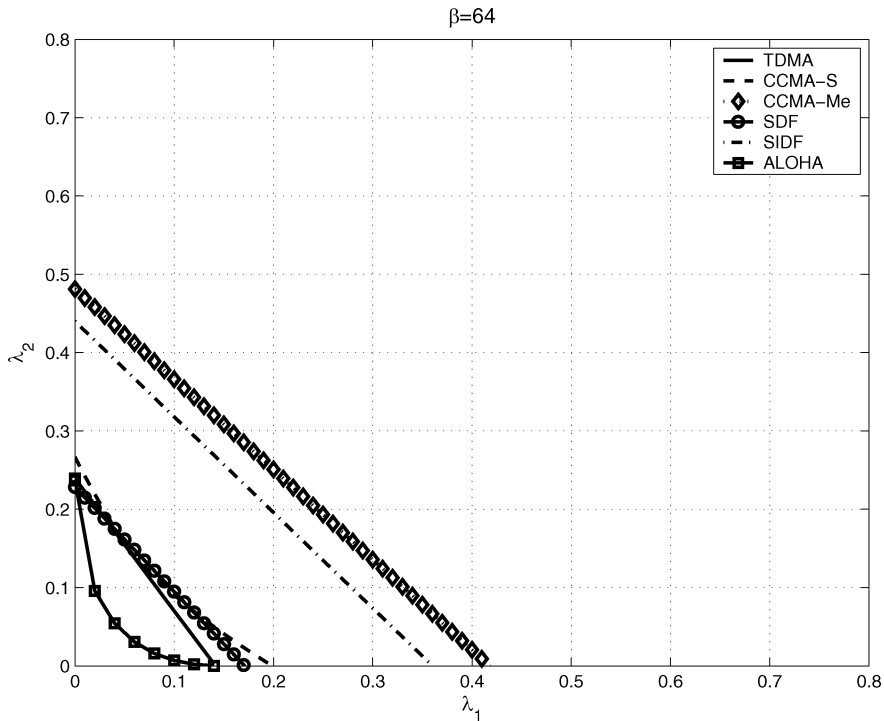

Fig. 4. Stability regions for the different considered protocols at an SNR threshold equal to $\beta = 64$. TDMA is contained in CCMA-S, and the gap between SIDF and CCMA-Me increases in this case.

where $\mathrm{Pr}_{\mathrm{SIDF}}(\tilde{O})$ has the same form as $\mathrm{Pr}_{\mathrm{SIDF}}(O)$ but evaluated at spectral efficiency $\tilde{R}$.

D. Numerical Results

We compare the stability regions of $M = 2$-user TDMA, CCMA-S, CCMA-Me, the two forms of adaptive relaying pro-

posed in [1] (selection and incremental relaying), and ALOHA as an example of random access.

In Figs. 3 and 4 we plot the stability regions for TDMA, CCMA-S, CCMA-Me, SDF, iSIDF, and ALOHA for an SNR threshold of $\beta = 35$, and $\beta = 64$, respectively. For SDF and SIDF we use the first scenario in which the packet structure is not changed. The parameters used to depict these results are
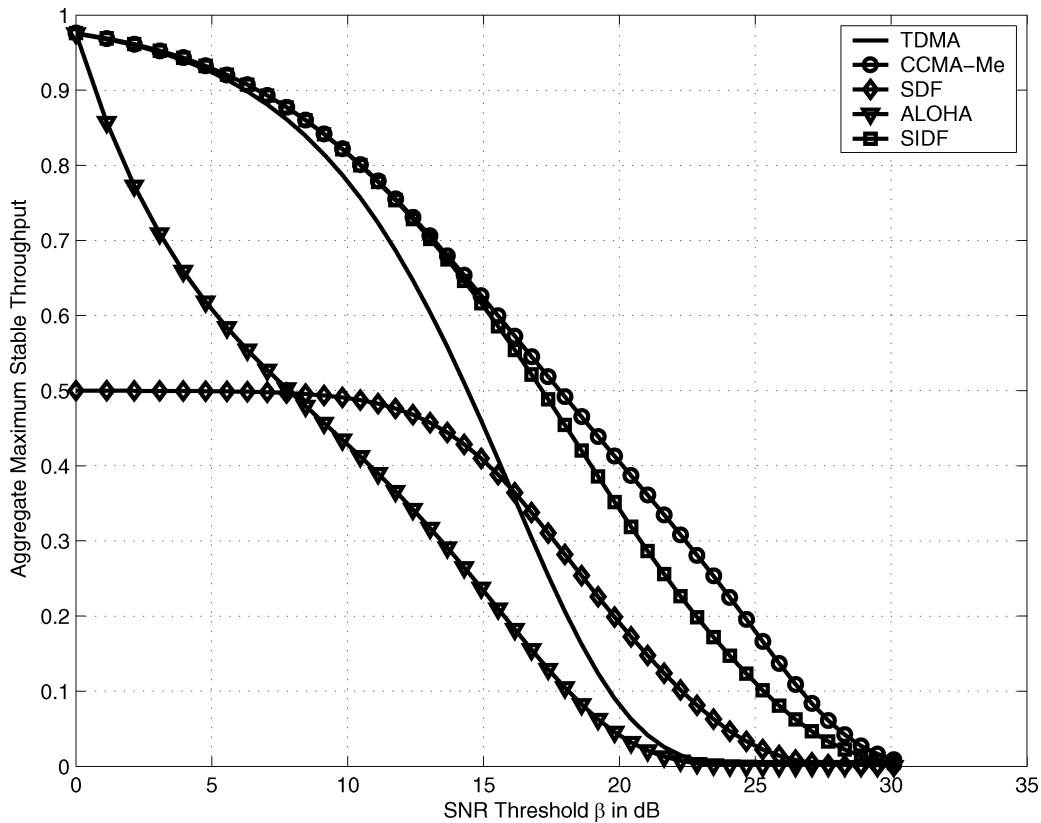
Fig. 5. Aggregate maximum stable throughput versus SNR threshold β in decibels. The propagation path loss is set to $\gamma = 3.5$. The first scenario is used for SDF and SIDF. CCMA-Me has the best tradeoff curve among all the other protocols.

as follows. The distances in meters between different terminals are given by $r_{1,d} = 120$, $r_{2,d} = 110$, $r_{l,d} = 40$, $r_{1,l} = 85$, $r_{2,l} = 80$. The propagation path loss is given by $\gamma = 3.6$, the transmit power $G = 0.01$ watt, and $N_o = 10^{-11}$. In both Figs. 3 and 4, CCMA-Me has the largest stable throughput region. In Fig. 3, CCMA-S and TDMA have identical stable throughput region, and it can be checked that conditions (24) are satisfied for $\beta = 35$. In Fig. 4, TDMA is contained inside CCMA-S. Both CCMA-S and CCMA-Me provide larger stable throughput region over SDF and ALOHA. This is because of the lost bandwidth efficiency in SDF and the interference in ALOHA due to collisions. SIDF is very close to CCMA-Me for smaller values of β, and the gap between them increases with increasing β as depicted in Fig. 4. This is because the bandwidth efficiency of SIDF reduces with increasing β which increases the probability of using a second time slot by the relay.

Next we demonstrate the tradeoff between the maximum stable throughput (MST) versus the SNR threshold β and the transmission rate R. For SDF and SIDF, we consider the two scenarios described in Section IV-C1, where in the first scenario the incoming packet structure is not changed and hence the two protocols have less bandwidth efficiency compared to TDMA and CCMA. While in the second strategy, the packet structure is changed to preserve the bandwidth efficiency. The maximum stable throughput results for the two scenarios are depicted in Figs. 5 and 6, respectively. In both figures, the relative distance between terminals are $r_{1,d} = r_{2,d} = 130$, $r_{l,d} = 50$, and $r_{1,l} = r_{2,l} = 80$. In Fig. 5, the MST is plotted against the SNR

threshold β and the propagation path loss is set to $\gamma = 3.5$. For SDF and SIDF, we use the first scenario. CCMA-Me has the best tradeoff for the whole range. TDMA and CCMA-S have identical performance as proven before for the symmetric case. The maximum attained MST for SDF is 0.5 as one can expect because of the time slot repetition. ALOHA has better performance over SDF for low-SNR threshold, but for medium and high values of β, SDF has better performance. SIDF has close performance to CCMA-Me for low values of β, and CCMA-Me outperforms it for the rest of the SNR threshold range.

In Fig. 6, the MST is plotted against the transmission rate R, and we use the second scenario for SDF and SIDF. For this case, the MST of SDF starts from 1 for low rates R but decays exponentially after that. SIDF and SDF have the best performance for low spectral efficiency regimes where sacrificing the bandwidth by transmitting at higher rate is less significant than the gains achieved by diversity. SIDF performs better than SDF because it is more bandwidth efficient. For higher spectral efficiency regimes, the proposed CCMA-Me provides significantly higher stable throughput compared to SDF or SIDF. An important point to observe from Fig. 6 is the graceful degradation in the performance of CCMA-Me, while the sudden catastrophic performance loss in SDF and SIDF. The rationale here is that the cognitive feature of the proposed CCMA-Me results in no bandwidth loss because cooperation is done in the idle time slots, while both SDF and SIDF suffer from a bandwidth loss which increases for SIDF with increasing R. Our results reveal
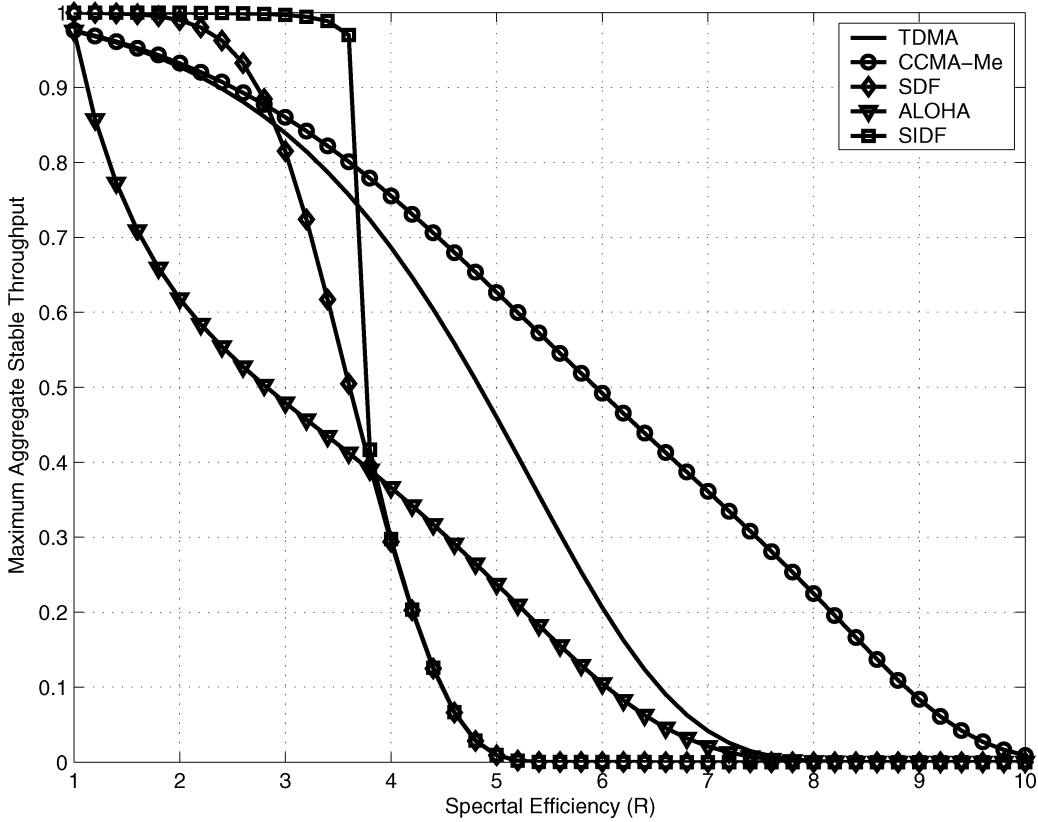
Fig. 6. Aggregate maximum stable throughput versus spectral efficiency R in bits per second per hertz (b/s/Hz). The propagation path loss is set to $\gamma = 3.5$. The second scenario is used for SDF and SIDF. SIDF has the best performance for low spectral efficiency but it suffers from a catastrophic degradation when increasing the spectral efficiency R. CCMA-Me has a graceful degradation due to its bandwidth efficiency, and it has the best tradeoff for medium-to-high spectral efficiency.

a very interesting observation that utilizing empty time slots to increase system reliability via cooperation is a very promising technique in designing cooperative relaying strategies for wireless networks.

V. THROUGHPUT REGION

In the characterization of the stable throughput region in the previous section, the source burstiness is taken into account. Consider now the scenario under which all terminal queues are saturated, i.e., each terminal has infinite number of packets waiting transmission. The maximum throughput supported by any terminal can be defined under such a scenario by the average maximum number of packets that can be transmitted successfully by that terminal [28]. The set of all such saturated throughputs for different resource–sharing vectors defines the throughput region.

Since in both CCMA-S and CCMA-Me the relay role depends on having empty time slots to enable cooperation, there is no surprise that under such saturated queues scenario the relay loses its role and CCMA-S and CCMA-Me reduce to TDMA without relaying. We state this in the following corollary.

Corollary 1: The throughput regions of TMDA, CCMA-S, and CCMA-Me are equivalent

$$\mathcal{C}\,(\text{TDMA}) \equiv \mathcal{C}\,(\text{CCMA-S}) \equiv \mathcal{C}\,(\text{CCMA-Me})\,. \qquad (55)$$

From the preceding corollary, we conclude that the saturated throughput region is a subset of the stability region for both CCMA-S and CCMA-Me. This is an important observation because for ALOHA systems it is conjectured in [28] that the maximum stable throughput region is identical to the throughput region. It is of interest then to point out that CCMA-S and CCMA-Me are examples of multiple-access protocols where the stable throughput region is different from the throughput region.

VI. DELAY ANALYSIS

In this section, we characterize the delay performance of the proposed cognitive cooperative multiple-access protocols, CCMA-S and CCMA-Me.

A. Delay Performance for CCMA-S

In CCMA-S, a packet does not depart a terminal's queue until it is successfully transmitted to the destination. Therefore, the delay encountered by a packet is the one encountered in the terminal's queue. Delay analysis for interacting queues in ALOHA has been studied in [25], [26] and more recently for ALOHA with multipacket reception (MPR) channels in [27], and it turns out to be a notoriously hard problem. Most of the known results are only for the two-user ALOHA case. In this section we consider a symmetric two-user CCMA-S scenario and characterize its delay performance.

Define the moment generating function of the joint queues' sizes processes (Q_1^t, Q_2^t) as follows:

$$G(u,v) = \lim_{t \to \infty} E\left[u^{Q_1^t} v^{Q_2^t}\right]. \qquad (56)$$

From the queue evolution equations in (10), we have

$$E\left[u^{Q_1^{t+1}} v^{Q_2^{t+1}}\right] = E\left[u^{X_1^t} v^{X_2^t}\right] E\left[u^{(Q_1^t - Y_1^t)^+} v^{(Q_2^t - Y_2^t)^+}\right] \qquad (57)$$

where the above equation follows from the independence assumption of the future arrival processes from the past departure and arrival processes. Since the arrival processes are assumed to follow Bernoulli random process, the moment generating function of the joint arrival processes is given by

$$A(u,v) \triangleq \lim_{t \to \infty} E\left[u^{X_1^t} v^{X_2^t}\right] = (u\lambda + 1 - \lambda)(v\lambda + 1 - \lambda) \qquad (58)$$

where due to the symmetry of the two terminals, each has an arrival rate λ. From the definition of the service process of CCMA-S (13), it follows that

$$\begin{aligned} E&\left[u^{(Q_1^t - Y_1^t)^+} v^{(Q_2^t - Y_2^t)^+}\right] \\ &= E\left[\mathbf{1}(Q_1^t = 0, Q_2^t = 0)\right] \\ &\quad + B(u)E\left[\mathbf{1}(Q_1^t > 0, Q_2^t = 0)u^{Q_1^t}\right] \\ &\quad + B(v)E\left[\mathbf{1}(Q_1^t = 0, Q_2^t > 0)v^{Q_2^t}\right] \\ &\quad + D(u,v)E\left[\mathbf{1}(Q_1^t > 0, Q_2^t > 0)u^{Q_1^t} v^{Q_2^t}\right] \quad (59) \end{aligned}$$

where

$$\begin{aligned} B(z) &= \frac{wf_{1d} + w^2 f_{1l}f_{ld}(1 - f_{1d})}{z} \\ &\quad + 1 - \left(wf_{1d} + w^2 f_{1l}f_{ld}(1 - f_{1d})\right) \\ D(u,v) &= wf_{1d}\left(\frac{1}{u} + \frac{1}{v}\right) + 2w(1 - f_{1d}), \qquad (60) \end{aligned}$$

in which we use w to denote symmetric resource-sharing portion of each terminal.

Substituting (58), (59), into (57) and taking the limits we get

$$\begin{aligned} &G(u,v) \\ &= A(u,v)\left(G(0,0) + B(u)\left[G(u,0) - G(0,0)\right] + B(v)[G(0,v) \\ &\quad - G(0,0)] + D(u,v)\left[G(u,v) + G(0,0) \\ &\quad - G(u,0) - G(0,v)\right]\right). \qquad (61) \end{aligned}$$

We can rewrite the above equation as follows: $G(u,v) = \frac{H(u,v)}{F(u,v)}$, where

$$\begin{aligned} H(u,v) &= G(0,0) + B(u)\left[G(u,0) - G(0,0)\right] \\ &\quad + B(v) \times \left[G(0,v) - G(0,0)\right] \\ &\quad + D(u,v)(G(0,0) - G(u,0) - G(0,v)) \end{aligned}$$

and

$$F(u,v) = 1 - A(u,v)D(u,v). \qquad (62)$$

Define $G_1(u,v) \triangleq \frac{\partial G(u,v)}{\partial u}$. Due to symmetry, the average queue size is given by $G_1(1,1)$. Hence, to find the average queuing delay, we need to compute $G_1(1,1)$.

First we find a relation between $G(0,0)$ and $G(1,0)$ using the following two properties which follow from the symmetry of the problem: $G(1,1) = 1$ and $G(1,0) = G(0,1)$. Applying these two properties to (61) along with a simple application of L'Hopital limit theorem we get

$$\begin{aligned} &\left(w^2 f_{1l}f_{ld}(1 - f_{1d})\right) G(0,0) \\ &\quad + \left(wf_{1d} - w^2 f_{1l}f_{ld}(1 - f_{1d})\right) G(1,0) = wf_{1d} - \lambda. \quad (63) \end{aligned}$$

Taking the derivative of (61) with respect to u, applying L'Hopital twice, and using the relation in (63), we get

$$G_1(1,1) = \frac{\lambda(1 - \lambda)}{wf_{1d} - \lambda} - \frac{w^2 f_{1l}f_{ld}(1 - f_{1d})}{wf_{1d} - \lambda}G_1(1,0). \quad (64)$$

To find another equation relating $G_1(1,1)$ and $G_1(1,0)$, we compute $\frac{\partial G(u,u)}{\partial u}$ at $u = 1$. After some tedious but straightforward calculations, we get

$$\begin{aligned} \frac{\partial G(u,u)}{\partial u}\Big|_{u=1} &= 2\lambda - 1 + \frac{wf_{1d} - w^2 f_{1l}f_{ld}(1 - f_{1d})}{wf_{1d} - \lambda}G_1(1,0) \\ &\quad - \frac{4wf_{1d}\lambda - 2wf_{1d} - \lambda^2}{2(wf_{1d} - \lambda)}. \qquad (65) \end{aligned}$$

Due to the symmetry of the problem, we have the following property:

$$\frac{\partial G(u,u)}{\partial u}\Big|_{u=1} = 2G_1(1,1). \qquad (66)$$

Using the above equation, and solving (64) and (65) we get

$$G_1(1,1) = \frac{-\left(2wf_{1d} + w^2 f_{1l}f_{ld}(1 - f_{1d})\right)\lambda^2 + 2wf_{1d}\lambda}{2(wf_{1d} + w^2 f_{1l}f_{ld}(1 - f_{1d}))(wf_{1d} - \lambda)}. \qquad (67)$$

The queueing delay for system CCMA-S can thus be determined as in the following theorem.

Theorem 5: The average queuing delay for a symmetrical two-terminal CCMA-S system is given by

$$\begin{aligned} D(\text{CCMA-S}) &= \frac{G_1(1,1)}{\lambda} \\ &= \frac{-\left(2wf_{1d} + w^2 f_{1l}f_{ld}(1 - f_{1d})\right)\lambda + 2wf_{1d}}{2(wf_{1d} + w^2 f_{1l}f_{ld}(1 - f_{1d}))(wf_{1d} - \lambda)}. \qquad (68) \end{aligned}$$

From Theorem 5, it can be observed that at $\lambda = \omega f_{1d}$ the delay of the system becomes unbounded, i.e., the system becomes saturated. This confirms our previous results in Corollary 1 that for a symmetrical system, both TDMA and CCMA-S have the same maximum stable throughput of $\lambda = \omega f_{1d}$.

B. Delay Performance of CCMA-Me

Due to the symmetrical scenario considered in analyzing the delay performance, if the relay helps one terminal then it helps all terminals, in which case both CCMA-M, CCMA-Me become equivalent. In CCMA-M, a packet can encounter two queuing delays; the first in the terminal's queue and the second in the relay's queue. If a packet successfully transmitted by a terminal
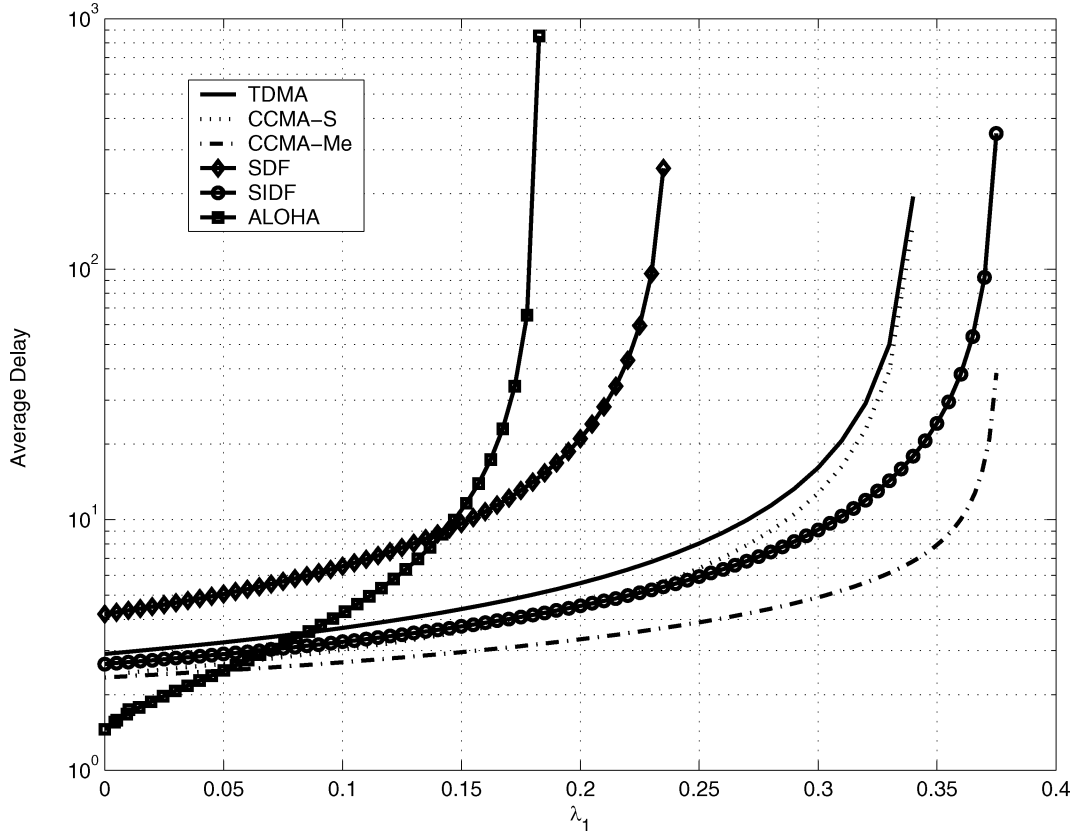
Fig. 7. Average queueing delay per terminal versus the arrival rate for an SNR threshold of $\beta = 15$.

directly goes to the destination, then this packet is not stored in the relay's buffer. Denote this event by ξ. The total delay encountered by a packet in CCMA-M can thus be modeled as

$$T(\text{CCMA-M}) = \begin{cases} T_t, & \xi \\ T_t' + T_l, & \bar{\xi} \end{cases} \quad (69)$$

where T_t is the queueing delay in the terminal's queue, and T_l is the queuing delay at the relay's queue. We can elaborate more on (69) as follows. For a given packet in the terminal's queue, if the first successful transmission for this packet is to the destination, then the delay encountered by this packet is only the queuing delay in the terminal's queue. On the other hand, if the first successful transmission for this packet is not to the destination, then the packet will encounter the following delays: queuing delay in the terminal's queue in addition to the queuing delay in the relay's queue.

First, we find the queuing delay either in the terminal's or relay's queue, as both queues have similar evolution equations, with the difference being in the average arrival and departure rates. Using the same machinery utilized in the analysis of the queueing delay in CCMA-S to analyze the delay performance of CCMA-M, the average queue size can be found as

$$E[N] = \frac{\lambda(1 - \lambda)}{\mu - \lambda} \quad (70)$$

where λ denotes the average arrival rate and μ denotes the average departure rate. We now compute the average delay in (69).

The probability that, for any packet, the first successful transmission from the terminal's queue is to the destination is given by

$$\Pr[\xi] = \frac{f_{1d}}{f_{1d} + f_{1l} - f_{1d}f_{1l}} = \frac{f_{1d}}{P_1}. \quad (71)$$

From (69)–(71), the average delay for system CCMA-M is thus given by

$$D(\text{CCMA-M})$$
$$= \frac{f_{1d}}{P_1} \frac{1 - \lambda}{wP_1 - \lambda} + \frac{f_{1l}(1 - f_{1d})}{P_1} \left(\frac{1 - \lambda}{wP_1 - \lambda} + \frac{1 - \lambda_l}{\mu_l - \lambda_l} \right) \quad (72)$$

where λ_l and μ_l are the average arrival and departure rates, respectively, for the relay's queue defined in (34) and (36). After simplifying the above equation, the average queueing delay for system CCMA-M can be summarized in the following theorem

Theorem 6: The average queuing delay for a packet in a symmetrical 2-terminal CCMA-M system is given by

$$D(\text{CCMA-M}) = \frac{1 - \lambda}{wP_1 - \lambda} + \frac{f_{1l}(1 - f_{1d})}{P_1} \left(\frac{1 - \lambda_l}{\mu_l - \lambda_l} \right). \quad (73)$$

C. Numerical Results for Delay Performance

We illustrate the delay performance of the proposed multiple-access schemes with varying SNR threshold β through some numerical examples. As in the case for the stability region, we include the delay performance of ALOHA, TDMA without
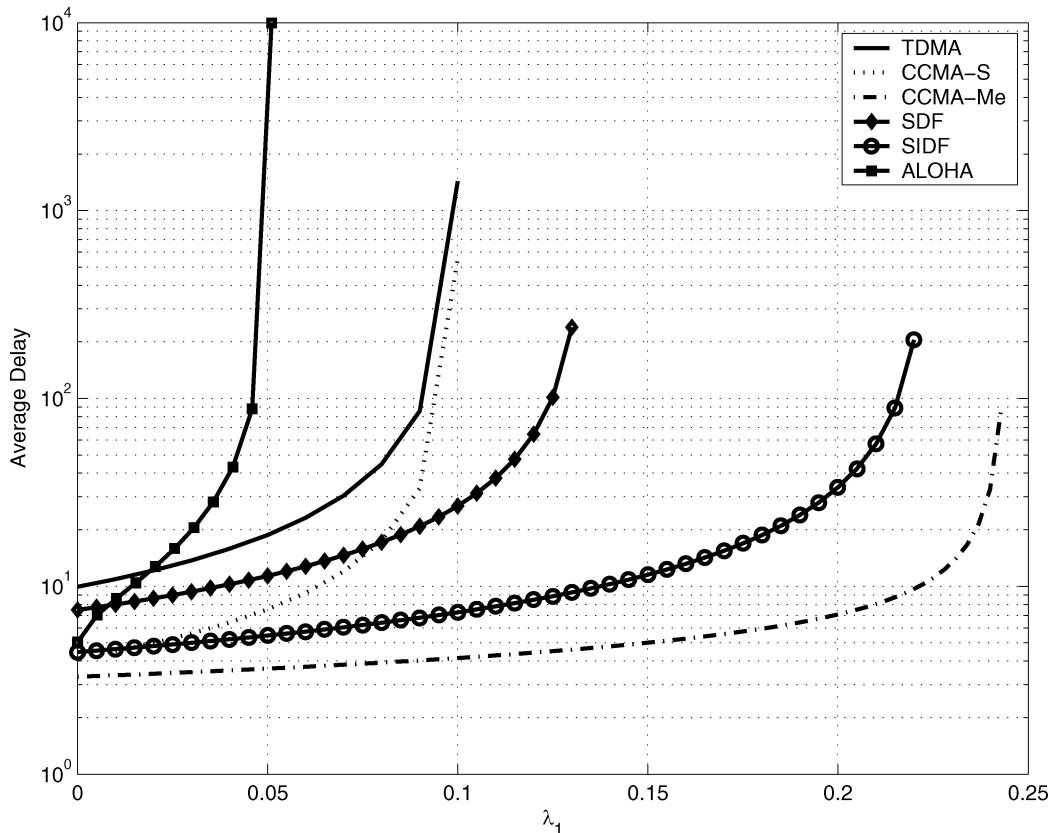
Fig. 8. Average queueing delay per terminal versus the arrival rate for an SNR threshold of $\beta = 64$.

relaying, SDF, and SIDF in our results. Delay performance results for the two-terminal symmetric ALOHA with capture has been derived in [27].

To compare the delay performance of the different multiple-access protocols considered in this paper, we plotted the analytical expressions we got for the queueing delay. The system parameters are the same used to generate the MST plots in Figs. 5 and 6. Figs. 7 and 8 depict the delay results for SNR thresholds $\beta = 15$ and $\beta = 64$, respectively. From Fig. 7, at very low arrival rates, ALOHA has the best delay performance. Increasing the arrival rate λ, both CCMA-S and SDF outperforms other strategies. For higher values of λ, CCMA-Me has the best performance .

The situation changes in Fig. 8 for $\beta = 64$ as both CCMA-S and SIDF outperform ALOHA even for very small arrival rates. The intuition behind this is the more stringent system requirements reflected by the higher SNR threshold $\beta = 64$, which makes the interference in ALOHA more severe. This makes our cognitive multiple-access protocol CCMA-S and CCMA-M perform better than ALOHA because of its high bandwidth efficiency and the gains of cooperation. Another important remark is that although CCMA-S and TDMA has the same MST for the symmetric case, as proven before and as clear from Figs. 7 and 8 where both protocols saturate at the same arrival rate, CCMA-S has always better delay performance than TDMA.

VII. CONCLUSION

In this paper, we have studied the impact of cooperative communications at the multiple-access layer. We introduced a new cognitive multiple-access protocol in the presence of

a relay in the network. The relay senses the channel for idle channel resources and exploits them to cooperate with the terminals in forwarding their packets. We developed two protocols to implement the proposed multiple-access strategy, namely, CCMA-S and CCMA-Me. We characterized the maximum stable throughput region of the proposed protocols and compared them to some existing adaptive relaying strategies, noncooperative TDMA, and random-access ALOHA. Moreover, we studied the delay performance of the proposed protocols. Our analysis reveals significant performance gains of the proposed protocols over their noncognitive counterparts. This is because the proposed multiple-access strategies do not result in any bandwidth loss, as cooperation is enabled only in idle "*unused*" channel resources, which results in a graceful degradation of the maximum stable throughput when increasing the communication rate. On the other hand, the maximum stable throughput of noncognitive relaying strategies such as selection and incremental relaying suffers from catastrophic degradation with increasing the communication rate, because of their inherent bandwidth inefficiency.

Furthermore, we show that the throughput region of the proposed protocols is a subset from their maximum stable throughput region, which is different from the case in ALOHA where it is conjectured that both regions are identical.

APPENDIX A
PROOF OF THEOREM 1

The stable throughput region of CCMA-S for a fixed resource-sharing vector (ω_1, ω_2) is specified in Lemma 1. In order to find the whole stability region of the protocol, we

need to take the union over all possible values of (ω_1, ω_2) in F. One method to characterize this union is to solve a constrained optimization problem to find the maximum feasible λ_2 corresponding to each feasible λ_1. For a fixed λ_1, the maximum stable arrival rate for queue 2 is given by solving the following optimization problem:

$$\max_{w_1, w_2} \lambda_2 = w_2 f_{2d} + w_1 w_2 K_2 - \frac{\lambda_1 w_2 K_2}{f_{1d}},$$
$$\text{s.t. } w_1 + w_2 \leq 1, \qquad \lambda_1 \leq w_1 f_{1d} \qquad (74)$$

where $K_i = (1 - f_{2d}) f_{2l} f_{ld}$, and $i \in \{1, 2\}$. To put this problem in a standard form, we can equivalently write it as the minimization of $-\lambda_2$. The Lagrangian of this optimization problem is given by

$$\begin{bmatrix} -w_2 K_2 \\ -w_1 K_2 + \frac{\lambda_1 K_2}{f_{1d}} - f_{2d} \end{bmatrix} + u_1 \begin{bmatrix} 1 \\ 1 \end{bmatrix} + u_2 \begin{bmatrix} -f_{1d} \\ 0 \end{bmatrix} = 0 \quad (75)$$

where u_1, u_2 are the complementary slackness variables that are nonnegative. Solving for the complementary slackness variables we get the following relation between u_1 and u_2:

$$u_1 = w_2 K_2 + u_2 f_{1d}. \qquad (76)$$

This shows that $u_1 > 0$, i.e., the first constraint ($w_1 + w_2 = 1$) in (74) is met with equality.

Substituting $w_2 = 1 - w_1$ in the cost function in (74), we get

$$\lambda_2 = (1 - w_1) f_{2d} + w_1 (1 - w_1) K_2 - (1 - w_1) \frac{\lambda_1 K_2}{f_{1d}}. \quad (77)$$

Taking the first derivative of the above equation with respect to w_1, we get

$$\frac{\partial \lambda_2}{\partial w_1} = -f_{2d} + K_2 - 2 w_1 K_2 + \frac{\lambda_1 K_2}{f_{1d}}. \qquad (78)$$

Since K_2 is nonnegative, the second derivative is negative, hence, the cost function in (77) is concave in w_1 and the necessary conditions (the Karush–Kuhn–Tucker conditions) for optimality [32] are also sufficient. Equating the first derivative in (78) to zero, the solution w_1^* is given by

$$w_1^* = \frac{1}{2 K_2} \left(K_2 - f_{2d} + \frac{\lambda_1 K_2}{f_{1d}} \right). \qquad (79)$$

From the first constraint in (74), the minimum value for w_1 that guarantees the stability of queue 1 in system S_1 is given by $w_{1,\min} = \frac{\lambda_1}{f_{1d}}$. Hence, if $w_1^* > w_{1,\min}$, and given concavity of the cost function, the optimal solution is just w_1^*, otherwise it is given by $w_{1,\min}$. Characterizing this in terms of the channel parameters, the optimal solution for the optimization problem in (74) is given by

$$w_1^* = \begin{cases} \frac{1}{2 K_2} \left(K_2 - f_{2d} + \frac{\lambda_1 K_2}{f_{1d}} \right), & \lambda_1 \leq f_{1d} - \frac{f_{1d} f_{2d}}{K_2} \\ \frac{\lambda_1}{f_{1d}}, & f_{1d} - \frac{f_{1d} f_{2d}}{K_2} < \lambda_1 < f_{1d}. \end{cases} \quad (80)$$

If $\lambda_1 > f_{1d}$, then the first queue can never be stable by construction of S_1.

Now we solve for the other branch in the stability region given by the dominant system S_2. The equations for this branch are

given by (16). Similar to the first stability region branch, solving for the Lagrangian of this branch gives the necessary condition $w_1 + w_2 = 1$. First, we find the maximum achievable stable rate for the first queue, which is achieved when $\lambda_2 = 0$. Substituting in (16), the arrival rate λ_1 can be written as

$$\lambda_1 = w_1 f_{1d} + w_1 (1 - w_1) K_1. \qquad (81)$$

The above equation is obviously concave. Taking the first derivative and equating to zero we get

$$w_1^* |_{\lambda_2 = 0} = \frac{1}{2 K_1} (f_{1d} + K_1). \qquad (82)$$

Since $w_1 \leq 1$, then if $w_1^* |_{\lambda_2 = 0} > 1$, i.e., $f_{1d} > K_1$, and given the concavity of the cost function, we let $w_1^* |_{\lambda_2 = 0} = 1$. Substituting $w_1^* |_{\lambda_2 = 0}$ in (81), the maximum achievable rate λ_1 can be given by

$$\lambda_1^* = \begin{cases} f_{1d}, & f_{1d} > K_1 \\ \frac{1}{4 K_1} (f_{1d} + K_1)^2, & 0 \leq f_{1d} \leq K_1. \end{cases} \quad (83)$$

Next for a fixed λ_1, we solve for the optimal λ_2 that can be achieved from the second branch. We can write (16) in terms of λ_2 as follows:

$$\lambda_2 = \frac{f_{1d} f_{2d}}{K_1} + (1 - w_1) f_{2d} - \lambda_1 \frac{f_{2d}}{w_1 K_1}. \qquad (84)$$

Taking the first derivative with respect to w_1 we get

$$\frac{\partial \lambda_2}{\partial w_1} = -f_{2d} + \lambda_1 \frac{f_{2d}}{K_1 w_1^2}. \qquad (85)$$

The second derivative is negative, which renders the whole function concave. Equating the first derivative to zero we get

$$w_1^* = \sqrt{\frac{\lambda_1}{(1 - f_{1d}) f_{1l} f_{ld}}}. \qquad (86)$$

From (16), we have the following constraint for the stability of the second queue, $\lambda_2 \leq w_2 f_{2d}$, which can be written in terms of w_1 as follows:

$$w_1 \leq 1 - \frac{\lambda_2}{f_{2d}}. \qquad (87)$$

Substituting λ_2 from (87) into (84), we get that the maximum value for w_1 in terms of λ_1 is given by $w_1 \leq \frac{\lambda_1}{f_{1d}}$. Therefore, the optimal value for w_1 can be written as

$$w_1^* = \min \left(\sqrt{\frac{\lambda_1}{(1 - f_{1d}) f_{1l} f_{ld}}}, \frac{\lambda_1}{f_{1d}} \right). \qquad (88)$$

The value of the expression in (88) can be shown never to exceed 1 by substituting the maximum value of λ_1 given by (83) in the above equation. After some manipulations, the optimum w_1^* in (88) can be further simplified as follows

$$w_1* = \begin{cases} \frac{\lambda_1}{f_{1d}}, & \text{if } \lambda_1 \leq \frac{f_{1d}^2}{(1 - f_{1d}) f_{1l} f_{ld}} \\ \sqrt{\frac{\lambda_1}{(1 - f_{1d}) f_{1l} f_{ld}}}, & \text{otherwise.} \end{cases} \quad (89)$$

We summarize equation (77), (80), (84), (35) describing the envelopes of the first and second branches as follows. For the
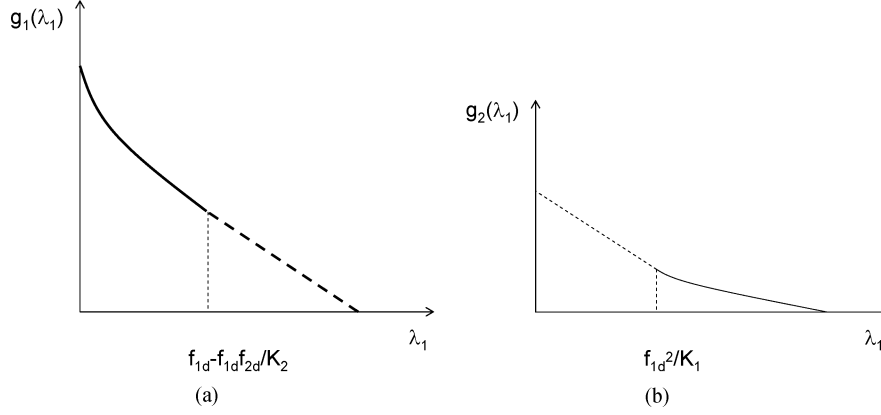
Fig. 9. Envelopes for the stability region of CCMA-S. (a) First branch. (b) Second branch.

first branch given by (77), (80), substituting (80) in (77) we get (90) at the bottom of the page. Similarly, by substituting (89) in (84), we get for the second branch

$$g_2(\lambda_1)$$
$$= \begin{cases} f_{2d} - \frac{f_{2d}}{f_{1d}}, & \lambda_1 < \frac{f_{1d}^2}{(1-f_{1d})f_{1l}f_{ld}} \\ \frac{f_{1d}f_{2d}}{K_1} + f_{2d} - 2f_{2d}\sqrt{\frac{\lambda_1}{K_1}}, & \frac{f_{1d}^2}{(1-f_{1d})f_{1l}f_{ld}} \leq \lambda_1 \leq \lambda_1^*. \end{cases}$$
(91)

The value for λ_1^* is given by (83). Since $g_1(\lambda_1)$ and $g_2(\lambda_2)$ are achieved by the dominant systems S_1 and S_2, respectively, then they are both achieved by CCMA-S. This proves Theorem 1. □

APPENDIX B
PROOF OF LEMMA 2

We prove the convexity of the envelope of region $\mathcal{R}(\text{CCMA-}S)$. First we consider the envelope $g_1(\lambda_1)$ for the first branch. As shown in Fig. 9(a), $g_1(\lambda_1)$ is defined over two regions: For $\lambda_1 \in [0, f_{1d} - \frac{f_{1d}f_{2d}}{K2}]$

$$g_1(\lambda_1) \triangleq g_{11}(\lambda_1) = K_2 \left(\frac{\lambda_1}{2f_{1d}} + 0.5 - \frac{f_{2d}}{2K_2} \right)^2 - \frac{K_2\lambda_1}{f_{1d}} + f_{2d}$$

while for $\lambda_1 \in (f_{1d} - \frac{f_{1d}f_{2d}}{K2}, f_{1d}]$, $g_1(\lambda_1) \triangleq g_{12}(\lambda_1)$ is a straight line given by $f_{2d} - \frac{f_{2d}}{f_{1d}}$. It can be readily seen that $g_1(\lambda_1)$ is convex over both regions because its second derivative is nonnegative. Thus, to prove that $g_1(\lambda_1)$ is convex over the whole region, we check the continuity and the first derivative at the intersection point $f_{1d} - \frac{f_{1d}f_{2d}}{K2}$. It is simple to show that $g_1(\lambda_1)$ is continuous at this point. Now, it remains to check the slope of the tangent at the intersection point. Taking the first derivative of g_{11} and g_{12}, we can show that

$$\frac{\partial g_{11}(\lambda_1)}{\partial \lambda_1} = \left. \frac{\partial g_{12}(\lambda_1)}{\partial \lambda_1} \right|_{\lambda_1 = f_{1d} - \frac{f_{1d}f_{2d}}{K2}}.$$
(92)

Therefore, $g_1(\lambda_1)$ is also differentiable, which proves that $g_1(\lambda_1)$ is convex over its domain of definition. Similar arguments apply for $g_2(\lambda_1)$ depicted in Fig. 9(b) to prove its convexity.

The envelope of the stability region of $\mathcal{R}(\text{CCMA-S})$ is given by $\max[g_1(\lambda_1), g_2(\lambda_2)]$. The maximum of two convex functions can be shown to be convex, and this proves that the envelope of the stability region for system CCMA-S is a convex function. □

APPENDIX C
PROOF OF THEOREM 4

In this appendix, we prove that the stability region of CCMA-S is a subset of that of CCMA-Me. In the proof of this theorem, we use two facts: the envelope of $\mathcal{R}(\text{CCMA-S})$ is convex as proved in Lemma 2, and the envelope of $\mathcal{R}(\text{CCMA-Me})$ is a straight line. Hence, to prove Theorem 3, it suffices to show that the intersections of the envelope of region $\mathcal{R}(\text{CCMA-S})$ with the λ axes are not greater than those for region $\mathcal{R}(\text{CCMA-Me})$. We consider the scenario when both CCMA-S and CCMA-Me have larger stability regions that TDMA, or equivalently $f_{ld}f_{il}(1-f_{id}) \geq f_{id}, i = 1, 2$, because if this condition is not satisfied the stability region of CCMA-S becomes identical to that of TDMA , and CCMA-Me was shown to outperform TDMA, and hence our theorem is true.

We will consider only the intersections with the λ_1 axis, and similar arguments follow for the λ_2 axis. The intersection of $\mathcal{R}(\text{CCMA-S})$ with the λ_1 axis, or the maximum stable arrival rate for the first queue is given by (83), while that for $\mathcal{R}(\text{CCMA-Me})$ is given by (39). Denote the difference between these two quantities by δ given as follows :

$$\delta = \frac{f_{ld}(f_{1d} + f_{1l} - f_{1d}f_{1l})}{f_{ld} + (1-f_{1d})f_{1l}} - \frac{[f_{1d} + (1-f_{1d})f_{1l}f_{ld}]^2}{4(1-f_{1d})f_{1l}f_{ld}}.$$
(93)

If we prove that δ is nonnegative, then we are done. For an arbitrary fixed value for the pair (f_{1d}, f_{ld}), we consider the range of

$$g_1(\lambda_1) = \begin{cases} K_2 \left(\frac{\lambda_1}{2f_{1d}} + 0.5 - \frac{f_{2d}}{2K_2} \right)^2 - \frac{K_2\lambda_1}{f_{1d}} + f_{2d}, & 0 \leq \lambda_1 \leq f_{1d} - \frac{f_{1d}f_{2d}}{K_2} \\ f_{2d} - \frac{f_{2d}}{f_{1d}}\lambda_1, & f_{1d} - \frac{f_{1d}f_{2d}}{K_2} < \lambda_1 \leq f_{1d}. \end{cases}$$
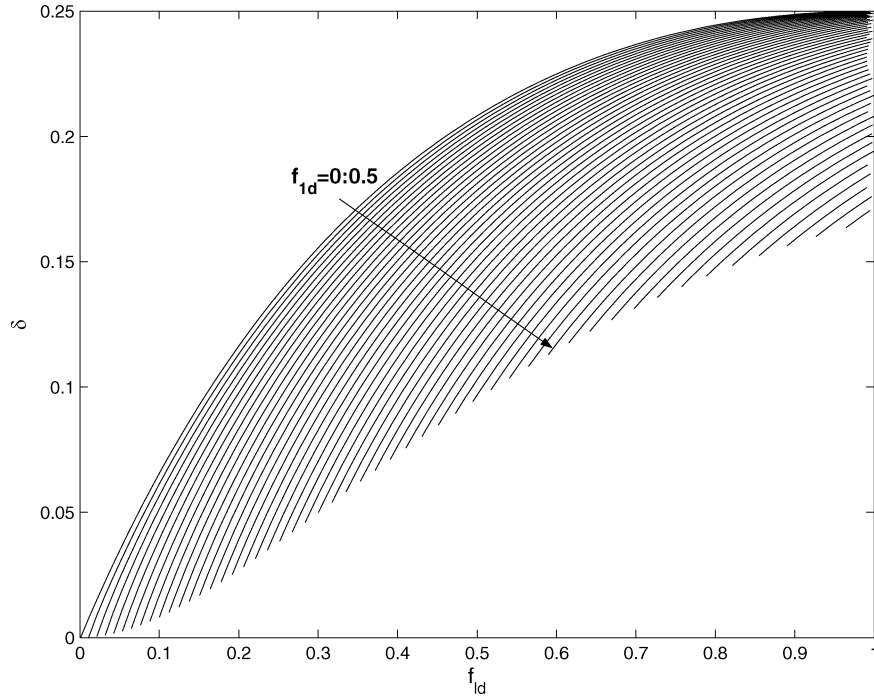(90)

Fig. 10. Numerical evaluation of δ for $f_{1l} = 1$.

f_{1l} that satisfies the constraint $f_{ld}f_{1l}(1 - f_{1d}) \geq f_{1d}$. The first derivative of δ with respect to f_{1l} is given by

$$\frac{\partial \delta}{\partial f_{1l}} = \frac{f_{ld}(1 - f_{1d})(f_{ld} - f_{1d})}{(f_{ld} + (1 - f_{1d})f_{1l})^2}$$
$$- \frac{1}{4}(1 - f_{1d})f_{ld} + \frac{f_{1d}^2}{4(1 - f_{1d})f_{1l}^2 f_{ld}}. \quad (94)$$

Since $f_{ld} > f_{1d}$, the second derivative of the above function is easily seen to be negative, hence, δ is concave in f_{1l} in the region of interest.

f_{1l} takes values in the range $\left[\frac{f_{1d}}{(1-f_{1d})f_{ld}}, 1\right]$. The function δ evaluated at the minimum value of f_{1l} is

$$\delta|_{f_{1l,min}} = \frac{f_{ld}f_{1d} + f_{1d}}{f_{ld} + (1 - f_{1d})f_{1l}} - f_{1d}$$
$$= \frac{f_{1d} - f_{1d}(1 - f_{1d})f_{1l}}{f_{ld} + (1 - f_{1d})f_{1l}} > 0. \quad (95)$$

Hence, δ is positive at the left-most boundary of the feasible region of f_{1l}. For the maximum feasible value of f_{1l} which is equal to 1, the function δ is given by

$$\delta = \frac{f_{ld}}{1 - f_{1d} + f_{ld}} - \frac{(f_{1d} + (1 - f_{1d})f_{ld})^2}{4(1 - f_{1d})f_{ld}}. \quad (96)$$

Fig. 10 depicts δ for $f_{1l} = 1$ over the feasible region of the pair (f_{1d}, f_{ld}). As shown in the figure, the value of δ is always positive. Hence, for an arbitrary feasible value of the pair (f_{1d}, f_{ld}), δ is concave in f_{1l} and positive at the extreme endpoints of the feasible region of f_{1l}, therefore, δ is positive for the whole range of f_{1l} for an arbitrary value of the pair (f_{1d}, f_{ld}). This proves that δ is always positive over the feasible region of success probabilities, and therefore, $\mathcal{R}(\text{CCMA-S})$ is a subset from $\mathcal{R}(\text{CCMA-Me})$. This proves Theorem 3.[5] ☐

REFERENCES

[1] J. N. Laneman, D. N. C. Tse, and G. W. Wornell, "Cooperative diversity in wireless networks: Efficient protocols and outage behavior," *IEEE Trans. Inf. Theory*, vol. 50, no. 12, pp. 3062–3080, Dec. 2004.

[2] J. N. Laneman and G. W. Wornell, "Distributed space-time coded protocols for exploiting cooperative diversity in wireless networks," *IEEE Trans. Inf. Theory*, vol. 49, pt. 10, pp. 2415–2525, Oct. 2003.

[3] G. Kramer, M. Gastpar, and P. Gupta, "Cooperative strategies and capacity theorems for relay networks," *IEEE Trans. Inf. Theory*, vol. 51, no. 9, pp. 3037–3063, Sep. 2005.

[4] G. Scutari and S. Barbarossa, "Distributed space-time coding for regenerative relay networks," *IEEE Trans. Wireless Commun.*, vol. 4, no. 5, pp. 2387–2399, Sep. 2005.

[5] K. Azarian, H. El Gamal, and P. Schniter, "On the achievable diversity–multiplexing tradeoff in half-duplex cooperative channels," *IEEE Trans. Inf. Theory*, vol. 51, no. 12, pp. 4152–4172, Dec. 2005.

[6] P. A. Anghel and M. Kaveh, "Exact symbol error probability of a cooperative network in a Rayleigh fading environment," *IEEE Trans. Wireless Commun.*, vol. 3, no. 5, pp. 1416–1422, Sep. 2004.

[7] W. Su, A. K. Sadek, and K. J. R. Liu, "SER performance analysis and optimum power allocation for decode-and-forward cooperation protocol in wireless networks," in *Proc. IEEE Wireless Communications and Networking Conf. (WCNC'05)*, New Orleans, LA, Mar. 2005, pp. 984–989.

[8] A. K. Sadek, W. Su, and K. J. R. Liu, "Performance analysis for multi-node decode-and-forward relaying in collaborative wireless networks," in *Proc. IEEE Int. Conf, Acoustics, Speech, and Signal Processing (ICASSP)*, Philadelphia, PA, Mar. 2005, pp. 521–524.

[9] A. K. Sadek, W. Su, and K. J. R. Liu, "Multinode cooperative communications in wireless networks," *IEEE Trans. Signal Process.*, vol. 55, no. 1, pp. 341–355, Jan. 2007.

[10] A. K. Sadek, Z. Han, and K. J. R. Liu, "A distributed relay-assignment algorithm for cooperative communications in wireless networks," in *Proc. IEEE Int. Conf. Communications (ICC 2006)*, Istanbul, Turkey, Jun. 2006, pp. 1592–1597.

---

[5]Theorem 3 has another proof that does not need numerical evaluation, however, it is complicated and will not add new insights to the results.

[11] A. K. Sadek, Z. Han, and K. J. R. Liu, "An efficient cooperation protocol to extend coverage area in cellular networks," in *Proc. IEEE Wireless Communications and Networking Conf. (WCNC)*, Las Vegas, NV, Apr. 2006, pp. 1687–1692.

[12] R. Lin and A. P. Petropulu, "Cooperative transmission for random access wireless networks," in *Proc. Int. Conf. Acoustics, Speech and Signal Processing*, Philadelphia, PA, Mar. 2005, pp. 373–376.

[13] G. Scutari, S. Barbarossa, and D. Ludovici, "Cooperation diversity in multihop wireless networks using opportunistic driven multiple access," in *Proc. IEEE Int. Workshop on Signal Processing Advances for Wireless Communications*, Rome, Italy, Jun. 2003, pp. 170–174.

[14] A. Ribeiro, N. D. Sidiropoulos, G. B. Giannakis, and Y. Yu, "Achieving wireline random access throughput in wireless networking via user cooperation," *IEEE Trans. Inf. Theory*, vol. 53, no. 2, pp. 732–758, Feb. 2007.

[15] W. Szpankowski, "Stability conditions for some multiqueue distributed systems: Buffered random access systems," *Adv. Appl. Probab.*, vol. 26, pp. 498–515, 1994.

[16] R. M. Loynes, "The stability of a queue with non-independent interarrival and service times," *Proc. Cambridge Philos. Soc.*, vol. 58, pp. 497–520, 1962.

[17] S. Ghez, S. Verdú, and S. Schwartz, "Stability properties of slotted ALOHA with multipacket reception capability," *IEEE Trans. Autom. Control*, vol. 33, no. 7, pp. 640–649, Jul. 1988.

[18] M. Zorzi and R. Rao, "Capture and retransmission control in mobile radio," *IEEE J. Sel. Areas Commun.*, vol. 12, no. 8, pp. 1289–1298, Oct. 1994.

[19] B. Hajek, A. Krishna, and R. O. LaMaire, "On the capture probability for a large number of stations," *IEEE Trans. Commun.*, vol. 45, no. 2, pp. 254–260, Feb. 1997.

[20] B. Tsybakov and W. Mikhailov, "Ergodicity of slotted ALOHA system," *Probl. Pered. Inform.*, vol. 15, pp. 73–87, 1979.

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

[22] W. Luo and A. Ephremides, "Stability of N interacting queues in random-access systems," *IEEE Trans. Inf. Theory*, vol. 45, no. 5, pp. 1579–1587, Jul. 1999.

[23] T. Cover and A. El Gamal, "Capacity theorems for the relay channel," *IEEE Trans. Inf. Theory*, vol. IT-25, no. 5, pp. 572–584, Sep. 1979.

[24] D. Bertsekas and R. Gallager, *Data Networks*, 2nd ed. Englewood Cliffs, NJ: Prentice-Hall, 1987.

[25] P. Nain, "Analysis of a two node ALOHA network with infinite capacity buffers," in *Proc. Int. Seminar on Computer Networking, Performance Evaluation*, Tokyo, Japan, 1985, pp. 2.2.1–2.2.16.

[26] M. Sidi and A. Segall, "Two interfering queues in packet-radio networks," *IEEE Trans. Commun.*, vol. COM-31, no. 1, pp. 123–129, Jan. 1983.

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

[28] J. Luo and A. Ephremides, "On the throughput, capacity and stability regions of random multiple access," *IEEE Trans. Inf. Theory*, vol. 52, no. 6, pp. 2593–2607, Jun. 2006.

[29] L. H. Ozaroq, S. Shamai (Shitz), and A. D. Wyner, "Information theoretic considerations for cellular mobile radio," *IEEE Trans. Veh. Technol.*, vol. 43, no. 3, pp. 359–378, May 1994.

[30] M. Kobayashi, G. Caire, and D. Gesbert, "Impact of multiple transmit antennas in a queued SDMA/TDMA downlink," in *Proc. SPAWC 2005, 6th IEEE Workshop on Signal Processing Advances in Wireless Communications*, New York City, Jun. 2005.

[31] M. K. Tsatsanis, R. Zhang, and S. Banerjee, "Network-assisted diversity for random access wireless networks," *IEEE Trans. Signal Process.*, vol. 48, no. 2, pp. 702–711, Mar. 2000.

[32] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge, U.K.: Cambridge Univ. Press.

[33] H. Zheng and C. Peng, "Collaboration and fairness in opportunistic spectrum access," in *Proc. IEEE Int. Conf. Communications (ICC)*, May 2005, pp. 3132–3136.