On the Capacity of Wireless CSMA/CA Multihop Networks

Rafael Laufer
Bell Laboratories, Alcatel-Lucent
Holmdel, NJ

Leonard Kleinrock
University of California, Los Angeles
Los Angeles, CA

***Abstract*—Due to a poor understanding of the interactions among transmitters, wireless multihop networks have commonly been stigmatized as unpredictable in nature. Even elementary questions regarding the throughput limitations of these networks cannot be answered in general. In this paper we investigate the behavior of wireless multihop networks using carrier sense multiple access with collision avoidance (CSMA/CA). Our goal is to understand how the transmissions of a particular node affect the medium access, and ultimately the throughput, of other nodes in the network. We introduce a theory which accurately models the behavior of these networks and show that, contrary to popular belief, their performance is easily predictable and can be described by a system of equations. Using the proposed theory, we provide the analytical expressions necessary to fully characterize the capacity region of *any* wireless CSMA/CA multihop network. We show that this region is nonconvex in general and entirely agnostic to the probability distributions of all network parameters, depending only on their expected values.**

I. Introduction

Wireless multihop networks have been considered a difficult modeling problem because transmissions from a particular node affect the medium access of several other nodes in an intricate way. Basically, whenever a node transmits in a CSMA/CA wireless network, any other node that overhears this transmission should remain silent and wait for it to finish before attempting to access the medium again [1]. This silence, in turn, may be interpreted by its own neighbors as an indication that the medium is idle, and thus trigger new transmissions. Due to this strong interdependence among the state of transmitters across the network, a theory which fully characterizes and predicts the behavior of wireless multihop networks has only been a vision so far.

The difficulty in creating such a theory mostly comes from (1) the distributed nature of the CSMA/CA MAC protocol itself, which dictates that transmitters should back off from each other to avoid collisions; (2) the buffer dynamics of unsaturated traffic sources, which occasionally cause queues to become empty and result in a time-varying subset of nodes contending for the channel; and (3) the dependence of downstream links on upstream traffic, which couples together the queue occupancy of all links in a multihop flow. The first issue induces some correlation among neighbor transmitters because of their physical proximity; the second and third issues, on the other hand, correlate transmitters throughout the network because of the traffic pattern. These problems are further exacerbated if the network serves several flows at the same time, and therefore transmissions of a given flow affect the medium access, as well as the throughput, of other flows. For accuracy, a throughput model must then consider both the physical proximity of transmitters and their respective interference constraints as well as the end-to-end traffic requirements of the network flows.

In this paper our goal is to propose such a model in order to understand the fundamental throughput limitations of wireless CSMA/CA multihop networks. In particular, we would like to answer specific questions regarding the network capacity. For instance, if the throughput of flow x increases by a specific amount, how much can flow y still achieve? Or even, if a new flow starts between nodes s and t, by how much must the other flows reduce their rates in order to keep the network stable? To our knowledge, even after significant research, the answers to these elementary questions are still unknown in general.

For this purpose, we develop a theory which models the behavior of a given CSMA/CA wireless multihop network and is able to accurately predict its throughput performance. It has the unique ability to model the buffer dynamics of unsaturated sources and handle multihop flows, while still respecting the interference constraints imposed by the wireless medium. The proposed theory is general and has no restrictions on the node placement, being thus suitable for arbitrary topologies. Its key feature is the ability to fully characterize the capacity region (i.e., the set of feasible input rates) of *any* wireless CSMA/CA network. We prove that this region is convex for the case where nodes are all within carrier-sense range, but nonconvex in general. We also show that the capacity region is completely *agnostic* to the probability distributions of all network parameters, such as the backoff, the transmission, and the interarrival times, depending only on their expected values.

To achieve these results, we determine the conditions under which a CSMA/CA wireless multihop network is stable and converges to a steady state. The probabilities π_S, that an independent link set S is transmitting, are well characterized through analytical expressions. We show that the problem of finding this steady-state solution can be formulated as two separate systems of equations, each with a unique solution. The first system defines the common format of the solution, and it is always linear; the second system determines the stability factors, and it is nonlinear in general. Using simulations, we show that our model is accurate.

The reminder of this paper is organized as follows. In Section II, we present the key assumptions used to derive our theory. Section III introduces our throughput model using a gradual build-up approach, starting from simpler problems and moving on to the more complex cases. In Section IV, we present our simulation results over the MIT Roofnet topology and show the accuracy of our model. Section V presents the related work, and Section VI concludes the paper.

II. System Model and Assumptions

We consider a wireless network where nodes forward traffic for each other in a multihop fashion [2]. End-to-end routes are established according to a given routing metric before the network becomes operational, and are assumed to remain fixed over time. Using these routes, each multihop flow is then allowed to traverse the network. Both the flows and their respective average rates do not change, at least for a sufficient amount of time, to allow convergence to a steady state. For ease of presentation, we assume that nodes are equipped with an omnidirectional antenna communicating in a single channel.

Nodes are assumed to have a unique transmission queue for each flow. As a packet arrives at the node, it is routed and placed in the corresponding flow queue for future transmission. Packet scheduling across the different flow queues within a node is realized with the CSMA/CA MAC protocol, as described in more detail below. Basically, each queue acts as an individual collocated transmitter, with its own backoff counter, and operates as if it was a different node altogether.

An idealized CSMA/CA MAC protocol is assumed to control the medium access among the transmitters [3]–[6]. In CSMA/CA, before transmitting a packet, each node first verifies whether the medium is idle via carrier sensing [1]. If the received power is above a given threshold, the medium is considered busy and the transmitter waits until the ongoing transmission is finished. Otherwise, the medium is considered idle. In this case, a transmitter i independently samples a random backoff interval B_i from a given continuous probability distribution, which can be different for each transmitter, and waits at least this long before transmitting. We do not require the backoff interval to be exponentially distributed as in [3]–[5], [7]. In fact, we place no assumptions whatsoever on the distribution of the backoff intervals. We assume that each queue within a node is a transmitter and has an individual backoff counter to store the remaining time until the scheduled transmission. If the medium becomes busy during the backoff interval, the transmitter freezes its counter and resumes the countdown only after the medium becomes idle again. When the counter is decremented to zero, the packet is finally transmitted.

The duration of a packet transmission is modeled as follows. Each transmission from i takes a random time T_i, which depends both on the packet size and on the bit rate. The bit rate r_i of each transmitter is assumed to be fixed, and thus the randomness of T_i comes only from the different packet sizes generated by the flow source. We do not require exponentially distributed packet sizes or independent regeneration of packets at relay nodes [3], [4], [7]. Instead, we assume that packet sizes are generated by the flow source according to a given discrete distribution, possibly different for each flow, and that packets *retain* their sizes as they traverse the network. In addition, we do not assume that all transmitters are saturated [3], [5]–[11]. Instead, packets are generated at each source i with a random interarrival time A_i, following a given probability distribution, which can also be different for each flow. As before, we place no assumption on these distributions.

During a transmission, each packet is susceptible to errors. As in previous work [3], [6], [7], [9], we assume packets are received without interference, and therefore the random noise in the wireless channel is the only source of error. This implies two important assumptions on the network model.

First, there are no hidden terminals in the network, and therefore, if two transmitters interfere at a common receiver, both are able to sense each other's transmission and back off accordingly. This is proven to occur if the carrier-sense range is large enough and if receivers can abort an ongoing reception to lock onto a new signal with a sufficiently higher power [12]. Atheros chipsets already allow this kind of preemption in the so-called restart mode, and thus the hidden terminal problem can be avoided [10].

Second, the carrier sensing is instantaneous, and thus, as soon as a transmission starts, it is immediately detected by the neighbors. This implies that both the propagation delay and the carrier-sense delay are zero. This is reasonable since nodes are usually physically close to each other and carrier sensing takes only a few microseconds in current wireless cards. With instantaneous carrier sensing, collisions due to two or more transmitters finishing their backoff intervals at the same time are also not possible, since these intervals are continuous random variables.

With these assumptions, each packet is received with a given probability p_i, the packet delivery ratio of transmitter i at the chosen bit rate r_i. If the transmission fails, the transmitter samples another backoff interval and rebroadcasts the packet as many times as necessary. This model has shown to approximate well the behavior of transmitters [3], [6], [7], [9]. Nonetheless, even if hidden interferers and collisions do exist in the network, their effect is considerably reduced in the unsaturated conditions we consider [13].

In CSMA/CA networks, several links may transmit together if they cannot hear each other's transmission. We define a set of links able to simultaneously transmit as a *feasible set*, and we use S or K to represent it throughout this paper. We assume that all feasible link sets of the network are known. With this knowledge, the *network state* is defined as the feasible set S of links which are currently transmitting. We define π_S as the probability or the fraction of time that the network is in state S (i.e., links in S are simultaneously transmitting), and thus $\sum_S \pi_S = 1$. We use $\pi_\emptyset$ to represent the fraction of time that no link is transmitting across the entire network. With a slight abuse of notation, the probability $\pi_{\{i,j\}}$ that links i and j are both transmitting is written as $\pi_{i,j}$.

At last, we let $\theta_i = E[T_i]/E[B_i]$ be the ratio between the

expected transmission time $E[T_i]$ and the expected backoff time $E[B_i]$ for transmitter i, and x_f be the average throughput of a flow f in bits per second. Table I summarizes our notation.

Notation	Definition
A_i	random variable for the interarrival time at a source i
B_i	random variable for the backoff interval of transmitter i
T_i	random variable for the transmission time of transmitter i
θ_i	ratio between $E[T_i]$ and $E[B_i]$, i.e., $\theta_i = E[T_i]/E[B_i]$
r_i	bit rate of transmitter i, in bits per second (assumed fixed)
p_i	packet delivery ratio of transmitter i (assumed fixed)
x_f	average throughput of a flow f, in bits per second
S	a set of links which may transmit at the same time
π_S	probability that all links in S are transmitting
$\pi_\emptyset$	the probability that no link is transmitting in the network
$\pi_{i,j}$	the probability that two links i and j are both transmitting

Table I
THE NOTATION USED IN OUR MODEL.

III. THROUGHPUT MODELING

In this section we describe our approach for estimating the throughput of each flow as well as the capacity region of a CSMA/CA network. First, the throughput of saturated single-hop flows is described in Section III-A. We introduce the notion of unfinished work in CSMA/CA networks, and use it to show the earlier results of Liew *et al.* [6]. Then, in Section III-B, we generalize these results for unsaturated single-hop flows, whose sources do not always have a packet to transmit. Finally, in Section III-C, we generalize our theory further to address multihop flows in an arbitrary network.

A. Saturated Single-Hop Flows

In this section we assume that nodes communicate only with their direct neighbors, and thus flows are transmitted over a single hop. This model applies, for instance, to enterprise wireless LANs (WLANs), where we have several access points deployed across a campus network.

Let the network have n links able to carrier sense each other and assume that each link is saturated with an infinite backlog. In these conditions, whenever a link is transmitting, the others freeze their backoff counter and wait until the ongoing transmission is over. Figure 1 depicts this scenario for a network of three nodes and shows the unfinished work $U_i(t)$ of each transmitter i at time t. The *unfinished work* represents the remaining time before the current node state changes, and it can be either the remaining backoff or the remaining transmission time. We know from the saturation condition that the transmitter must always be either backing off, frozen, or transmitting. For each packet, a backoff interval B_i is sampled, and the node waits at least this long before transmitting. If during this interval a neighbor starts transmitting, the node freezes its backoff counter and waits for it to finish. When the counter reaches zero, the node transmits the packet for T_i seconds, after which the cycle restarts.

There are $n+1$ states in which such a network can be. The first state is $S = \emptyset$, which occurs when nobody is transmitting; the other n states $S = \{i\}$ are when a transmitter i is active while the others are frozen. The steady-state solution then defines the probabilities $\pi_\emptyset, \pi_1, \ldots, \pi_n$ of each network state.

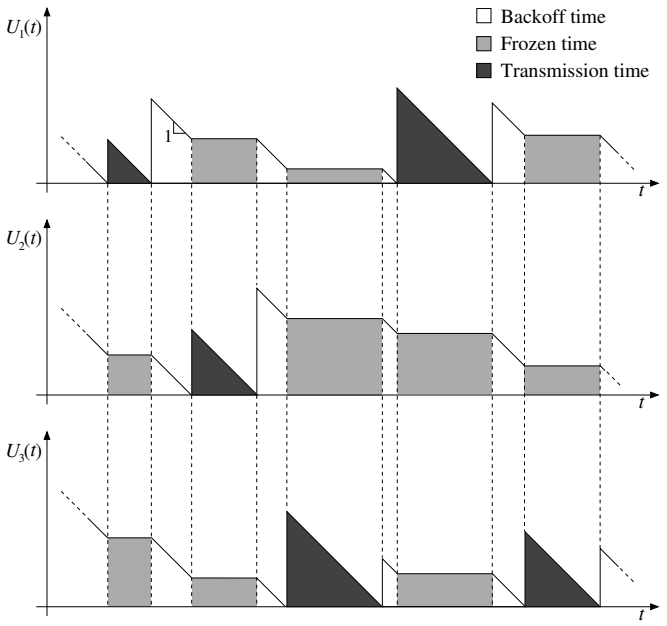


Figure 1. The operation of three saturated links within carrier-sense range. The graphs show the unfinished work $U_i(t)$ of each transmitter i at time t, which can be either the remaining backoff or the remaining transmission time.

Let $c_i(t)$ be the transmission count from node i in a large time window $[0, t)$. If within this time this node made $c_i(t)$ transmissions, then it also backed off $c_i(t)$ times, since for each transmitted packet there is a backoff interval. Realizing that each node only decreases its backoff counter when nobody is transmitting (i.e., when the network state is $S = \emptyset$), the ratio $\pi_i/\pi_\emptyset$ can be computed as

$$\frac{\pi_i}{\pi_\emptyset} = \lim_{t\to\infty} \frac{\frac{1}{t}\sum_{j=1}^{c_i(t)} T_i(j)}{\frac{1}{t}\sum_{j=1}^{c_i(t)} B_i(j)} = \lim_{t\to\infty} \frac{\frac{1}{c_i(t)}\sum_{j=1}^{c_i(t)} T_i(j)}{\frac{1}{c_i(t)}\sum_{j=1}^{c_i(t)} B_i(j)} = \frac{E[T_i]}{E[B_i]} = \theta_i, \tag{1}$$

where $B_i(j)$ and $T_i(j)$ are the duration of the j-th backoff interval and the j-th transmission of node i, respectively. We see that $\pi_i/\pi_\emptyset$ does not depend on the individual distributions of T_i and B_i, but rather only on the ratio θ_i between their expected values.

From (1), a system of linear equations can be written as

$$\pi_\emptyset = \frac{\pi_1}{\theta_1} = \frac{\pi_2}{\theta_2} = \ldots = \frac{\pi_n}{\theta_n}, \tag{2}$$

which, along with the normalizing condition $\sum_S \pi_S = 1$, can be solved to find the steady-state probabilities

$$\pi_\emptyset = \frac{1}{1+\theta_1+\theta_2+\ldots+\theta_n} \quad \pi_i = \frac{\theta_i}{1+\theta_1+\theta_2+\ldots+\theta_n}. \tag{3}$$

The average throughput x_i of a link i is then computed as

$$x_i = \left(\frac{\theta_i}{1+\theta_1+\theta_2+\ldots+\theta_n}\right) r_i p_i. \tag{4}$$

Now assume that there are still n links in the network, but *not all* links are within carrier-sense range. As a result, two or more links may transmit at the same time. To compute the steady-state probabilities in this general case, it is first necessary to establish the relation between two general states π_S and $\pi_{S\cup\{i\}}$, which differ only by a single transmitter i. A saturated CSMA/CA network has been proven to be a time-reversible Markov field in [6], and therefore detailed balance holds. In this case, the relation between any two adjacent network states, S and $S \cup \{i\}$, is shown to be

$$\pi_S = \frac{\pi_{S\cup\{i\}}}{\theta_i}. \tag{5}$$

Note that (5) generalizes the relation in (2) for the case where transmitters are not necessarily within carrier-sense range. From (5), a system of linear equations can then be written as

$$\pi_\emptyset = \frac{\pi_1}{\theta_1} = \ldots = \frac{\pi_n}{\theta_n} = \ldots = \frac{\pi_{i,j}}{\theta_i\theta_j} = \ldots = \frac{\pi_S}{\prod_{k\in S}\theta_k}, \tag{6}$$

which, along with the normalizing condition $\sum_S \pi_S = 1$, can be solved to find the steady-state probabilities

$$\pi_\emptyset = \frac{1}{\sum_K \prod_{k\in K}\theta_k} \qquad \pi_S = \frac{\prod_{k\in S}\theta_k}{\sum_K \prod_{k\in K}\theta_k}, \tag{7}$$

where the summation in the denominator is over all feasible sets K. The fraction of time a given set S transmits is thus proportional to the product of the θ_k ratios of each transmitter k within this set. From (7), the average throughput x_i of transmitter i can then be computed as $x_i = \left(\sum_{S:i\in S}\pi_S\right) r_i p_i$, where the summation is over all sets S where i is transmitting.

B. Unsaturated Single-Hop Flows

We now generalize the previous results for the case of unsaturated single-hop flows. This reflects the case where nodes communicate only with direct neighbors, but now each source does not always have a packet to transmit. As a result, the subset of nodes contending for the channel significantly changes over time, imposing an extra challenge to the model.

Let the network have n links able to carrier sense each other. Each transmitter i generates packets with a random interarrival time A_i, following a given probability distribution, and these packets are then placed into the respective flow queue for transmission. Consider the time line shown in Figure 2, where we have three links within carrier-sense range. The queue backlogs are not infinite anymore, and thus transmitters have a packet to send only part of the time. In the figure, when a new packet arrives at an empty queue, the backoff counter B_i is sampled and the countdown begins. The behavior is then similar to the saturated network, where each transmitter freezes its counter whenever a neighbor node transmits. After the counter is decremented to zero, the node transmits for T_i seconds. The time during which a transmitter could be counting down, but it is not because the queue is empty, is what we call the *idle time*. The idle time of the third transmitter is shown right below the time axis.

Given that transmitters are within carrier-sense range, the countdown only occurs when nobody is transmitting, i.e.,

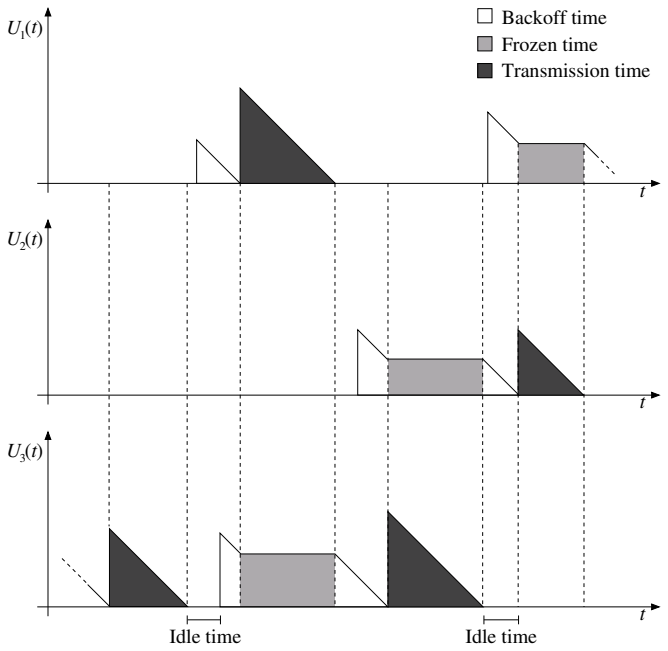


Figure 2. The operation of three unsaturated links within carrier-sense range. The graphs show the unfinished work $U_i(t)$ of each transmitter i at time t. A transmitter is active when its queue is non-empty, but remains idle otherwise.

$S = \emptyset$. However, since the sources are not saturated, each transmitter counts down only a fraction of this time. If this fraction is ρ_i for a transmitter i, such that $0 < \rho_i \leq 1$, then, noting (2) and reducing $\pi_\emptyset$ by ρ_i, we have

$$\rho_i\pi_\emptyset = \frac{\pi_i}{\theta_i}, \tag{8}$$

where $\pi_i/\theta_i = \pi_i E[B_i]/E[T_i]$ is the fraction of time that transmitter i counts down, and $\rho_i\pi_\emptyset$ reflects that i counts down only a fraction of $\pi_\emptyset$. If we define a new variable

$$\gamma_i = \rho_i\theta_i \tag{9}$$

for each transmitter, a system of equations can be written as

$$\pi_\emptyset = \frac{\pi_1}{\gamma_1} = \frac{\pi_2}{\gamma_2} = \ldots = \frac{\pi_n}{\gamma_n}, \tag{10}$$

which, along with the normalizing condition $\sum_S \pi_S = 1$, can be solved for the steady-state probabilities

$$\pi_\emptyset = \frac{1}{1+\gamma_1+\gamma_2+\ldots+\gamma_n} \quad \pi_i = \frac{\gamma_i}{1+\gamma_1+\gamma_2+\ldots+\gamma_n}. \tag{11}$$

The average throughput x_i of a transmitter i is then

$$x_i = \left(\frac{\gamma_i}{1+\gamma_1+\gamma_2+\ldots+\gamma_n}\right) r_i p_i. \tag{12}$$

One would expect the steady-state solution in (11) for unsaturated sources to be different than the solution in (3) for saturated sources. However, both are remarkably similar. The only difference is that each component θ_i is replaced with γ_i.

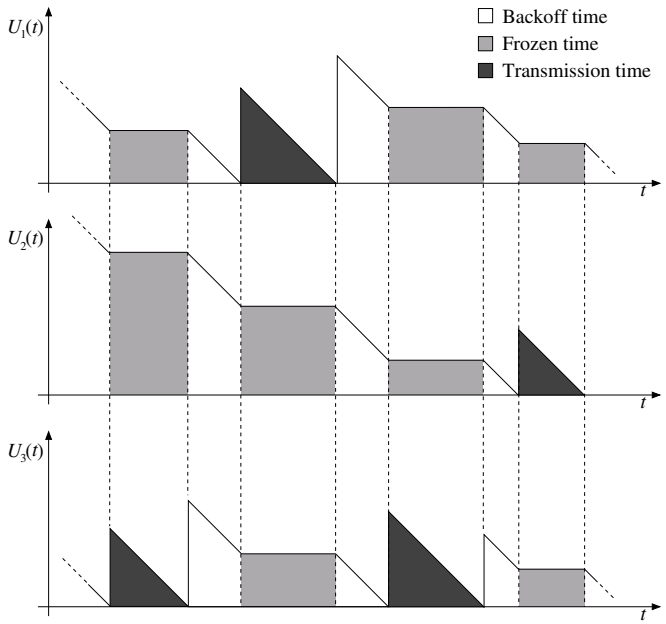


Figure 3. The *dual* saturated network for the unsaturated network depicted in Figure 2. The backoff intervals are now stretched such that transmitters have no idle time. The average backoff time increases from $E[B_i]$ to $E[B_i]/\rho_i$.

The intuition here is that (11) is also the solution of another network, with saturated sources. To see this, note that

$$\gamma_i = \rho_i \theta_i = \rho_i \left(\frac{E[T_i]}{E[B_i]} \right) = \frac{E[T_i]}{E[B_i]/\rho_i}. \tag{13}$$

Therefore, the solution in (11) is equivalent to a network where each source is saturated and has a larger average backoff time $E[B_i]/\rho_i$. This scenario is depicted in Figure 3, which shows the *dual* saturated network for the unsaturated network depicted in Figure 2. Basically, the backoff intervals are stretched such that each transmitter has no idle time. In both networks, nodes transmit during exactly the same time, and thus the steady-state solution must be the same.

Given that the steady-state solution in (11) depends only on the average values, the probability distribution of the stretched backoff interval in the dual network (cf. Figure 3) does not need to be determined. However, for the solution to be unique, the ρ_i factors must still be defined.

The ρ_i factor is similar to the utilization factor in queueing theory and is directly related to stability. To see this, note that each transmitter i is idle for a fraction $(1-\rho_i)\pi_\emptyset$ of time. If this idle time is positive, then the queue of transmitter i must frequently go to zero. This implies that the empty queue is a positive recurrent state, and therefore the queue is stable. Since we know $\pi_\emptyset > 0$ even for saturated sources, the condition for stability must be $\rho_i < 1$. As a result, if the source generates more traffic than the CSMA/CA MAC protocol can deliver, then ρ_i tends to 1, and (11) falls back to the case of saturated sources in (3). On the other hand, if the source generates too little traffic, then ρ_i tends to 0, and the network behaves almost as if i does not exist at all.

With the knowledge of the stability condition, we are able to determine the range of input rates under which the network is stable, and can therefore characterize the *capacity region*. Let $y_i = x_i/(r_i p_i)$ be the throughput x_i of a transmitter i normalized with regard to its link capacity $r_i p_i$. From (12), we know that $y_i = \pi_i$ and as a result $\pi_\emptyset = 1 - \sum_j y_j$. Since $y_i = \pi_i = \pi_\emptyset(\rho_i \theta_i)$, the ρ_i factor can be expressed as

$$\rho_i = \left(\frac{1}{\theta_i} \right) \frac{y_i}{1 - \sum_j y_j}, \tag{14}$$

and, from the stability condition $\rho_i < 1$, an inequality

$$y_i < \frac{\theta_i}{1+\theta_i} \left(1 - \sum_{j \neq i} y_j \right) \tag{15}$$

can be derived for each transmitter i. The $1 - \sum_{j \neq i} y_j$ factor in (15) is the fraction of time that i is not frozen. Within this time, i must transmit for strictly less time than what it would in the saturated case to guarantee stability.

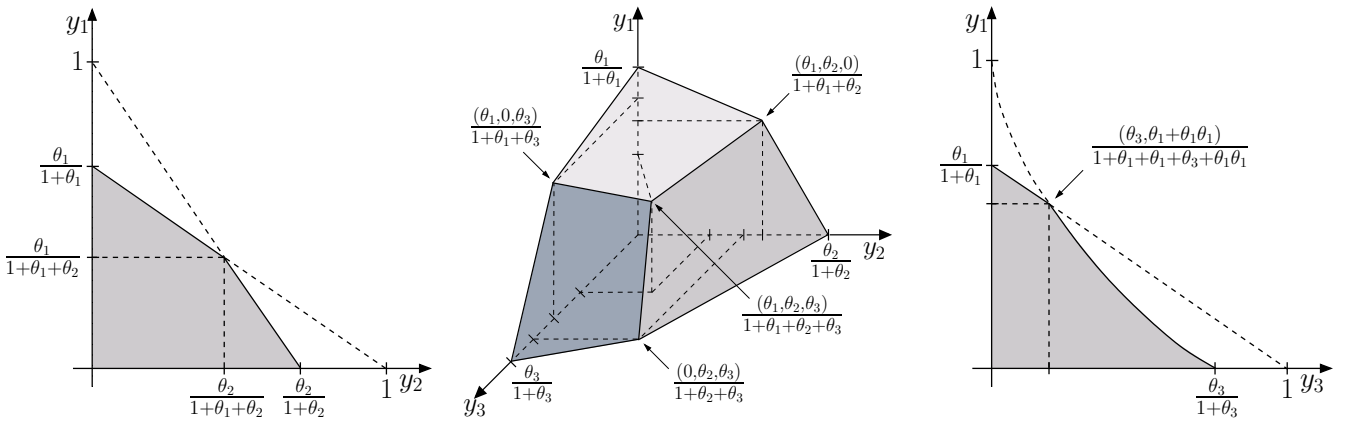
The relation among the throughputs of each transmitter is clearly linear from (15) and can be easily visualized. Figure 4 depicts the capacity region for three different scenarios. In the first scenario, we consider a simple network with only two transmitters within carrier-sense range; Figure 4(a) shows the capacity region for this case. Each transmitter imposes a linear constraint for stability and the capacity region is the intersection of both areas. The upper boundary is defined by the input rates where at least one transmitter is saturated, and the intersection point is the case where both are saturated. Figure 4(b) depicts the capacity region for the case of three links within carrier-sense range. The linear constraint from each transmitter i now represents a plane, which crosses the axis y_i at $\theta_i/(1+\theta_i)$ and the other axes y_j at 1, for $j \neq i$. In general, when transmitters are within range, the capacity region is the intersection of the half-spaces defined by (15) and by $y_i \geq 0$, and is therefore *convex*. Similar to the previous case, the upper boundary is defined by the input rates where at least one transmitter is saturated. The line segments intersecting two planes result from the rates where at least two transmitters are saturated, and the intersection point of the three planes is when all transmitters are saturated.

Under stability, we can also derive an expression for the average interarrival time $E[A_i]$ between generated packets at a transmitter i. Recall that $c_i(t)$ is the transmission count from i in a large time window $[0, t)$. From the law of large numbers, the total transmission time of this node gets closer to $c_i(t)E[T_i]$ as $t \to \infty$. Since the network is stable, the number of generated packets within this window must be $p_i c_i(t)$, and therefore $p_i c_i(t) E[A_i]$ approximates the total time t. The fraction of time y_i that a node i transmits must then be

$$y_i = \left(\frac{1}{p_i} \right) \frac{E[T_i]}{E[A_i]}. \tag{16}$$

As a result, given the parameters $E[A_i]$, $E[T_i]$, $E[B_i]$, and p_i of each transmitter, one can determine if the network is stable by checking if condition (15) holds for every transmitter.

From the particular case where transmitters are within carrier-sense range, we are now able to generalize the previous



(a) 2 transmitters within carrier-sense range. (b) 3 transmitters within carrier-sense range. (c) 3 transmitters, not all within carrier-sense range.

Figure 4. The capacity region for three different network topologies. (a) Two transmitters within carrier-sense range. (b) Three transmitters within carrier-sense range. (c) Three transmitters, not all within carrier-sense range. Transmitter τ_3 hears both τ_1 and τ_2, but τ_1 and τ_2 do not hear each other. For ease of visualization, only the cross section at $y_1 = y_2$ is showed.

results for the case where such an assumption does not hold. Note that the idea of stretching the backoff intervals to saturate the network, depicted in Figure 3, is general and can also be applied when nodes are not necessarily within range. That is, if we stretch the backoff interval of every transmitter, such that the average increases from $E[B_i]$ to $E[B_i]/\rho_i$, for some $0 < \rho_i \leq 1$, the result is a dual saturated network where nodes transmit at exactly the same time. Since the solution for a saturated network is known from (7), the steady-state probabilities for the unsaturated network can be derived as

$$\pi_\emptyset = \frac{1}{\sum_K \prod_{k\in K} \gamma_k} \qquad \pi_S = \frac{\prod_{k\in S} \gamma_k}{\sum_K \prod_{k\in K} \gamma_k}, \tag{17}$$

where $\gamma_k = \rho_k \theta_k$. The average throughput x_i of a transmitter i can then be computed as $x_i = \left(\sum_{S:i\in S} \pi_S\right) r_i p_i$ by summing over all sets S where i is transmitting.

The stability condition $\rho_i < 1$ is still the same when nodes are not all within range. Let $\mathcal{B}(S)$ be the set of transmitters allowed to back off (i.e., neither transmitting nor frozen) when the links in S are active. Each transmitter i is then idle for a fraction $(1-\rho_i)\sum_{S:i\in\mathcal{B}(S)} \pi_S$ of time, which to be positive requires $\rho_i < 1$. The expression for the ρ_i factors, however, is not the solution to a linear system anymore as in (14). Instead, we must now solve a *nonlinear* system of equations to find these factors. Each transmitter provides a variable ρ_i and an equation $y_i = \sum_{S:i\in S} \pi_S$ to the system, and it can be shown that, if the system is stable, it has a unique solution. However, one must now resort to either symbolical computation or numerical methods to find it.

For instance, consider a topology with three links, such that a transmitter τ_3 is within carrier-sense range of the two other transmitters τ_1 and τ_2, but τ_1 and τ_2 cannot hear each other. In this case, from (17) we know that

$$\pi_S = \frac{\prod_{i\in S} \rho_i \theta_i}{1 + \rho_1\theta_1 + \rho_2\theta_2 + \rho_3\theta_3 + (\rho_1\theta_1)(\rho_2\theta_2)}, \tag{18}$$

from which a system of three equations can be built

$$y_1 = \pi_1 + \pi_{1,2} \qquad y_2 = \pi_2 + \pi_{1,2} \qquad y_3 = \pi_3. \tag{19}$$

This system can be symbolically solved for ρ_1, ρ_2, and ρ_3 as

$$\begin{aligned} \rho_1 &= \left(\frac{1}{\theta_1}\right) \frac{y_1}{1-y_1-y_3} \\ \rho_2 &= \left(\frac{1}{\theta_2}\right) \frac{y_2}{1-y_2-y_3} \\ \rho_3 &= \left(\frac{1}{\theta_3}\right) \frac{y_3(1-y_3)}{(1-y_1-y_3)(1-y_2-y_3)}. \end{aligned} \tag{20}$$

Using the general result from (16) as well as the stability condition $\rho_i < 1$ for each equation in (20), we are then able to determine if the network is stable only from the parameters $E[A_i]$, $E[T_i]$, $E[B_i]$, and p_i of each transmitter. The capacity region for this case is depicted in Figure 4(c), where we only show the cross section at $y_1 = y_2$ for ease of visualization. Transmitters τ_1 and τ_2 impose linear constraints from (20), but τ_3 imposes an elliptical constraint. This region is clearly *nonconvex*, and thus the convexity of the capacity region does not necessarily hold when nodes are not all within range.

C. Multihop Flows

In the previous section we only considered one-hop flows between direct neighbor nodes. However, in real wireless networks, each flow may need to traverse multiple hops to reach the destination. In this section we generalize our earlier results for multihop networks.

Modeling wireless multihop flows has been considered a difficult problem due to the strong interdependence among links. In particular, downstream links depend on the upstream links for traffic, and therefore the queue state among them is tightly coupled. If multiple flows traverse the network, this problem is even harder, since transmissions of a given flow affect the medium access of other flows. As a result, the buffer dynamics are usually ignored at relay nodes and each link is assumed saturated, with independent packet regeneration at each relay [3], [4], [7], [8]. We, however, make no such assumptions and provide a general solution for wireless multihop flows. The link coupling naturally appears in our equations,

and thus the exact behavior of the network can be modeled without approximations.

In order to derive this model, we rely on two important observations. First, we can think about each transmitter in a flow as a traffic source. Therefore, if we use the same idea of stretching the backoff intervals of each transmitter, as depicted in Figure 3, the result is a dual saturated network where nodes transmit at exactly the same time. The theory developed in Section III-B can thus be also applied for the case of wireless multihop networks, and the steady-state probabilities are the same as in (17). Since the ratio θ_i of each transmitter is given, the only remaining question is how to find the ρ_i factors. Our second observation is that this can be addressed by assuming that the entire network is stable. In this case, the traffic received by a node must eventually be transmitted to its next hop, and no packets accumulate in the network queues.

From the stability assumption, a nonlinear system of equations can be built to find the ρ_i factors as follows. If τ_i^f is the i-th transmitter in the k-hop path used by flow f, then, in steady state, the following $k-1$ equations must hold

$$\left(\sum_{S:\tau_1^f \in S} \pi_S\right) r_1^f p_1^f = \left(\sum_{S:\tau_2^f \in S} \pi_S\right) r_2^f p_2^f = \ldots = \left(\sum_{S:\tau_k^f \in S} \pi_S\right) r_k^f p_k^f, \tag{21}$$

where r_i^f and p_i^f are the bit rate and the delivery ratio of τ_i^f, respectively. The physical interpretation of (21) is that the amount of information transmitted over each link of a stable flow must be exactly the same in steady state. We also know from (16) that, at the source τ_1^f of a flow f, we have

$$y_1^f = \sum_{S:\tau_1^f \in S} \pi_S = \left(\frac{1}{p_1^f}\right) \frac{E[T_1^f]}{E[A_1^f]}. \tag{22}$$

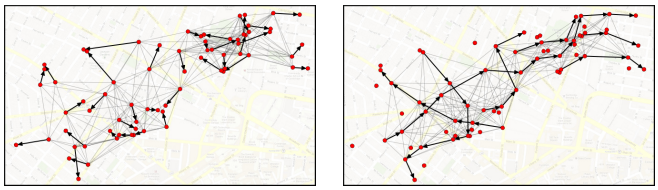
From (21) and (22), we then have k equations for each k-hop flow f traversing the network. This system of equations can be solved for the unknown variables ρ_i^f. From this solution, one can then determine if the network is stable by checking whether the condition $0 \leq \rho_i^f < 1$ holds for every transmitter. If not or if the system does not converge, the set of flow input rates is not feasible, and thus the average interarrival time $E[A_1^f]$ at the sources must be increased until the stability condition is satisfied. Under stability, the throughput of each flow f can be computed as $x_f = \left(\sum_{S:\tau_1^f \in S} \pi_S\right) r_1^f p_1^f$.

As an example, consider a simple 4-hop chain topology with a single flow and assume that the first and the last links in the chain are not within carrier-sense range. Since we only have one flow, the f notation is dropped in the following equations for convenience. From (17), the steady-state probability of a link set S in this network is

$$\pi_S = \frac{\prod_{i \in S} \rho_i \theta_i}{1 + \rho_1\theta_1 + \rho_2\theta_2 + \rho_3\theta_3 + \rho_4\theta_4 + \rho_1\theta_1\rho_4\theta_4}. \tag{23}$$

Assuming stability, the amount of information transmitted by each link of the flow must be equal, and therefore

$$(\pi_1 + \pi_{1,4})\, r_1 p_1 = (\pi_2)\, r_2 p_2 = (\pi_3)\, r_3 p_3 = (\pi_4 + \pi_{1,4})\, r_4 p_4. \tag{24}$$



(a) Single-hop topology. (b) Multihop topology.

Figure 5. The MIT Roofnet topology used in our simulations, composed of 70 nodes and 35 links with an interference range of 500 meters. Wireless links are represented by arrows and interference is represented by gray lines.

For simplicity, assume that the capacity $r_i p_i$ of each link is the same. From (22), (23), and (24), we build a nonlinear system of four equations, which can be symbolically solved for the four variables ρ_1, ρ_2, ρ_3, and ρ_4 as

$$\begin{aligned} \rho_1 &= \left(\frac{1}{\theta_1}\right) \frac{y_1}{1-3y_1} & \rho_2 &= \left(\frac{1}{\theta_2}\right) \frac{y_1(1-2y_1)}{(1-3y_1)^2} \\ \rho_3 &= \left(\frac{1}{\theta_3}\right) \frac{y_1(1-2y_1)}{(1-3y_1)^2} & \rho_4 &= \left(\frac{1}{\theta_4}\right) \frac{y_1}{1-3y_1} \end{aligned}, \tag{25}$$

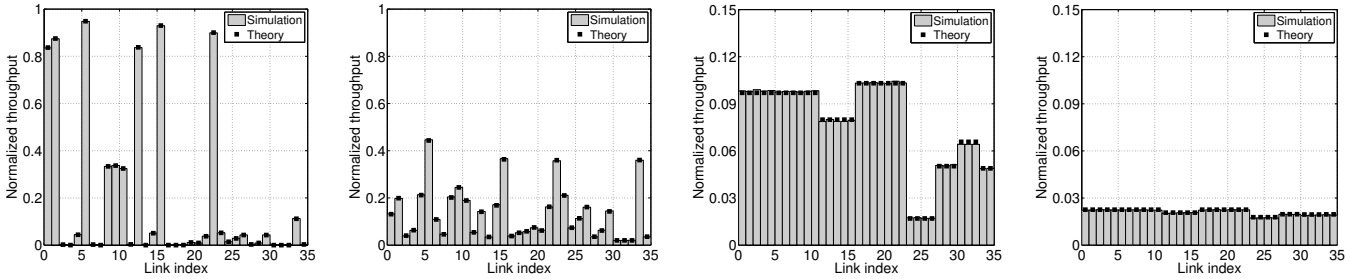
where y_1 is the ratio between $E[T_1]$ and $p_1 E[A_1]$ as in (22). From the stability condition $0 \leq \rho_i < 1$, one can then perform a binary search over the average interarrival time $E[A_1]$ to determine the lower bound required to stabilize the network.

D. Summary

In summary, the harder problem of characterizing a wireless multihop network and its capacity region through the steady-state probabilities π_S can be decomposed into two subproblems. The first problem is finding the general form of these probabilities, which is achieved by a linear system of equations whose solution is showed in (17); the second problem is finding the ρ_i^f factors, which is achieved by solving a nonlinear system of equations described by (21) and (22). In general, the solution is *exact* and holds if $0 \leq \rho_i^f < 1$ for every transmitter. It is worth noting that only the parameters $E[A_1^f]$, $E[T_i^f]$, $E[B_i^f]$, r_i^f, p_i^f are required to compute this solution. As a result, the steady-state probabilities π_S and the capacity region are completely *agnostic* to the individual probability distributions of these parameters; only the averages are relevant.

IV. Evaluation

We evaluate the accuracy of the proposed theory using network simulations over the MIT Roofnet topology, depicted in Figure 5. This topology is composed of 70 nodes spread over an area of roughly 2.5 km^2. We assume that nodes are able to carrier sense transmissions from neighbors up to 500 meters away. Each node implements the CSMA/CA MAC described in Section II, and freezes its backoff counter during any transmission within this range. The backoff interval of each transmitter is uniformly sampled from 25 to 50 μs. Packets are generated at each flow source with a uniform interarrival time and with a uniform size varying from 1,000 to 1,500 bytes. As they traverse the network, each packet retains its original size. For simplicity, the bit rates and the delivery



(a) Single-hop flows, $\rho = 1.00$ (b) Single-hop flows, $\rho = 0.01$ (c) Multihop flows, $\rho = 0.001$ (d) Multihop flows, $\rho = 0.0001$

Figure 6. The normalized throughput (i.e, the fraction of time) of each link in the network for different traffic loads. For each graph, the stability factor of the flow sources is set to the same value ρ. The vertical bars represent the simulation results and the square dots represent the predicted theoretical results.

ratios of all links are fixed at 1 Mbps and 90%, respectively. Simulations using different probability distributions, but the same average, for these parameters provided *identical* results. Due to space constraints, we only present a portion of our results.

As predicted, a perfect convergence between theoretical and simulation results always occurs, and the relative error between the two can be consistently reduced by increasing the simulation time. Our simulations ran until the average relative error $(1/n)\sum_{i=1}^{n}|y_i^s - y_i^t|/y_i^t$ became lower than 1%, where y_i^s and y_i^t are the fraction of time a node i transmits in the simulation and in the proposed theoretical model, respectively.

A. Single-Hop Flows

We first perform simulations with single-hop flows to show the accuracy of the throughput model presented in Sections III-A and III-B. The network topology for these experiments is depicted in Figure 5(a), where the 70 nodes are arranged in 35 links. In the figure, wireless links are shown using dark arrows and interfering transmitters are connected by gray lines. With an interference range of 500 meters, we have 5744 possible link sets S in this topology. In order to ensure that the input rates are feasible and within the capacity region, we fix the ρ_i factor of each transmitter i to the same value $0 < \rho < 1$. The average interarrival time $E[A_i]$ at each source is then computed from the theoretical model with (16) and (17), and used in the simulation.

Figures 6(a) and 6(b) depict the normalized throughput of each link (i.e, the fraction of time y_i that node i transmits) in the network for two values of ρ. Simulation results are shown using vertical bars, and theoretical results are shown using square dots. From these figures, we see a perfect agreement between the theoretical and simulation results. In addition, the unfairness of the CSMA/CA MAC protocol is also evident. In Figure 6(a), all sources are saturated and the unfairness is higher, with a few flows achieving a high throughput while all others starve. This occurs because a saturated network stays, most of the time, in states where the number of active links is maximum (i.e., the maximum independent sets) [6], [7], [9]. This can be seen from (7); since in practice every θ_i is large, the probability π_S of a maximum independent set S is much higher than the probability $\pi_{S'}$ of a non-maximum set S'. As a result, it is reasonable to assume $\pi_{S'} \approx 0$ and to approximate the flow throughputs using only the probabilities π_S of the maximum sets. In our topology, the maximum independent sets are composed of 7 links and there are only 3 of these sets, each with a high probability of 27%. Flows 1, 2, 6, 13, 16, and 23 are active in all of the 3 maximum sets, achieving a throughput higher than 80% in Figure 6(a). Flows 9, 10, and 11 appear once in each set, achieving roughly 33%.

The maximum independent set approximation works well for saturated networks. For unsaturated networks, however, this approximation is not valid. As ρ decreases to 0.01 in Figure 6(b), the aforementioned flows become less dominant, which results in more time available for other flows to transmit. The probability $\pi_{S'}$ of a non-maximum set S' thus becomes non-negligible, and the analytical equations in (17) must be used to accurately compute the steady-state probabilities and the throughput of each flow. From this figure, we also see that using a lower ρ provides higher fairness, although not perfect. Perfect fairness can be achieved in our scenario by setting the average interarrival time $E[A_i] = A$ of all nodes to the same value. A binary search can then be used to find the minimum value of A supported by the network. In our topology, this value is 133 ms, which achieves a normalized throughput of 8.3% per node under stability and perfect fairness.

B. Multihop Flows

We now show the accuracy of our throughput model in the presence of multihop flows, as described in Section III-C. Figure 5(b) depicts the network topology for this case, which is composed of 35 links arranged in 7 flows with up to 11 hops. Using the same 500-meter interference range, we have 4211 possible link sets in this topology. To ensure operation within the capacity region, we fix the ρ_1^f factor of each flow source to a small value $0 < \rho < 1$, and make sure the other ρ_i^f factors are within the valid range $(0, 1)$ by solving the nonlinear system of equations in (21). Once stability is guaranteed and the ρ_i^f factors are known, the interarrival time $E[A_1^f]$ at each source is computed from (22) and used in the simulation. Packets then flow from the source to the destination.

Figure 6(c) and 6(d) depict the normalized throughput of each link in the network for two values of ρ. Once again, we see an agreement between simulation and theoretical results. In addition, there are several links which present the same throughput in Figure 6(c). This is expected since these links

belong to the same flow and carry the same amount of information. From this figure, we also see that the throughput of each flow is not the same, even though the ρ_1^f factors at the sources are equal. This occurs because sources with a higher number of neighbors are frozen for a longer time, and cannot pump too much traffic into the network. Lower values of ρ reduce this unfairness, as show in Figure 6(d). However, in order to guarantee perfect fairness among flows, each source must have the same average interarrival time $E[A_1^f] = A$. As before, a binary search can be used to find the minimum value of A supported by the network. In our topology, this value is 112 ms, which achieves a normalized throughput of 9.9% per flow under stability and perfect fairness.

V. Related Work

Since their conception [1], CSMA networks have been the topic of intense research. Recently, a few models for wireless multihop networks have been studied [3]–[11], [14], [15], which can be classified into node-centric or set-centric [4].

In node-centric approaches, the throughput of each node is expressed as a function of the throughput of its interfering neighbors. Using these expressions, a system of equations is built and solved to find the individual throughputs. Examples of node-centric models include [8], [10], [11], [14], [15].

By design, node-centric approaches do not consider the global network state, which results in serious challenges to finding the overlapping transmission times. To address this, simplifying assumptions (e.g., pairwise interference) are made and elaborate techniques (e.g., inclusion-exclusion principle) must be employed, often resulting in complex models with limited insight into CSMA operation.

In contrast, set-centric approaches model the global network state using independent link sets, resulting in simple product-form expressions able to characterize the behavior of CSMA networks. In a seminal work, Boorstyn *et al.* [3] model the network as a continuous-time Markov chain whose states are the independent link sets. More recently, this model has been used in fairness studies [7], [9], showing that unfairness is mainly caused by topology inequalities, and also in extensions for 802.11 operation [4], [5].

Although providing insight into CSMA, the aforementioned works assume exponential backoff intervals, which is not realistic in 802.11. Only recently, Liew *et al.* [6] proved that the product-form solution also holds for *any* backoff distribution, a result we consider in our analysis. Different than previous work, however, we do not assume single-hop flows, saturated sources, exponential packet lengths, or independent packet regeneration at relays [3]–[11]. As a result, our model is more general and applies to arbitrary wireless CSMA/CA networks. In addition, a full characterization of the capacity region of these networks is still missing, and we believe that the equations and methodology introduced in this paper are the first attempt to do so.

VI. Conclusions

Even after significant research, the throughput limitations of wireless CSMA/CA multihop networks are still poorly understood in general. In this paper we introduce a theory able to not only predict the behavior of these networks, but also fully characterize their capacity region using analytical expressions. As a result, fundamental tradeoffs between the input rates of the various traffic sources can now be analyzed. Our theory has no restrictions on the node placement and can be applied to *any* CSMA/CA network, providing support for unsaturated sources, multihop flows, and arbitrary probability distributions for the packet size, backoff, and interarrival times. We show that the capacity region is entirely *agnostic* to the distributions of these parameters, depending only on their average values. The proposed theory respects the interference constraints among nodes and incorporates the buffer dynamics of unsaturated sources and multihop flows. Now, with knowledge of the underlying process governing CSMA networks, we believe additional insights into network design and performance can be obtained. In particular, we believe that our results open up new directions in areas such as routing optimization, network provisioning, and fairness.

References

[1] L. Kleinrock and F. Tobagi, "Packet Switching in Radio Channels: Part I - Carrier Sense Multiple-Access Modes and Their Throughput-Delay Characteristics," *IEEE Trans. on Communications*, vol. COM-23, no. 12, pp. 1400–1416, Dec. 1975.

[2] M. Campista, L. H. Costa, and O. C. Duarte, "A Routing Protocol Suitable for Backhaul Access in Wireless Mesh Networks," *Computer Networks*, vol. 56, no. 2, pp. 703–718, Feb. 2012.

[3] R. Boorstyn, A. Kershenbaum, B. Maglaris, and V. Sahin, "Throughput Analysis in Multihop CSMA Packet Radio Networks," *IEEE Trans. on Communications*, vol. 35, no. 3, pp. 267–274, Mar. 1987.

[4] M. Garetto, T. Salonidis, and E. W. Knightly, "Modeling Per-Flow Throughput and Capturing Starvation in CSMA Multi-Hop Wireless Networks," *IEEE Trans. on Networking*, vol. 16, no. 4, pp. 864–877, Aug. 2008.

[5] B. Nardelli and E. Knightly, "Closed-form Throughput Expressions for CSMA Networks with Collisions and Hidden Terminals," in *Proc. IEEE INFOCOM*, Mar. 2012.

[6] S. Liew, C. Kai, H. Leung, and P. Wong, "Back-of-the-Envelope Computation of Throughput Distributions in CSMA Wireless Networks," *IEEE Trans. on Mobile Computing*, vol. 9, no. 9, pp. 1319–1331, Sep. 2010.

[7] X. Wang and K. Kar, "Throughput Modelling and Fairness Issues In CSMA/CA Based Ad-Hoc Networks," in *Proc. IEEE INFOCOM*, Mar. 2005, pp. 23–34.

[8] Y. Gao, D.-M. Chiu, and J. C. S. Lui, "Determining the End-to-End Throughput Capacity in Multi-Hop Networks: Methodology and Applications," in *Proc. ACM SIGMETRICS*, Jun. 2006.

[9] M. Durvy, O. Dousse, and P. Thiran, "Self-Organization Properties of CSMA/CA Systems and Their Consequences on Fairness," *IEEE Trans. on Information Theory*, vol. 55, no. 3, pp. 931–943, Mar. 2009.

[10] P. C. Ng and S. C. Liew, "Throughput Analysis of IEEE802.11 Multi-Hop Ad Hoc Networks," *IEEE Trans. on Networking*, vol. 15, no. 2, pp. 309–322, Apr. 2007.

[11] M. M. Carvalho and J. J. Garcia-Luna-Aceves, "A Scalable Model for Channel Access Protocols in Multihop Ad Hoc Networks," in *Proc. ACM MobiCom*, Sep. 2004, pp. 330–344.

[12] L. Jiang and S. C. Liew, "Removing Hidden Nodes in IEEE 802.11 Wireless Networks," in *Proc. IEEE VTC*, Sep. 2005, pp. 1127–1131.

[13] H. Zhai, X. Chen, and Y. Fang, "How Well Can the IEEE 802.11 Wireless LAN Support Quality of Service?" *IEEE Trans. on Wireless Communications*, vol. 4, no. 6, pp. 3084–3094, Nov. 2005.

[14] K. Medepalli and F. A. Tobagi, "Towards Performance Modeling of IEEE 802.11 based Wireless Networks: A Unified Framework and its Applications," in *Proc. IEEE INFOCOM*, Apr. 2006.

[15] A. Jindal and K. Psounis, "The Achievable Rate Region of 802.11-Scheduled Multihop Networks," *IEEE Trans. on Networking*, vol. 17, no. 4, pp. 1118–1131, Aug. 2009.