The Capacity of Wireless CSMA/CA Networks

Rafael Laufer, *Member, IEEE, ACM*, and Leonard Kleinrock, *Life Fellow, IEEE, Fellow, ACM*

***Abstract*—Due to a poor understanding of the interactions among transmitters, wireless networks using carrier sense multiple access with collision avoidance (CSMA/CA) have been commonly 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 CSMA/CA networks 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 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 network. We show that this region is nonconvex in general and agnostic to the probability distributions of all network parameters, depending only on their expected values. Our theory is also shown to extend naturally to time division multiple access (TDMA) networks and to predict how the network responds to infeasible input rates.**

***Index Terms*—Capacity, CSMA/CA, wireless networks.**

I. Introduction

Wireless CSMA/CA 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 wireless CSMA/CA 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 CSMA/CA networks has only been a vision so far.

The difficulty in creating such a theory mainly stems from: 1) the distributed nature of the CSMA/CA protocol itself, which dictates that transmitters should back off from each other to avoid collisions; 2) the limited radio range of nodes, which creates different broadcast domains whose behaviors are interdependent; and 3) 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. The first issue induces some correlation among neighbor transmitters because of their physical proximity; the second and third issues correlate transmitters throughout the network because of the traffic pattern. For accuracy, a throughput model must then consider both the proximity of transmitters, with their respective interference constraints, and the 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 networks. In particular, we answer specific questions regarding the network capacity. For instance, if the throughput of flow f_1 increases by 10%, how much can an interfering flow f_2 still achieve? Or even, if a new flow f_3 starts, by how much must f_1 and f_2 reduce their rates to keep the network stable? To the best of our knowledge, even after decades of research, the answers to these quite elementary questions are still unknown in general.

To address this, we develop a theory which models the behavior of wireless CSMA/CA networks and also predicts their throughput performance. It has the unique ability to model the buffer dynamics of unsaturated sources, while still respecting the interference constraints imposed by the wireless medium. Our 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 wireless CSMA/CA 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 these steady-state probabilities 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. Finally, our theory is also shown to extend naturally to TDMA networks, and to predict how the network responds to infeasible input rates.

The remainder of this paper is organized as follows. In Section II, we present the key assumptions used to derive the proposed theory. Section III presents our throughput model, and Section IV discusses network stability. We show how the capacity region can be characterized in Section V and how to predict the network behavior under infeasible input rates in

Manuscript received April 23, 2014; revised November 07, 2014, February 09, 2015; accepted March 13, 2015; approved by IEEE Transactions on Networking Editor X. Lin. Date of publication June 10, 2015; date of current version June 14, 2016. A preliminary version of this paper was published in the Proceedings of the IEEE Conference on Computer Communications (INFOCOM), Turin, Italy, April 14–19, 2013.

R. Laufer is with Bell Laboratories, Alcatel-Lucent, Holmdel, NJ 07733 USA (e-mail: rafael.laufer@alcatel-lucent.com).

L. Kleinrock is with the Computer Science Department, University of California, Los Angeles, CA 90095 USA (e-mail: lk@cs.ucla.edu).

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

Digital Object Identifier 10.1109/TNET.2015.2415465

Section VI. In Section VII, we present simulations to demonstrate our theoretical results. Section VIII presents the related work, and Section IX concludes the paper.

II. System Model and Assumptions

We consider a wireless network where nodes are not all within range of each other. Single-hop flows are assumed, with each node transmitting traffic to a given neighbor. Both the flows and their average input rates do not change, at least for a sufficient amount of time, to allow convergence to a steady state. For ease of presentation, nodes are assumed to communicate in a single radio channel and to have a unique transmission queue for each flow. Packet scheduling across the different queues within a node is realized with the CSMA/CA MAC protocol, as described 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 [2]–[9]. In CSMA/CA, before sending a packet, each transmitter τ_i 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 τ_i waits for the ongoing transmission to finish. Otherwise, the medium is considered idle and τ_i samples a random backoff interval B_i from a given continuous probability distribution (possibly different for each node) and waits at least this long before transmitting. The backoff interval is not required to be exponentially distributed as in [3]–[9]. In fact, we place no assumptions on its distribution. Each queue within a node is considered a separate transmitter with an individual backoff counter to store the remaining time until the scheduled transmission. If the medium becomes busy during the backoff interval, then τ_i freezes its counter and resumes the countdown only after the medium is 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 τ_i takes a random time T_i, depending both on the packet size and on the bit rate. The bit rate r_i of each transmitter τ_i is assumed fixed, and thus the randomness of T_i comes only from the different packet sizes generated by the flow source. We do not require packet sizes to be exponentially distributed [3]–[8]; instead, packets are generated according to a given discrete probability distribution of sizes (possibly different for each node). Transmitters are also not necessarily saturated [2], [5]–[11]. In fact, packets are generated at each node following an exogenous arrival process. After a newly arrived packet, each transmitter τ_i samples an interarrival interval A_i from a given continuous probability distribution (possibly different for each node). A counter is used to store the remaining time until the next arrival and, similar to the backoff case, this counter also freezes when the medium becomes busy. After the arrival counter is decremented to zero, a new packet is placed into the queue, and the process repeats. No assumption is made on the interarrival time distributions[1].

TABLE I
The Notation Used in Our Model

Notation	Definition
A_i	random variable for the interarrival time of transmitter τ_i
B_i	random variable for the backoff interval of transmitter τ_i
T_i	random variable for the transmission time of transmitter τ_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 τ_i, in bits per second (assumed fixed)
p_i	average packet delivery ratio of transmitter τ_i (assumed fixed)
ρ_i	stability factor of transmitter τ_i
S or K	set of links which may transmit at the same time
π_S	probability that all links in S are transmitting
$\pi_\emptyset$	probability that no link is transmitting in the network
$\pi_{i,j}$	probability that both τ_i and τ_j are transmitting
λ_i	fraction of time that τ_i transmits, i.e., $\lambda_i = \sum_{S:i\in S} \pi_S$

During transmissions, packets are susceptible to reception errors. As in previous work [2], [3], [5], [6], [8], packets are assumed to be received without interference, with the random noise and fading in the wireless channel being the only error sources. This imposes two 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 sufficiently large and if receivers can abort an ongoing reception to lock onto a new signal with 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 [11].

Second, the carrier sensing is instantaneous, and thus, as soon as a transmission starts, it is immediately detected by 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 transmitters finishing their backoff intervals at the same time are not possible, since these intervals are continuous random variables.

With these assumptions, each packet transmitted by τ_i is received with a probability p_i, the packet delivery ratio at the chosen bit rate r_i. If the transmission fails, the transmitter samples another backoff interval and rebroadcasts the same packet as many times as necessary. This model is known to approximate the behavior of transmitters well [2], [5], [6]. Nonetheless, even if hidden interferers and collisions do exist in the network, their effect is considerably reduced in the unsaturated conditions considered in this work.

In CSMA/CA networks, several links may transmit together if their transmitters cannot hear each other. 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 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 both τ_i and τ_j are 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 interval $E[B_i]$ of τ_i. Table I summarizes our notation.

[1] Freezing the arrival process is required for our model to be analytically tractable and still allow arbitrary interarrival time distributions. This constraint is missing in our preliminary conference paper, but it is required for the product-form solution in (11) to hold.

III. Throughput Modeling

In this section we describe our approach for calculating the throughput of each transmitter in a CSMA/CA network. First, the case of saturated transmitters is described in Section III-A. We introduce the notion of unfinished work in CSMA/CA networks, and use it to show the results of Liew *et al.* [2]. Then, in Section III-B, we generalize these results for unsaturated transmitters. In this case, sources do not always have a packet to transmit, resulting in a time-varying subset of nodes contending for the channel.

A. Saturated Transmitters

Let the network have n transmitters able to carrier sense each other and assume that each transmitter is saturated with an infinite backlog. In these conditions, whenever someone is transmitting, the others freeze their backoff counter and wait for the ongoing transmission to finish. Fig. 1 depicts this scenario for a network of three nodes and shows the unfinished work $U_i(t)$ of each transmitter τ_i at time t. The *unfinished work* represents the remaining time before the state of τ_i changes, and it can be either the remaining backoff or the remaining transmission time. We know from the saturation condition that τ_i must always be either backing off, frozen, or transmitting. For each packet, a backoff interval B_i is sampled, and τ_i waits at least this long before transmitting. If during this interval a neighbor starts transmitting, then τ_i freezes its backoff counter and waits for the neighbor to finish. When the counter reaches zero, τ_i 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 τ_i is active while the others are frozen. The steady-state solution $\boldsymbol{\pi}$ then defines the probabilities $\pi_\emptyset, \pi_1, \ldots, \pi_n$ of each state.

Let $c_i(t)$ be the transmission count from node τ_i in a large time window $[0,t)$. If, within this time, τ_i completed $c_i(t)$ 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 jth backoff interval and the jth transmission of τ_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

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

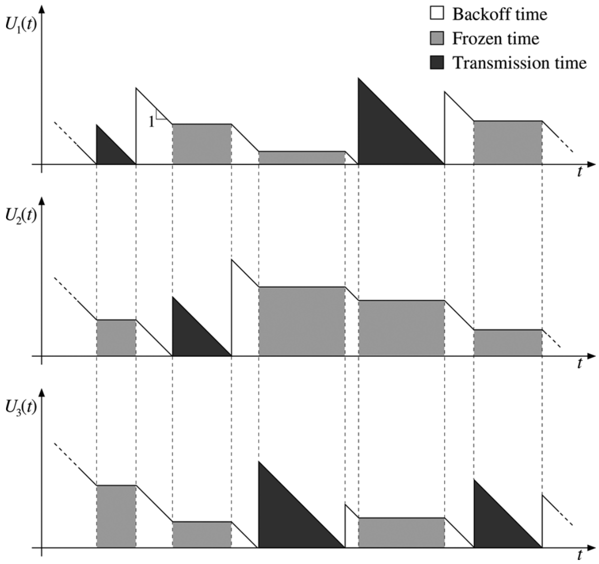


Fig. 1. The operation of three saturated links within carrier-sense range. The graphs show the unfinished work $U_i(t)$ of each transmitter τ_i at time t, which can be either the remaining backoff or the remaining transmission time.

which, along with the normalizing condition $\sum_S \pi_S=1$, can be solved to find the steady-state solution $\boldsymbol{\pi}$ as

$$\pi_\emptyset=\frac{1}{1+\theta_1+\cdots+\theta_n} \qquad \pi_i=\frac{\theta_i}{1+\theta_1+\cdots+\theta_n}. \tag{3}$$

The throughput of τ_i is then $\pi_i r_i p_i$.

Now assume that there are still n links in the network, but *not all* transmitters are within carrier-sense range. As a result, two or more links may transmit at the same time. A saturated CSMA/CA network is proven to be a Markov random field in [2], and the relation between any two adjacent network states, S and $S\cup\{i\}$, is shown to be

$$\pi_S=\frac{\pi_{S\cup\{i\}}}{\theta_i} \tag{4}$$

with an equivalent result also being achieved in [3]. Note that (4) generalizes the relation in (2) for the case where transmitters are not necessarily within carrier-sense range. From (4), a system of linear equations can then be written as

$$\pi_\emptyset=\frac{\pi_1}{\theta_1}=\ldots=\frac{\pi_n}{\theta_n}=\ldots=\frac{\pi_{i,j}}{\theta_i\theta_j}=\ldots=\frac{\pi_S}{\prod_{k\in S}\theta_k} \tag{5}$$

which, along with the normalizing condition $\sum_S \pi_S=1$, can be solved to find the steady-state solution $\boldsymbol{\pi}$ as

$$\pi_\emptyset=\frac{1}{\sum_K\prod_{k\in K}\theta_k} \qquad \pi_S=\frac{\prod_{i\in S}\theta_i}{\sum_K\prod_{k\in K}\theta_k} \tag{6}$$

where the summation in the denominator is over all feasible sets K. The throughput of τ_i can then be computed as $(\sum_{S:i\in S}\pi_S)r_i p_i$, where the summation is over all sets S where τ_i transmits.

B. Unsaturated Transmitters

The previous results are now generalized for unsaturated transmitters. In this case, after a newly arrived packet, each transmitter τ_i samples a random interarrival interval A_i for the next arrival. A countdown then begins, with the arrival counter freezing whenever the medium becomes busy. Once the counter reaches zero, a new packet (with a random size) is placed at the back of the transmission queue.

Let the network have n links able to carrier sense each other. Consider the timeline shown in Fig. 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. When a new packet arrives at the empty queue of τ_i, a 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 the network is idle, i.e., $S = \emptyset$. 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 τ_i, such that $0 \le \rho_i < 1$, then, noting (2) and reducing $\pi_\emptyset$ by ρ_i, results in

$$\rho_i \pi_\emptyset = \frac{\pi_i}{\theta_i} \tag{7}$$

where $\pi_i/\theta_i = \pi_i E[B_i]/E[T_i]$ is the fraction of time that transmitter τ_i counts down, and $\rho_i\pi_\emptyset$ reflects that τ_i counts down only a fraction of $\pi_\emptyset$. A system of linear equations can then be written as

$$\pi_\emptyset = \frac{\pi_1}{\rho_1\theta_1} = \frac{\pi_2}{\rho_2\theta_2} = \ldots = \frac{\pi_n}{\rho_n\theta_n} \tag{8}$$

which, along with the normalizing condition $\sum_S \pi_S = 1$, is solved with the following steady-state probabilities:

$$\pi_\emptyset = \frac{1}{1+\rho_1\theta_1+\ldots+\rho_n\theta_n} \qquad \pi_i = \frac{\rho_i\theta_i}{1+\rho_1\theta_1+\ldots+\rho_n\theta_n}. \tag{9}$$

The throughput of τ_i is then $\pi_i r_i p_i$.

One would expect the solution in (9) 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 $\rho_i\theta_i$. The intuition here is that (9) is also the solution of another network, with saturated sources. To see this, note that

$$\rho_i\theta_i = \rho_i\left(\frac{E[T_i]}{E[B_i]}\right) = \frac{E[T_i]}{E[B_i]/\rho_i}. \tag{10}$$

Therefore, the solution in (9) 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 Fig. 3, which shows the *dual* saturated network for the unsaturated network of Fig. 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.

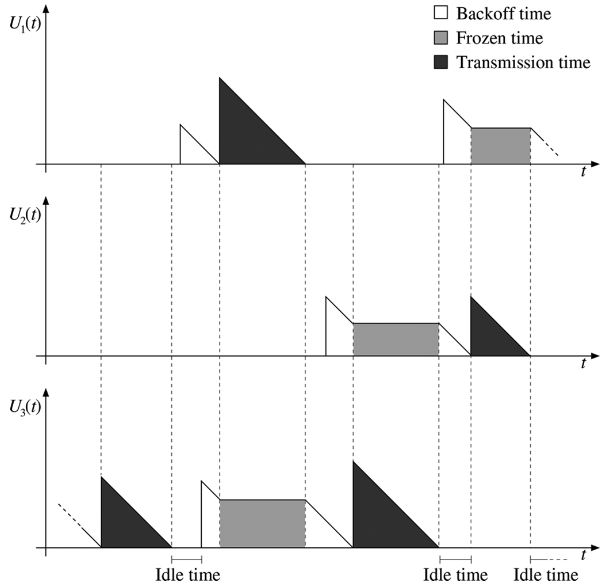


Fig. 2. The operation of three unsaturated links within carrier-sense range. The graphs show the unfinished work $U_i(t)$ of each transmitter τ_i at time t. A transmitter is active when its queue is non-empty, but remains idle otherwise.

The idea of stretching the backoff intervals to saturate the network can also be applied when nodes are not necessarily within range. By stretching the backoff interval of every transmitter, such that the average increases from $E[B_i]$ to $E[B_i]/\rho_i$, for some $0 \le \rho_i < 1$, the result is a dual saturated network where nodes transmit at exactly the same time. The steady-state probability $\boldsymbol{\pi}$ for the unsaturated network must be therefore similar to (6). However, when nodes are not all within range, the arrival process must freeze during neighbor transmissions for $\boldsymbol{\pi}$ to have a product form. By doing so, each node τ_i behaves as an independent $GI/G/1$ queue while unfrozen, and ρ_i does not depend on the network state S, as shown in the following theorem. The proof is in the Appendix.

Theorem 1: By freezing the arrival process, the probability π_S that a link set S is active in an unsaturated network is

$$\pi_S = \frac{\prod_{i\in S}\rho_i\theta_i}{\sum_K \prod_{k\in K}\rho_k\theta_k}. \tag{11}$$

The throughput of τ_i is then $(\sum_{S:i\in S}\pi_S) r_i p_i$, where the summation is over all sets S that τ_i transmits.

Given that the steady-state solution $\boldsymbol{\pi}$ depends only on the average values $E[B_i]/\rho_i$ of each transmitter τ_i, the probability distribution of the stretched backoff interval in the dual saturated network (cf. Fig. 3) does not need to be determined. However, the $\boldsymbol{\rho} = (\rho_1, \rho_2, \ldots, \rho_n)$ vector must still be found to characterize $\boldsymbol{\pi}$. We defer the expression of $\boldsymbol{\rho}$ to Section VI and instead discuss its relation to stability in the next section.

IV. Network Stability

For a wireless CSMA/CA network to be stable, two conditions must hold: 1) link set stability, i.e., the steady-state solution $\boldsymbol{\pi}$ must exist, and 2) queue stability, i.e., queues must not grow without limitation.

Link set stability is guaranteed to occur in any wireless CSMA/CA network, even if queues have infinite backlogs.

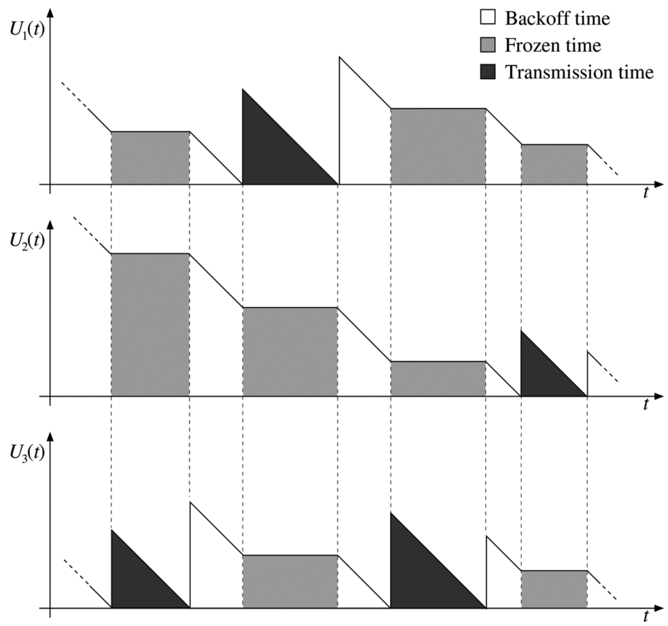


Fig. 3. The *dual* saturated network for the unsaturated network depicted in Fig. 2. The backoff intervals are now stretched such that transmitters have no idle time. The average backoff time increases from $E[B_i]$ to $E[B_i]/\rho_i$.

This can be easily seen if we realize that $\boldsymbol{\pi}$ in (11) is also the solution of a finite irreducible Markov process where the detailed balance equation $\pi_S = \pi_{S\cup\{i\}}/(\rho_i\theta_i)$ holds for any two adjacent states S and $S \cup \{i\}$. As a result, the steady-state solution $\boldsymbol{\pi}$ always exists and is unique.

Queue stability, on the other hand, is not always guaranteed. In fact, only under certain conditions are the transmission queues stable. We now discuss these conditions and extend the stability concept to notions of strong and weak stability.

A. Strong Stability

To better understand queue stability, one must first realize the central role played by the idle time (cf. Fig. 2). If a $GI/G/1$ queue is idle for a strictly positive fraction of time, then its arrival rate must be strictly lower than its service rate, guaranteeing stability. In our case, if $\rho_i < 1$ for a given transmitter τ_i, then $E[B_i]/\rho_i$ is strictly larger than $E[B_i]$, implying that the node must be idle for some time. Since the ρ factors are non-negative, queue stability then occurs if $0 \le \rho_i < 1$ for each transmitter τ_i, or equivalently, if the $\boldsymbol{\rho} = (\rho_1, \rho_2, \ldots, \rho_n)$ vector is bounded as $\mathbf{0} \preceq \boldsymbol{\rho} \prec \mathbf{1}$, with the curled symbols $\preceq$ and $\prec$ denoting componentwise inequalities. We refer to this as the *strong stability* condition.

Each ρ factor can thus be thought of as the utilization factor in queueing theory, and therefore as an indicator of how close to saturation a given transmitter is. If a transmitter τ_i generates or receives more traffic than its CSMA/CA MAC layer is able to deliver, then ρ_i tends to 1 and, if this occurs for all nodes, then (11) falls back to the case of saturated transmitters in (6). On the other hand, if τ_i generates too little traffic, then ρ_i tends to 0, and the network behaves almost as if τ_i does not exist at all.

B. Weak Stability

Consider now a given steady-state solution $\boldsymbol{\pi}$ where a factor $\rho_i > 1$ violates the strong stability condition. If $E[B_i]$ is assumed fixed, then we know that $\boldsymbol{\pi}$ can never be realized in practice. However, if this assumption is relaxed, then a possible interpretation for this case is that $\boldsymbol{\pi}$ is feasible as long as the average backoff interval $E[B_i]$ is reduced by a factor of $1/\rho_i$. In this case, for $\rho_i > 1$, the interval $E[B_i]/\rho_i$ becomes shorter and satisfies the steady-state solution $\boldsymbol{\pi}$.

While this reduction ensures that $\boldsymbol{\pi}$ is feasible, it does not guarantee strong stability. Nonetheless, the network becomes strongly stable if a strictly shorter average backoff interval is selected. In particular, if we choose $\rho_i'(E[B_i]/\rho_i)$ for any $0 \le \rho_i' < 1$, then the same steady-state distribution $\boldsymbol{\pi}$ is achieved and strong stability is also guaranteed. Therefore, any feasible wireless CSMA/CA network has a dual network which is strongly stable as long as $\boldsymbol{\rho} \succeq \mathbf{0}$. We refer to this as the *weak stability* condition, since it only holds when the average backoff interval is allowed to be reduced. Clearly, strong stability implies weak stability, but not vice versa.

V. Capacity Region Characterization

With knowledge of the stability condition, it is possible to determine the range of input rates under which the network is stable. Here, we describe the key application for the theory developed in the previous sections and show how it can be used to characterize the capacity region of wireless CSMA/CA networks.

A. All Transmitters Within Range

Consider at first a network where all nodes are within range and let λ_i be the fraction of time that τ_i transmits. In this case, $\lambda_i = \pi_i$ and $\pi_\emptyset = 1 - \sum_j \lambda_j$. Since $\lambda_i = \pi_i = \pi_\emptyset \times (\rho_i\theta_i)$, the ρ_i factor can be expressed as

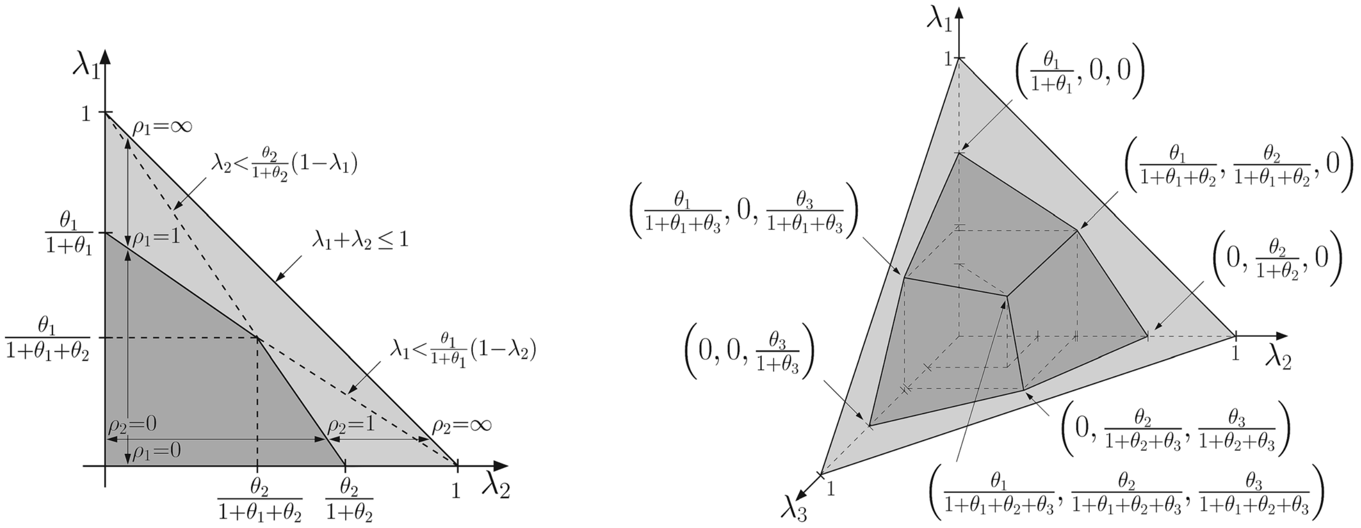
$$\rho_i = \left(\frac{1}{\theta_i}\right)\frac{\lambda_i}{1 - \sum_j \lambda_j}. \tag{12}$$

Applying the weak stability condition $\rho_i \ge 0$ to (12), each λ_i must be non-negative, and the sum of the normalized throughputs must also respect the constraint $\sum_j \lambda_j \le 1$. From the strong stability condition $\rho_i < 1$, the strict inequality

$$\lambda_i < \frac{\theta_i}{1+\theta_i}\left(1 - \sum_{j \ne i} \lambda_j\right) \tag{13}$$

must hold for each transmitter τ_i. Intuitively, the $1 - \sum_{j\ne i}\lambda_j$ factor in (13) is the fraction of time that τ_i is not frozen. Within this time, τ_i must transmit for strictly less time than what it would in the saturated case to guarantee strong stability.

The relation among the throughputs of each transmitter is clearly linear from (13) and therefore can be easily visualized. Fig. 4 depicts the capacity region for two fundamental scenarios. In the first one, we consider a simple network with only two transmitters within carrier-sense range; Fig. 4(a) shows the capacity region for this case. From weak stability, both non-negativity and the linear constraint $\lambda_1 + \lambda_2 \le 1$ must hold, creating the outer capacity region depicted in light gray. From the strong



λ_1
1
$\rho_1=\infty$
$\lambda_2<\frac{\theta_2}{1+\theta_2}(1-\lambda_1)$
$\frac{\theta_1}{1+\theta_1}$
$\rho_1=1$
$\lambda_1+\lambda_2\leq 1$
$\frac{\theta_1}{1+\theta_1+\theta_2}$
$\lambda_1<\frac{\theta_1}{1+\theta_1}(1-\lambda_2)$
$\rho_2=0$
$\rho_2=1$
$\rho_2=\infty$
$\rho_1=0$
$\frac{\theta_2}{1+\theta_1+\theta_2}$
$\frac{\theta_2}{1+\theta_2}$
1
λ_2
λ_1
1
$\left(\frac{\theta_1}{1+\theta_1},0,0\right)$
$\left(\frac{\theta_1}{1+\theta_1+\theta_2},\frac{\theta_2}{1+\theta_1+\theta_2},0\right)$
$\left(\frac{\theta_1}{1+\theta_1+\theta_3},0,\frac{\theta_3}{1+\theta_1+\theta_3}\right)$
$\left(0,\frac{\theta_2}{1+\theta_2},0\right)$
$\left(0,0,\frac{\theta_3}{1+\theta_3}\right)$
1
λ_2
$\left(0,\frac{\theta_2}{1+\theta_2+\theta_3},\frac{\theta_3}{1+\theta_2+\theta_3}\right)$
1
λ_3
$\left(\frac{\theta_1}{1+\theta_1+\theta_2+\theta_3},\frac{\theta_2}{1+\theta_1+\theta_2+\theta_3},\frac{\theta_3}{1+\theta_1+\theta_2+\theta_3}\right)$

(a) Two transmitters within carrier-sense range.

(b) Three transmitters within carrier-sense range.

Fig. 4. The capacity region for two different topologies. (a) Two transmitters within carrier-sense range. (b) Three transmitters within carrier-sense range. In both figures, the inner region in dark gray is derived under strong stability, and the outer region in light gray is derived under weak stability.

stability condition in (13), each transmitter imposes a linear constraint, and the inner capacity region shown in dark gray is the area which satisfies both constraints. Its upper boundary is defined by the input rates where at least one transmitter is saturated, and the intersection point is the case where both transmitters are saturated.

Interestingly, the region in light gray for weak stability corresponds to the capacity region of a TDMA network. This is always true, even when nodes are not all within range, since the condition $\boldsymbol{\rho} \succeq \mathbf{0}$ includes the case where $\boldsymbol{\rho} \to \boldsymbol{\infty}$. In this scenario, each node has an infinitesimal backoff interval, and thus it spends most of its time either frozen or transmitting. This corresponds to an ideal TDMA scheme which is able to perfectly schedule all transmissions across the network without any control overhead. As soon as a transmission ends, another one begins almost immediately, and the network state is always a maximal independent set. Different than traditional TDMA networks, however, in this case transmissions are not restricted to start at fixed time slots.

From this insight, the gap between the two regions in Fig. 4(a) is then the capacity toll paid due to the adoption of a distributed CSMA/CA coordination (dark gray region), as opposed to an ideal fully centralized scheduler using TDMA (light gray region). This gap is typically small in practical networks, but it can be further reduced by either shrinking the average backoff intervals (i.e., increasing θ_i), or equivalently, by increasing the ρ factors, as shown in the figure.

Fig. 4(b) depicts the capacity region for the case of three links within carrier-sense range. Weak stability constraints define the outer tetrahedral region in light gray. The linear constraint in (13) for each transmitter τ_i now represents a plane, which crosses the axis λ_i at $\theta_i/(1+\theta_i)$ and the other axes λ_j at 1, for $j \neq i$, resulting in the inner dark gray region. Similar to the previous case, its upper boundary is defined by the input rates where at least one transmitter is saturated. The line segments intersecting the planes represent two saturated transmitters, and the intersection point represents the case where all three transmitters are saturated. In general, whenever nodes are all within carrier-sense range, both the inner and the outer capacity regions are defined by the intersection of several half-spaces, and are therefore *convex*.

B. Not All Transmitters Within Range

Consider now the case where not all transmitters are within range. Let $\boldsymbol{\lambda} = (\lambda_1, \lambda_2, \dots, \lambda_n)$ be the throughput vector normalized with regard to link capacity, i.e., $\lambda_i = \sum_{S:i\in S} \pi_S$. In addition, define $\boldsymbol{\pi} = [\pi_S]_{m\times 1}$ as a column vector with the steady-state probabilities and $\mathbf{S} = [s_{ij}]_{m\times n}$ as a binary matrix describing the feasible link sets, with m being the number of sets and n being the number of links. Each element s_{ij} in $\mathbf{S}$ is 1 if the jth link is active in the ith set, and 0 otherwise. A throughput vector $\boldsymbol{\lambda}$ is then defined as *feasible* if there exists a steady-state solution $\boldsymbol{\pi}$ in the product-form of (11), with $\boldsymbol{\rho} \succeq \mathbf{0}$, such that $\boldsymbol{\lambda} = \mathbf{S}^T\boldsymbol{\pi}$. Let $\Lambda = \{\boldsymbol{\lambda} \in \mathbb{R}^n_+ \mid \boldsymbol{\lambda} = \mathbf{S}^T\boldsymbol{\pi}\}$ be the set of all feasible vectors, i.e., Λ is the capacity region.

From this notation, any feasible throughput vector $\boldsymbol{\lambda} = f(\boldsymbol{\rho})$ can then be obtained from a function $f : \mathbb{R}^n_+ \to \Lambda$ of $\boldsymbol{\rho}$, whose shape depends on the interference constraints of the wireless network. We now prove that f is *bijective*, i.e., there is a one-to-one correspondence between $\boldsymbol{\lambda}$ and $\boldsymbol{\rho}$. In fact, we show that any feasible $\boldsymbol{\lambda}$ is generated by only one steady-state solution $\boldsymbol{\pi}$, and that each solution $\boldsymbol{\pi}$ is generated by only one vector $\boldsymbol{\rho}$, i.e., $\boldsymbol{\lambda} \leftrightarrow \boldsymbol{\pi} \leftrightarrow \boldsymbol{\rho}$. The proof is given in the Appendix.

Theorem 2: There is a one-to-one correspondence between a feasible throughput vector $\boldsymbol{\lambda}$ and $\boldsymbol{\rho}$ via $\boldsymbol{\pi}$, i.e., $\boldsymbol{\lambda} \leftrightarrow \boldsymbol{\pi} \leftrightarrow \boldsymbol{\rho}$.

As a corollary, the function f must always have an inverse function $f^{-1} : \Lambda \to \mathbb{R}^n_+$, such that, if $\boldsymbol{\lambda} = f(\boldsymbol{\rho})$, then

$$\boldsymbol{\rho} = f^{-1}(\boldsymbol{\lambda}). \tag{14}$$

To characterize the capacity region, the first step is then to find this inverse function. However, f^{-1} cannot be easily found from the solution of a linear system anymore, as in (12). Instead, the system of equations $\boldsymbol{\lambda} = f(\boldsymbol{\rho})$ becomes *nonlinear* when nodes

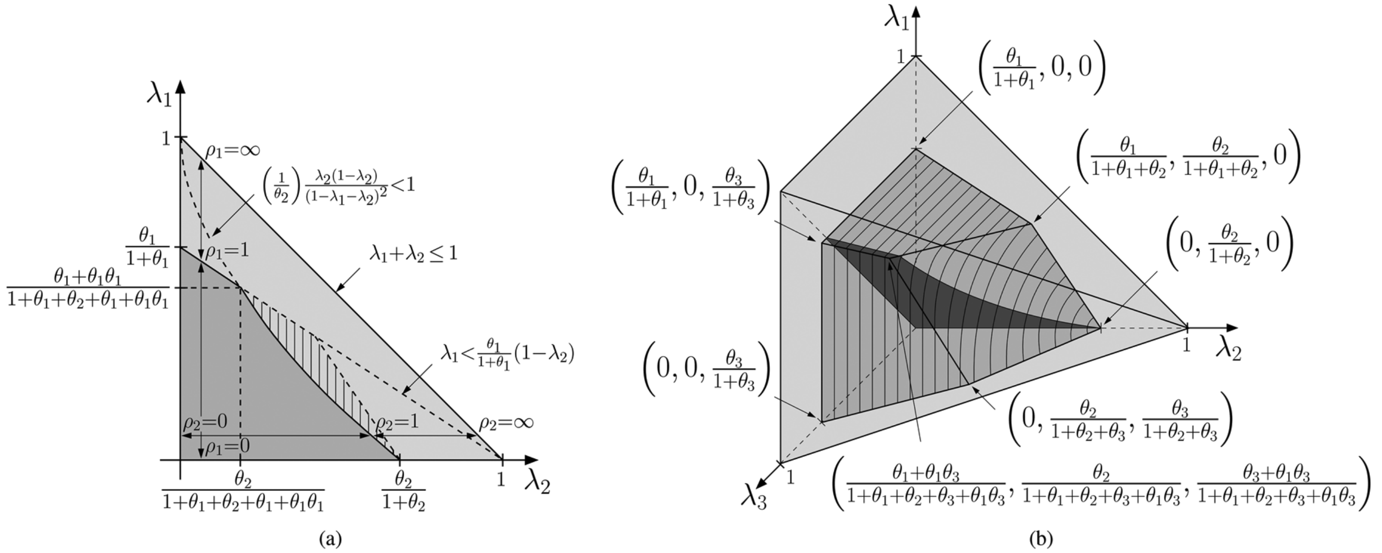


Fig. 5. The capacity region for three transmitters, not all within carrier-sense range. Transmitter τ_2 hears both τ_1 and τ_3, but τ_1 and τ_3 cannot hear each other. (a) The cross section of the capacity region for $\lambda_1 = \lambda_3$. (b) The capacity region, with the plane at $\lambda_1 = \lambda_3$ depicting the cross section shown at (a).

are not all within range. As a result, symbolical computation or numerical methods may have to be used to find f^{-1}. Once the inverse is known, the conditions

$$f^{-1}(\boldsymbol{\lambda}) \succeq \mathbf{0} \quad \text{or} \quad \mathbf{0} \preceq f^{-1}(\boldsymbol{\lambda}) \prec \mathbf{1} \tag{15}$$

are used to characterize the capacity region under weak or strong stability, respectively.

As an example, consider a topology with three transmitters, such that τ_2 is within carrier-sense range of both τ_1 and τ_3, but τ_1 and τ_3 cannot hear each other. In this case, from (11),

$$\pi_S = \frac{\prod_{i\in S}\rho_i\theta_i}{1+\rho_1\theta_1+\rho_2\theta_2+\rho_3\theta_3+(\rho_1\theta_1)(\rho_3\theta_3)} \tag{16}$$

from which the system $\boldsymbol{\lambda} = f(\boldsymbol{\rho})$ can be built as

$$\lambda_1 = \pi_1 + \pi_{1,3} \qquad \lambda_2 = \pi_2 \qquad \lambda_3 = \pi_3 + \pi_{1,3}. \tag{17}$$

This system can be symbolically solved for $\boldsymbol{\rho} = (\rho_1, \rho_2, \rho_3)$, and the inverse function $\boldsymbol{\rho} = f^{-1}(\boldsymbol{\lambda})$ is expressed as

$$\begin{aligned}\rho_1 &= \left(\frac{1}{\theta_1}\right)\frac{\lambda_1}{1-\lambda_1-\lambda_2}\\ \rho_2 &= \left(\frac{1}{\theta_2}\right)\frac{\lambda_2(1-\lambda_2)}{(1-\lambda_1-\lambda_2)(1-\lambda_2-\lambda_3)}\\ \rho_3 &= \left(\frac{1}{\theta_3}\right)\frac{\lambda_3}{1-\lambda_2-\lambda_3}.\end{aligned} \tag{18}$$

Applying the weak stability condition $\rho_i \geq 0$ to (18), each λ_i must be non-negative, and the normalized throughputs must satisfy the constraints $\lambda_1 + \lambda_2 \leq 1$ and $\lambda_2 + \lambda_3 \leq 1$. From the strong stability condition $\rho_i < 1$, the strict inequalities

$$\lambda_1 < \frac{\theta_1}{1+\theta_1}(1-\lambda_2) \qquad \lambda_3 < \frac{\theta_3}{1+\theta_3}(1-\lambda_2) \tag{19}$$

must hold for ρ_1 and ρ_3, and, for ρ_2, we must respect

$$\left(\frac{1}{\theta_2}\right)\frac{\lambda_2(1-\lambda_2)}{(1-\lambda_1-\lambda_2)(1-\lambda_2-\lambda_3)} < 1. \tag{20}$$

From these inequalities, the capacity region of the network can be fully characterized, and it is depicted in Fig. 5. For ease of visualization, Fig. 5(a) shows the cross section for the case where $\lambda_1 = \lambda_3$. The outer capacity region in light gray depicts the area limited by weak stability constraints, i.e., non-negativity and $\lambda_1 + \lambda_2 \leq 1$. From strong stability, only one of the two inequalities in (19) is active when $\lambda_1 = \lambda_3$ (i.e., the other inequality is always satisfied), and (20) becomes an elliptical constraint, creating the inner capacity region depicted in dark gray. This region is clearly *nonconvex*, and thus the convexity of the capacity region does not necessarily hold when nodes are not all within carrier-sense range.

Compared with Fig. 4(a), the dashed area in Fig. 5(a) represents the capacity lost due to the lack of synchronization between τ_1 and τ_3. If both nodes were perfectly synchronized and transmitting at exactly the same time, then they would behave as a single transmitter to τ_2. In this case, the dashed area would be feasible and Figs. 4(a) and 5(a) would be the same. However, due to the nature of CSMA/CA, transmissions from both neighbors partially overlap at τ_2, significantly reducing its medium access and capacity.

Fig. 5(b) shows the capacity region of the network, with the plane at $\lambda_1 = \lambda_3$ depicting the cross section of Fig. 5(a). The outer pyramidal region is defined by the weak stability constraints $\lambda_1 + \lambda_2 \leq 1$ and $\lambda_2 + \lambda_3 \leq 1$, and corresponds to the capacity region of a TDMA network. The inner region is defined by the strong stability constraints (19) and (20), and is depicted with contour lines.

VI. Feasibility and Infeasibility

In addition to characterizing the capacity region, our theory has other applications. In Section VI-A, we provide the analytical expression for each ρ_i and introduce tests to verify feasibility and, in Section VI-B, we show how to predict the network response to an infeasible input rate.

A. Feasibility Testing

We first provide an expression for $\boldsymbol{\rho}$, from which feasibility can be tested. Consider the unfinished work $U_i(t)$ of node τ_i (cf. Fig. 2) and assume that the frozen time of τ_i is *removed* from the timeline. Since both the backoff and the arrival counters freeze during neighbor transmissions, the behavior of τ_i in this timeline is the same as if it was transmitting *alone* in the network. Let $a_i(t)$ be the number of packets generated by τ_i in a large window $[0, t)$ and $E[A_i]$ be the average interarrival time of τ_i. From the strong law of large numbers for renewal processes, $\lim_{t\to\infty} a_i(t)/t = 1/E[A_i]$; therefore, the total time t gets closer to $a_i(t)E[A_i]$ as $t \to \infty$. Assuming stability, no packets accumulate in the queues, and the total time that τ_i transmits in $[0, t)$ is approximately $a_i(t)E[T_i]/p_i$, since each packet is transmitted $1/p_i$ times on average. The fraction of time that τ_i transmits in the unfrozen timeline is then

$$\frac{E[T_i]/p_i}{E[A_i]} = \frac{\rho_i\theta_i}{1+\rho_i\theta_i} \tag{21}$$

where the right-hand side is the fraction of time that τ_i transmits when it is alone, from (11). Isolating ρ_i in (21) results in

$$\rho_i = \frac{E[B_i]/p_i}{E[A_i] - E[T_i]/p_i}. \tag{22}$$

Any average interarrival time $E[A_i] \geq E[T_i]/p_i$ is then weakly stable (i.e., $\rho_i \geq 0$). If $E[A_i] > E[B_i]/p_i + E[T_i]/p_i$, then it is strongly stable as well (i.e., $0 \leq \rho_i < 1$).

We now provide a second test to determine if a given vector $\boldsymbol{\lambda} = (\lambda_1, \lambda_2, \ldots, \lambda_n)$ is feasible without having to resort to a graphical solution (cf. Section V). Let $E[\tilde{A}_i]$ be the overall average interarrival time at τ_i including its frozen time. Assuming stability and following the same logic as before, the fraction of time λ_i that τ_i transmits must then be

$$\lambda_i = \frac{E[T_i]/p_i}{E[\tilde{A}_i]} = \frac{E[A_i]}{E[\tilde{A}_i]}\left(\frac{\rho_i\theta_i}{1+\rho_i\theta_i}\right) \tag{23}$$

where the second equality holds from (21). From Theorem 1, we know that the relation $\boldsymbol{\lambda} \leftrightarrow \boldsymbol{\rho}$ is one-to-one, and thus, from (21) and (23), $E[A_i]$ and $E[\tilde{A}_i]$ must be unique for each $\boldsymbol{\lambda}$.

Packets are generated at each τ_i with an average size of $E[P_i] = E[T_i] \times r_i$. To know if a given input rate $E[P_i]/E[\tilde{A}_i]$ at each τ_i is feasible, the throughput λ_i of each transmitter must first be computed from (23) to derive $\boldsymbol{\lambda}$. After computing $\boldsymbol{\rho} = f^{-1}(\boldsymbol{\lambda})$, the signs of the ρ factors are checked. If $\boldsymbol{\rho} \succeq \mathbf{0}$, then $\boldsymbol{\lambda}$ is feasible and at least weakly stable; if $\boldsymbol{\rho} \prec \mathbf{1}$, then $\boldsymbol{\lambda}$ is also strongly stable.

From these tests we see that only $E[A_i]$, $E[B_i]$, $E[T_i]$, r_i, and p_i are required to determine feasibility. Both 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.

B. Network Response

Consider now the case where sources generate too much traffic, such that the injected rate is known to be outside the capacity region. In this case, some transmitters are not able to forward all the generated traffic and their queues eventually saturate. However, it is not clear at first which transmitters saturate and which do not, since their individual capacities are interdependent. In this section we show how the network responds to an infeasible input rate.

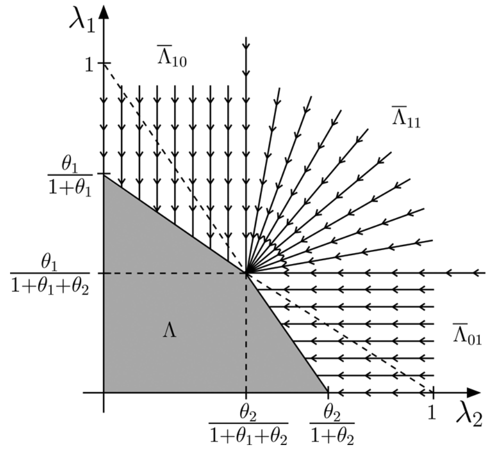

λ_1
1
$\overline{\Lambda}_{10}$
$\overline{\Lambda}_{11}$
$\frac{\theta_1}{1+\theta_1}$
$\frac{\theta_1}{1+\theta_1+\theta_2}$
Λ
$\overline{\Lambda}_{01}$
$\frac{\theta_2}{1+\theta_1+\theta_2}$
$\frac{\theta_2}{1+\theta_2}$
1
λ_2

Fig. 6. The projection lines for each input rate $\boldsymbol{\lambda}' \notin \Lambda$. The complement region $\overline{\Lambda}$ is divided into three subregions $\overline{\Lambda}_{10}$, $\overline{\Lambda}_{01}$, and $\overline{\Lambda}_{11}$ according to the transmitters that are saturated in the projection $\boldsymbol{\lambda} \in \Lambda$ of each subregion.

For ease of exposition, assume that each $E[B_i]$ cannot be reduced, i.e., the network is either strongly stable or unstable. In addition, let Λ be the capacity region and $\overline{\Lambda}$ its complement in $\mathbb{R}_+^n$. Assume that each $E[\tilde{A}_i]$ is known, and let $\boldsymbol{\lambda}'$ be the corresponding input rate computed from (23). Let $\boldsymbol{\lambda}$ be the rate at which the network operates, such that $\boldsymbol{\lambda} = \boldsymbol{\lambda}'$ if $\boldsymbol{\lambda}' \in \Lambda$, and $\boldsymbol{\lambda} \preceq \boldsymbol{\lambda}'$, with $\lambda_i < \lambda_i'$ for at least a transmitter τ_i, otherwise. In this case, the network can be viewed as a system and we call $\boldsymbol{\lambda}$ the *network response* to the input rate $\boldsymbol{\lambda}'$.

When $\boldsymbol{\lambda}' \notin \Lambda$, at least one node must be saturated, and therefore the network response $\boldsymbol{\lambda} \in \Lambda$ is a projection of $\boldsymbol{\lambda}'$ onto the border of the capacity region. As an example, consider again the case of two transmitters within carrier-sense range showed in Fig. 4(a). Fig. 6 shows the projection lines for each input rate $\boldsymbol{\lambda}' = (\lambda_1', \lambda_2')$ in $\overline{\Lambda}$. Clearly, there are three subregions in $\overline{\Lambda}$ whose projections are different. We define $\overline{\Lambda}_{b_1 b_2}$, with each $b_i \in \{0, 1\}$, as the subregion in $\overline{\Lambda}$ whose projections result in transmitter τ_i being saturated if $b_i = 1$.

At $\overline{\Lambda}_{10}$, any input $\boldsymbol{\lambda}'$ is vertically projected down, and the network response is $\boldsymbol{\lambda} = (\lambda_1, \lambda_2')$ for some $\lambda_1 < \lambda_1'$. This occurs because, when

$$0 \leq \lambda_2' < \frac{\theta_2}{1+\theta_1+\theta_2} \tag{24}$$

transmitter τ_2 cannot saturate. Its saturation line is outside the capacity region, and thus τ_1 can never inject enough traffic to saturate τ_2 without saturating itself first. The vertical gap $\lambda_1' - \lambda_1$ is the traffic that τ_1 injects in excess. At $\overline{\Lambda}_{01}$, the projection is horizontal to the left, and the response is $\boldsymbol{\lambda} = (\lambda_1', \lambda_2)$ for some $\lambda_2 < \lambda_2'$. Similarly, when

$$0 \leq \lambda_1' < \frac{\theta_1}{1+\theta_1+\theta_2} \tag{25}$$

transmitter τ_1 can never saturate. The horizontal gap $\lambda_2' - \lambda_2$ is the excess traffic injected by τ_2. Finally, for any input rate $\boldsymbol{\lambda}' \in \overline{\Lambda}_{11}$, the projection is to the point where both nodes are saturated, since in this case both τ_1 and τ_2 are injecting too much traffic. The gaps $\lambda'_1 - \lambda_1$ and $\lambda'_2 - \lambda_2$ are the excess traffic injected by τ_1 and τ_2, respectively.

The aforementioned procedure can be generalized to find the response of any wireless CSMA/CA network. For any input rate $\boldsymbol{\lambda}' \notin \Lambda$, the response $\boldsymbol{\lambda} \in \Lambda$ can be graphically determined solely from the location of $\boldsymbol{\lambda}'$. In particular, for a network with n nodes, if $\boldsymbol{\lambda}' \in \overline{\Lambda}_{b_1 b_2 \dots b_n}$, then any transmitter τ_i with $b_i = 1$ must be saturated in the projection $\boldsymbol{\lambda}$. However, this method requires knowledge of the $2^n - 1$ subregions in $\overline{\Lambda}$, each with a different projection pattern. This makes it hard to use a graphical solution in general. Therefore, inspired by [13], we show that the network response $\boldsymbol{\lambda}$ can be efficiently computed from the solution of a convex optimization problem.

Given an input rate $\boldsymbol{\lambda}'$, consider the following optimization problem over the n-dimensional variable $\boldsymbol{\nu} = (\nu_1, \dots, \nu_n)$

$$\begin{aligned} \underset{\boldsymbol{\nu} \in \mathbb{R}^n}{\text{maximize}} \quad & g(\boldsymbol{\nu}, \boldsymbol{\lambda}') = \sum_{i=1}^{n} \lambda'_i \nu_i - \log \sum_{S} \exp \sum_{i \in S} \nu_i \\ \text{subject to} \quad & \nu_i \le \log \theta_i, \quad i = 1, \dots, n. \end{aligned} \tag{26}$$

The log-sum-exp function is convex in $\mathbb{R}^n$ and thus $g(\boldsymbol{\nu}, \boldsymbol{\lambda}')$ is concave in $\boldsymbol{\nu}$ for each $\boldsymbol{\lambda}'$. Since (26) is an upper-bounded concave maximization, an optimal $\boldsymbol{\nu}^\star$ is attainable. From

$$\frac{\partial g(\boldsymbol{\nu}, \boldsymbol{\lambda}')}{\partial \nu_i} = \lambda'_i - \sum_{S: i \in S} \frac{\exp \sum_{j \in S} \nu_j}{\sum_K \exp \sum_{j \in K} \nu_j} \tag{27}$$

we see that $\nu_i = \log(\rho_i \theta_i)$, and the maximization in (26) is over the logarithmic transform of $(\rho_1\theta_1, \rho_2\theta_2, \dots, \rho_n\theta_n)$. The constraint $\nu_i \le \log \theta_i$ forces the solution to be within the capacity region under strong stability.

A maximizing sequence $(\boldsymbol{\nu}_k)$ where $g(\boldsymbol{\nu}_k, \boldsymbol{\lambda}') \to g(\boldsymbol{\nu}^\star, \boldsymbol{\lambda}')$ as $k \to \infty$ is given from (27) by the gradient algorithm

$$\boldsymbol{\nu}_{k+1} = [\boldsymbol{\nu}_k + \delta_k \nabla g(\boldsymbol{\nu}_k, \boldsymbol{\lambda}')]_{\mathcal{D}} \tag{28}$$

where $\delta_k \ge 0$ is a small step size and $[\cdot]_{\mathcal{D}}$ is a projection onto the feasible set $\mathcal{D} = \{\boldsymbol{\nu} \in \mathbb{R}^n \mid \nu_i \le \log \theta_i, i = 1, 2, \dots, n\}$. The algorithm converges when $\|\nabla g(\boldsymbol{\nu}_k, \boldsymbol{\lambda}')\| \le \varepsilon$, for some small $\varepsilon \ge 0$. For any input rate $\boldsymbol{\lambda}'$, the optimal solution $\boldsymbol{\nu}^\star$ is unique and generates the network response $\boldsymbol{\lambda}$, as shown in the following theorem. The proof is in the Appendix.

Theorem 3: The optimal $\boldsymbol{\nu}^\star = \arg\max_{\boldsymbol{\nu} \in \mathcal{D}} g(\boldsymbol{\nu}, \boldsymbol{\lambda}')$ *is unique and generates the network response* $\boldsymbol{\lambda}$ *to any input rate* $\boldsymbol{\lambda}'$.

As a corollary, the optimization in (26) can also be used as a simple test to determine if a given input rate $\boldsymbol{\lambda}'$ is feasible. In this case, let $\boldsymbol{\lambda}^\star$ be the network response generated by $\boldsymbol{\nu}^\star$. If $\boldsymbol{\lambda}^\star = \boldsymbol{\lambda}'$, then $\boldsymbol{\lambda}' \in \Lambda$; otherwise, $\boldsymbol{\lambda}^\star \preceq \boldsymbol{\lambda}'$ with $\lambda^\star_i < \lambda'_i$ for at least one transmitter τ_i, and thus $\boldsymbol{\lambda}' \notin \Lambda$.

Under weak stability, the feasible set $\mathcal{D} = \{\boldsymbol{\nu} \in \mathbb{R}^n\}$ is unconstrained. In this case, from $\partial g(\boldsymbol{\nu}^\star, \boldsymbol{\lambda}')/\partial \nu_i = 0$ in (27), $\boldsymbol{\nu}^\star = \arg\max_{\boldsymbol{\nu} \in \mathcal{D}} g(\boldsymbol{\nu}, \boldsymbol{\lambda}')$ is attainable only if $\boldsymbol{\lambda}' \in \Lambda$.

VII. Simulations

We use a discrete event simulator in MATLAB to demonstrate our theoretical results. In Section VII-A we focus on the throughput modeling results, and in Section VII-B we show results on the capacity region and network response.

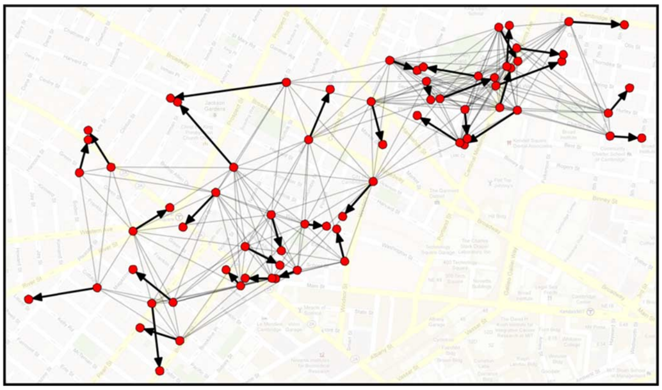


Fig. 7. The MIT Roofnet topology used in our simulations, composed of 70 nodes and 35 links. The interference range is set to 500 meters. Wireless links are represented by arrows and interference is represented by gray lines.

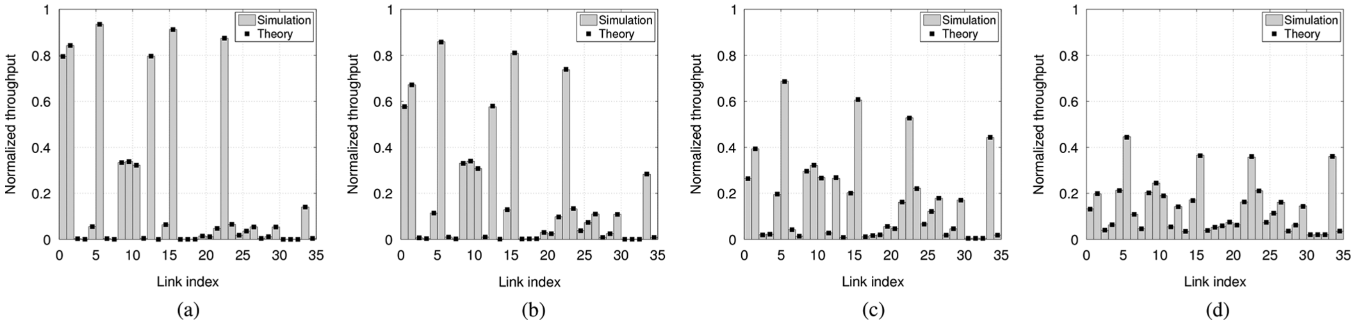
In the simulations, each node implements the CSMA/CA protocol described in Section II, and freezes its backoff counter during any transmission within this range. The backoff interval of each transmitter is uniformly sampled from 25 to 50 μs. Packets are generated at each node with a uniform interarrival time and with a uniform size varying from 1000 to 1500 bytes. For simplicity, the bit rates and the delivery 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.

As predicted, 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 $(1/n)\sum_{i=1}^{n} |\lambda^s_i - \lambda^t_i|/\lambda^t_i$ became lower than 1%, where λ^s_i and λ^t_i are the fraction of time a node τ_i transmits in the simulation and in the proposed theoretical model, respectively.

A. Throughput Modeling

Fig. 7 shows the MIT Roofnet topology used in the simulations. The topology is composed of 70 nodes arranged in 35 links spread over an area of roughly 2.5 km^2. Wireless links are shown using dark arrows and interfering transmitters are connected by gray lines. With a carrier-sense range of 500 meters, there are 5744 link sets in this topology. To ensure that the input rate is feasible, we select some $0 < \rho < 1$ and set $\rho_i = \rho$ for each transmitter τ_i. Its average interarrival time $E[A_i]$ is then computed from (21), and used in the simulation to generate packets at each transmitter τ_i using a uniform distribution.

Fig. 8(a)–(d) shows the normalized throughput λ_i of each link in the network for different values of ρ. Simulation results are shown using vertical bars, and theoretical results are shown using square dots. From these figures, a perfect agreement is seen between the theoretical and simulation results. In addition, the unfairness of the CSMA/CA protocol is also evident. In Fig. 8(a), all sources are closer to saturation and the unfairness is higher, with a few flows achieving high throughput while others starve. This occurs because a saturated network stays, most of the time, in states with the maximum number of active links, i.e., the maximum independent sets [2], [6], [8]. This can be seen from (6); since in practice every θ_i is large, the probability π_S of a maximum independent set S is significantly



(a) (b) (c) (d)

Fig. 8. The normalized throughput $\lambda_i = \sum_{S:i \in S} \pi_S$ of each link in the network under different traffic loads. For each graph, the ρ factor of all transmitters is set to the same value. The vertical bars represent the simulation results and the square dots represent the theoretical results from Section III. (a) $\rho = 0.75$, (b) $\rho = 0.25$, (c) $\rho = 0.05$, and (d) $\rho = 0.01$.

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 [2]. In our topology, the maximum independent sets are composed of seven links and there are only three of these sets, each with a high probability of approximately 27%. Flows 1, 2, 6, 13, 16, and 23 are active in all of the three maximum sets, achieving a throughput higher than 80% in Fig. 8(a). Flows 9, 10, and 11 appear once in each set, achieving roughly 33%.

The maximum independent set approximation works well for networks close to saturation. For unsaturated networks, however, this approximation is not valid. As ρ decreases to 0.25, 0.05, and 0.01 in Figs. 8(b)–(d), respectively, the aforementioned flows become less dominant, resulting in more time available for other flows to transmit. The probability $\pi_{S'}$ of a non-maximum set S' thus becomes non-negligible, and (11) must be used to accurately compute the steady-state probabilities and the throughput of each flow.

B. Capacity Region and Network Response

For ease of exposition, we provide capacity and network response results using the case of Fig. 5(a), i.e., transmitter τ_2 is within range of both τ_1 and τ_3, but τ_1 and τ_3 cannot hear each other. We vary the injected input rate $\boldsymbol{\lambda}' = (\lambda'_1, \lambda'_2)$ over the space $[0,1] \times [0,1]$ to compute the average interarrival times from (23), assuming $\lambda'_1 = \lambda'_3$, and measure the corresponding network response $\boldsymbol{\lambda} = (\lambda_1, \lambda_2)$ from simulation. The default parameters are changed to make the three subregions $\overline{\Lambda}_{01}$, $\overline{\Lambda}_{10}$, and $\overline{\Lambda}_{11}$ visible. In this case, transmitters have an average backoff interval of 50 μs and an average transmission time of $E[T_1] = E[T_3] = 125\,\mu$s and $E[T_2] = 262.5\,\mu$s.

Fig. 9 shows the network capacity and the projection lines for this network. The 'x' dots represent the injected input rate $\boldsymbol{\lambda}'$ and the 'o' dots represent the corresponding network response $\boldsymbol{\lambda}$. The lines connecting each input $\boldsymbol{\lambda}'$ to its response $\boldsymbol{\lambda}$ form the projection lines. The area Λ shown in gray is the strongly stable capacity region computed from (19) and (20). Within the capacity region, we see that the network is able to fully sustain the input rate and transmit all injected traffic, with $\boldsymbol{\lambda} = \boldsymbol{\lambda}'$. Outside the capacity region, however, the input rate $\boldsymbol{\lambda}'$ is not sustainable and it is thus projected onto the border of the capacity region, with $\boldsymbol{\lambda} \preceq \boldsymbol{\lambda}'$ and $\lambda_i < \lambda'_i$ for at least one transmitter τ_i. From the projection lines, a pattern similar to the one in Fig. 6 is also seen here. In particular, for any $0 \leq \lambda'_2 < \theta_2/(1 + \theta_1 + \theta_2 + \theta_1 + \theta_1\theta_1) \approx 0.3$, the saturation line of transmitter τ_2 is outside the capacity region and thus it cannot saturate. Any input rate $\boldsymbol{\lambda}' \in \overline{\Lambda}_{10}$ must then be vertically projected down. In a similar fashion, for any $0 \leq \lambda'_1 < (\theta_1 + \theta_1\theta_1)/(1 + \theta_1 + \theta_2 + \theta_1 + \theta_1\theta_1) \approx 0.5$, the saturation line of transmitters τ_1 and τ_3 is outside the capacity region, and thus they cannot saturate. Any input rate $\boldsymbol{\lambda}' \in \overline{\Lambda}_{01}$ is then horizontally projected to the left. Finally, as expected, any input rate $\boldsymbol{\lambda}' \in \overline{\Lambda}_{11}$ is projected to the saturation point $(0.5, 0.3)$, since in this case all transmitters are injecting too much traffic into the network.

VIII. Related Work

Following the classification in [4], models for wireless networks can be classified into node-centric or set-centric.

1) Node-centric models: In this approach, 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.

Ng and Liew [11] propose a node-centric approach to model the throughput of a single multihop flow. The condition to determine if the flow throughput is limited by hidden nodes or by carrier sensing is provided. Gao *et al.* [10] introduce a node-centric methodology to compute the capacity of several multihop flows. The authors assume that every transmitter is saturated, and compute the flow capacity as the minimum link capacity in the path. Medepalli and Tobagi [14] provide an alternative model based on the computation of the estimated service time of a packet once it reaches the head of the transmission queue. Jindal and Psounis [15] characterize the rate region of 802.11 multihop networks using a decompose-and-combine approach.

Node-centric approaches try to model the performance of each node individually without a global view of the network. The main difficulty in this case is to compute the fraction of time that transmissions overlap. To address this issue, a few simplifying assumptions, such as independence among transmitters [14] and pairwise interference [10], [15], must be made, or techniques, such as the inclusion-exclusion principle [10], [15], have to be employed. The resulting models, however, become rather complex and provide limited insight into the operation of CSMA/CA networks.

2) Set-centric models: In set-centric approaches, the global network behavior is modeled using the independent link sets. This method results in closed-form analytical solutions which provide a better understanding of the CSMA/CA operation.

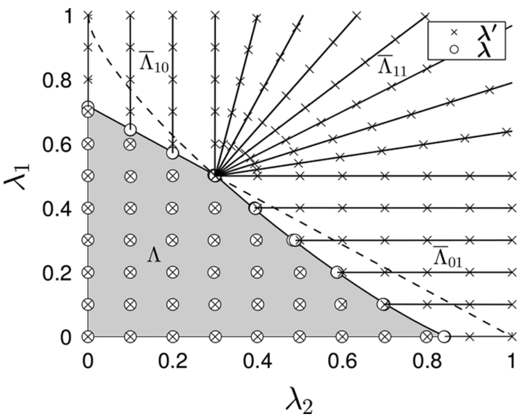


Fig. 9. The capacity region and network response of the $\tau_1 \leftrightarrow \tau_2 \leftrightarrow \tau_3$ topology. The 'x' dots represent the input rate $\boldsymbol{\lambda}'$ and the 'o' dots represent the corresponding network response $\boldsymbol{\lambda}$ (connected with a line to $\boldsymbol{\lambda}'$).

In a seminal work, Boorstyn *et al.* [5] model a wireless CSMA network as a continuous-time Markov process whose states are the independent link sets. For saturated queues with exponentially distributed medium access attempts, the authors prove that the steady-state probabilities have a product-form solution. Brazio [7] generalizes this result and shows that the product-form solution holds for a wider class of MAC protocols as long as the hearing matrix is symmetric. Wang and Kar [6], as well as Durvy *et al.* [8], use the same model to study the fairness problem in CSMA networks under saturation conditions. The authors show that unfairness is mainly caused by topology inequalities, with nodes at the network border being significantly favored. Garetto *et al.* [4] propose a node-centric model for 802.11 which employs the results in [5] to derive the time that each node counts down. Nardelli and Knightly [9] extend this model further to address collisions and hidden terminals in 802.11 networks, providing closed-form expressions to compute the network throughput.

The aforementioned works assume exponentially distributed backoff intervals. Recently, Liew *et al.* [2] proved that the product-form solution holds for *any* backoff distribution. Van de Ven *et al.* [3] present the same insensitivity result and the stability condition for two unsaturated topologies. Kai and Zhang [16] independently derive a model similar to ours (Section III) to approximate unsaturated networks. Different than previous work, however, we do not assume saturated sources [2], [4]–[11], exponential distributions [3]–[9], specific topologies [3], or approximations [16]. As a result, our model is more general and applies to arbitrary CSMA/CA networks. Nonetheless, for the steady-state solution $\boldsymbol{\pi}$ in (11) to be exact, we do require that the arrival process at each node freezes during neighbor transmissions. If the arrival process is not frozen, (11) can still be a good approximation under certain conditions. This is notably the case when the input rate $\boldsymbol{\lambda} \in \Lambda$ is relatively far from the boundary, i.e., the time that a node is frozen is much shorter than its average interarrival time.

3) Bounds and capacity: Significant research is also dedicated to derive asymptotic bounds for wireless multihop networks. In a seminal work, Gupta and Kumar [17] show that, in a network of n nodes, each communicating with another randomly selected node, the per-node throughput is upper bounded by $O(1/\sqrt{n})$. Toumpis and Goldsmith [18], Jain *et al.* [19], and Kodialam and Nandagopal [20] use a convex combination of the capacities of the feasible link sets to derive network capacity bounds. These works, however, assume a TDMA network with an optimal centralized scheduler and do not directly apply to CSMA networks. One exception is provided by Chau *et al.* [21], who show that the same per-node throughput upper bound of $O(1/\sqrt{n})$ can also be achieved by CSMA networks. Recently, Jiang and Walrand [13] propose that nodes adjust their backoffs based on queue lengths, and prove that this scheme achieves the network capacity. In a similar fashion, we show that TDMA is a particular case of CSMA/CA when $\boldsymbol{\rho} \to \boldsymbol{\infty}$. Therefore, both networks must have the same capacity, although this is only true in the ideal case where backoff intervals are infinitesimal.

To the best of our knowledge, a full characterization of the capacity region of wireless CSMA/CA networks is still missing, and we believe that the equations introduced in this work are the first attempt to do so.

IX. Conclusion

In this paper, we introduced a theory that is able to not only predict the behavior of wireless CSMA/CA 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 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. The theory also extends naturally to TDMA networks, shown to be a particular case of CSMA/CA when backoff intervals are infinitesimal. Finally, feasibility tests and a convex optimization that efficiently determines the network response to infeasible input rates are also introduced.

Appendix
Proofs

Theorem 1: By freezing the arrival process, the probability π_S that a link set S is active in an unsaturated network is

$$\pi_S = \frac{\prod_{i \in S} \rho_i \theta_i}{\sum_K \prod_{k \in K} \rho_k \theta_k}.$$

Proof: We first derive the forward Kolmogorov equation for an unsaturated wireless CSMA/CA network. For this purpose, the network state is supplemented with three extra variables, such that the resulting stochastic process becomes Markovian. Let $\mathbf{a} = (a_1, a_2, \ldots, a_n)$ be a vector containing the remaining time until the next arrival of each transmitter. Similarly, let $\mathbf{r} = (r_1, r_2, \ldots, r_n)$ contain the remaining time until each node finishes its current backoff or transmission, and let $\mathbf{q} = (q_1, q_2, \ldots, q_n)$ contain the number of packets in the queue of each transmitter. A binary vector $\mathbf{s} = (s_1, s_2, \ldots, s_n)$ is used to represent each feasible link set, such that $s_i = 1$ if the ith node transmits in this set and $s_i = 0$ otherwise. The network state $X(t)$ at time t is then a tuple $(\mathbf{s},\mathbf{q},\mathbf{a},\mathbf{r})$, which summarizes the history of the entire process, i.e., given $X(t)$, the future behavior is independent of the past. We set out to find how the probability density $p_{\mathbf{s},\mathbf{q}}(\mathbf{a},\mathbf{r},t)$, that at time t the network is at state $(\mathbf{s},\mathbf{q},\mathbf{a},\mathbf{r})$, evolves over time.

First, additional notation is introduced. For a vector $\mathbf{s}$, let $T(\mathbf{s})$ be the set of transmitting nodes, $B(\mathbf{s})$ be the set of nodes allowed to reduce their backoff counters, and $F(\mathbf{s})$ be the set of frozen nodes. We define $\mathbf{u}_\mathbf{s} = [u_i]_{n\times 1}$ as a binary vector representing the unfrozen nodes in the set $\mathbf{s}$, i.e., $u_i = 1$ if $i \notin F(\mathbf{s})$ and $u_i = 0$ otherwise. Let $C(\mathbf{s},\mathbf{q}) = \{i \notin F(\mathbf{s}) \,|\, q_i > 0\}$ be the set of nodes that are either actively counting down or transmitting in network state $(\mathbf{s},\mathbf{q})$. The vector $\mathbf{c}_{\mathbf{s},\mathbf{q}} = [c_i]_{n\times 1}$ is a binary vector that represents this relation, such that $c_i = 1$ if both $i \notin F(\mathbf{s})$ and $q_i > 0$, and $c_i = 0$ otherwise. At network state $(\mathbf{s},\mathbf{q})$, let $A(\mathbf{s},\mathbf{q}) = \{i \notin F(\mathbf{s}) |\, q_i > 1 \text{ if } i \in T(\mathbf{s}) \text{ or } q_i > 0 \text{ if } i \in B(\mathbf{s})\}$ be the set of nodes at which an arrival results in the queue vector $\mathbf{q}$, i.e., if $i \in A(\mathbf{s},\mathbf{q})$, then an arrival may occur at this node, changing the queue state from $\mathbf{q}-\mathbf{e}_i$ to $\mathbf{q}$, where $\mathbf{e}_i$ is a unit vector with the ith entry equal to 1 and all others equal to 0. Arrivals at a node $i \notin A(\mathbf{s},\mathbf{q})$ cannot occur because either $i \in F(\mathbf{s})$ or the queue vector $\mathbf{q}-\mathbf{e}_i$ is infeasible at $\mathbf{s}$. Finally, we define both $\tilde{\mathbf{a}}_i = (a_1,\ldots,a_{i-1},0,a_{i+1},\ldots,a_n)$ and $\tilde{\mathbf{r}}_i = (r_1,\ldots,r_{i-1},0,r_{i+1},\ldots,r_n)$ as the vectors $\mathbf{a}$ and $\mathbf{r}$ with their ith entry equal to 0, respectively.

Without loss of generality, each node samples a new backoff interval after a transmission regardless of its queue state; however, the node only counts down when its queue is nonempty. For simplicity, transmissions are assumed to be correctly received (the case with transmission errors is similar).

Let the functions $f_{A_i}(\cdot)$, $f_{B_i}(\cdot)$, and $f_{T_i}(\cdot)$ be the probability densities of the interarrival, backoff, and transmission times of τ_i, respectively. Then, for a small enough interval Δt and for a valid state $(\mathbf{s},\mathbf{q},\mathbf{a},\mathbf{r})$, the motion of the process is governed by the Chapman–Kolmogorov equation

$$\begin{aligned}
p_{\mathbf{s},\mathbf{q}}(\mathbf{a},\mathbf{r},t+\Delta t) &= p_{\mathbf{s},\mathbf{q}}(\mathbf{a}+\mathbf{u}_\mathbf{s}\Delta t,\mathbf{r}+\mathbf{c}_{\mathbf{s},\mathbf{q}}\Delta t,t)\\
&+\sum_{i\in A(\mathbf{s},\mathbf{q})} p_{\mathbf{s},\mathbf{q}-\mathbf{e}_i}(\tilde{\mathbf{a}}_i+\mathbf{u}_\mathbf{s}\Delta t,\mathbf{r}+\mathbf{c}_{\mathbf{s},\mathbf{q}-\mathbf{e}_i}\Delta t,t)f_{A_i}(a_i)\Delta t\\
&+\sum_{i\in T(\mathbf{s})} p_{\mathbf{s}-\mathbf{e}_i,\mathbf{q}}(\mathbf{a}+\mathbf{u}_{\mathbf{s}-\mathbf{e}_i}\Delta t,\tilde{\mathbf{r}}_i+\mathbf{c}_{\mathbf{s}-\mathbf{e}_i,\mathbf{q}}\Delta t,t)f_{T_i}(r_i)\Delta t\\
&+\sum_{i\in B(\mathbf{s})} p_{\mathbf{s}+\mathbf{e}_i,\mathbf{q}+\mathbf{e}_i}(\mathbf{a}+\mathbf{u}_{\mathbf{s}+\mathbf{e}_i}\Delta t,\mathbf{r}_i+\mathbf{c}_{\mathbf{s}+\mathbf{e}_i,\mathbf{q}+\mathbf{e}_i}\Delta t,t)f_{B_i}(r_i)\Delta t.
\end{aligned}\tag{29}$$

Assuming that at most one event occurs in the time interval $(t,t+\Delta t)$ and the network state is $X(t+\Delta t) = (\mathbf{s},\mathbf{q},\mathbf{a},\mathbf{r})$ at time $t+\Delta t$, then only a few states $X(t)$ are possible. The first term on the right-hand side (RHS) of (29) is the case where no events occur in $(t,t+\Delta t)$, and thus the counters just decrease by Δt. The vector $\mathbf{u}_\mathbf{s}$ enforces that only unfrozen nodes decrease the arrival counters and $\mathbf{c}_{\mathbf{s},\mathbf{q}}$ enforces that only unfrozen nodes with a nonempty queue decrease their backoff/transmission counters. The other terms represent the cases where an event occurs in $(t,t+\Delta t)$. The second term in the RHS of (29) is for the events where an arrival occurs at a node $i \in A(\mathbf{s},\mathbf{q})$, changing the queue vector from $\mathbf{q}-\mathbf{e}_i$ to $\mathbf{q}$; the third term is for the events where a node $i \in T(\mathbf{s})$ finishes its backoff period and starts a transmission; and the final term is for the events where a node $i \in B(\mathbf{s})$ finishes its transmission and then samples the backoff interval of the next packet, even if $q_i = 0$ after the transmission.

Expanding the first term on the RHS of (29) results in

$$\begin{aligned}
p_{\mathbf{s},\mathbf{q}}(\mathbf{a}+\mathbf{u}_\mathbf{s}\Delta t,\mathbf{r}+\mathbf{c}_{\mathbf{s},\mathbf{q}}\Delta t,t) &= p_{\mathbf{s},\mathbf{q}}(\mathbf{a},\mathbf{r},t)\\
+\sum_{i\notin F(\mathbf{s})}\frac{\partial p_{\mathbf{s},\mathbf{q}}(\mathbf{a},\mathbf{r},t)}{\partial a_i}\Delta t &+ \sum_{i\in C(\mathbf{s},\mathbf{q})}\frac{\partial p_{\mathbf{s},\mathbf{q}}(\mathbf{a},\mathbf{r},t)}{\partial r_i}\Delta t + o(\Delta t)
\end{aligned}\tag{30}$$

where $\lim_{\Delta t\to 0} o(\Delta t)/\Delta t = 0$. By substituting (30) back into (29), subtracting $p_{\mathbf{s},\mathbf{q}}(\mathbf{a},\mathbf{r},t)$ from both sides, dividing by Δt, and taking the limit as $\Delta t \to 0$, we obtain the forward Kolmogorov equation for wireless CSMA/CA networks as

$$\begin{aligned}
\frac{\partial p_{\mathbf{s},\mathbf{q}}(\mathbf{a},\mathbf{r},t)}{\partial t} &= \sum_{i\notin F(\mathbf{s})}\frac{\partial p_{\mathbf{s},\mathbf{q}}(\mathbf{a},\mathbf{r},t)}{\partial a_i} + \sum_{i\in C(\mathbf{s},\mathbf{q})}\frac{\partial p_{\mathbf{s},\mathbf{q}}(\mathbf{a},\mathbf{r},t)}{\partial r_i}\\
&+\sum_{i\in A(\mathbf{s},\mathbf{q})} p_{\mathbf{s},\mathbf{q}-\mathbf{e}_i}(\tilde{\mathbf{a}}_i,\mathbf{r},t)f_{A_i}(a_i)\\
&+\sum_{i\in T(\mathbf{s})} p_{\mathbf{s}-\mathbf{e}_i,\mathbf{q}}(\mathbf{a},\tilde{\mathbf{r}}_i,t)f_{T_i}(r_i)\\
&+\sum_{i\in B(\mathbf{s})} p_{\mathbf{s}+\mathbf{e}_i,\mathbf{q}+\mathbf{e}_i}(\mathbf{a},\tilde{\mathbf{r}}_i,t)f_{B_i}(r_i).
\end{aligned}\tag{31}$$

In steady state, convergence occurs and $\partial p_{\mathbf{s},\mathbf{q}}(\mathbf{a},\mathbf{r},t)/\partial t = 0$. Defining $\pi_{\mathbf{s},\mathbf{q}}(\mathbf{a},\mathbf{r}) = \lim_{t\to\infty} p_{\mathbf{s},\mathbf{q}}(\mathbf{a},\mathbf{r},t)$ and assuming that the limit exists, the global balance equation is then

$$\begin{aligned}
-\sum_{i\notin F(\mathbf{s})}\frac{\partial \pi_{\mathbf{s},\mathbf{q}}(\mathbf{a},\mathbf{r})}{\partial a_i} &- \sum_{i\in C(\mathbf{s},\mathbf{q})}\frac{\partial \pi_{\mathbf{s},\mathbf{q}}(\mathbf{a},\mathbf{r})}{\partial r_i}\\
&= \sum_{i\in A(\mathbf{s},\mathbf{q})} \pi_{\mathbf{s},\mathbf{q}-\mathbf{e}_i}(\tilde{\mathbf{a}}_i,\mathbf{r})f_{A_i}(a_i)\\
&+\sum_{i\in T(\mathbf{s})} \pi_{\mathbf{s}-\mathbf{e}_i,\mathbf{q}}(\mathbf{a},\tilde{\mathbf{r}}_i)f_{T_i}(r_i)\\
&+\sum_{i\in B(\mathbf{s})} \pi_{\mathbf{s}+\mathbf{e}_i,\mathbf{q}+\mathbf{e}_i}(\mathbf{a},\tilde{\mathbf{r}}_i)f_{B_i}(r_i).
\end{aligned}\tag{32}$$

In order to find the steady-state probability $\pi_\mathbf{s}$, the first step is to marginalize $\mathbf{q}$ out of the density $\pi_{\mathbf{s},\mathbf{q}}(\mathbf{a},\mathbf{r})$. Let $Q(\mathbf{s}) = \{\mathbf{q} \,|\, q_i > 0 \text{ if } i \in T(\mathbf{s}) \text{ and } q_i \geq 0 \text{ if } i \notin T(\mathbf{s})\}$ be the set of feasible queue vectors when the link set $\mathbf{s}$ is active. Summing (32) over all feasible queue vectors $\mathbf{q} \in Q(\mathbf{s})$ and interchanging the order of the summations results in

$$\begin{aligned}
-\sum_{i\notin F(\mathbf{s})}\sum_{\mathbf{q}\in Q(\mathbf{s})}\frac{\partial \pi_{\mathbf{s},\mathbf{q}}(\mathbf{a},\mathbf{r})}{\partial a_i} &- \sum_{i\in T(\mathbf{s})}\sum_{\mathbf{q}\in Q(\mathbf{s})}\frac{\partial \pi_{\mathbf{s},\mathbf{q}}(\mathbf{a},\mathbf{r})}{\partial r_i}\\
-\sum_{i\in B(\mathbf{s})}\sum_{\mathbf{q}\in Q(\mathbf{s}+\mathbf{e}_i)}\frac{\partial \pi_{\mathbf{s},\mathbf{q}}(\mathbf{a},\mathbf{r})}{\partial r_i}&\\
&= \sum_{i\notin F(\mathbf{s})}\sum_{\mathbf{q}\in Q(\mathbf{s})} \pi_{\mathbf{s},\mathbf{q}}(\tilde{\mathbf{a}}_i,\mathbf{r})f_{A_i}(a_i)\\
&+\sum_{i\in T(\mathbf{s})}\sum_{\mathbf{q}\in Q(\mathbf{s})} \pi_{\mathbf{s}-\mathbf{e}_i,\mathbf{q}}(\mathbf{a},\tilde{\mathbf{r}}_i)f_{T_i}(r_i)\\
&+\sum_{i\in B(\mathbf{s})}\sum_{\mathbf{q}\in Q(\mathbf{s}+\mathbf{e}_i)} \pi_{\mathbf{s}+\mathbf{e}_i,\mathbf{q}}(\mathbf{a},\tilde{\mathbf{r}}_i)f_{B_i}(r_i).
\end{aligned}\tag{33}$$

Let $\pi_{\mathbf{s}}(\mathbf{a},\mathbf{r}) = \sum_{\mathbf{q}\in Q(\mathbf{s})} \pi_{\mathbf{s},\mathbf{q}}(\mathbf{a},\mathbf{r})$ be the density with $\mathbf{q}$ marginalized out, and $\pi_{\mathbf{s},q_i^+}(\mathbf{a},\mathbf{r}) = \sum_{\mathbf{q}\in Q(\mathbf{s}+\mathbf{e}_i)}\pi_{\mathbf{s},\mathbf{q}}(\mathbf{a},\mathbf{r})$ be the density with $q_i > 0$, for $i \in B(\mathbf{s})$. Then, (33) is written as

$$\begin{aligned}
&-\sum_{i\notin F(\mathbf{s})}\frac{\partial \pi_{\mathbf{s}}(\mathbf{a},\mathbf{r})}{\partial a_i}-\sum_{i\in T(\mathbf{s})}\frac{\partial \pi_{\mathbf{s}}(\mathbf{a},\mathbf{r})}{\partial r_i}-\sum_{i\in B(\mathbf{s})}\frac{\partial \pi_{\mathbf{s},q_i^+}(\mathbf{a},\mathbf{r})}{\partial r_i}\\
&\quad=\sum_{i\notin F(\mathbf{s})}\pi_{\mathbf{s}}(\tilde{\mathbf{a}}_i,\mathbf{r})f_{A_i}(a_i)+\sum_{i\in T(\mathbf{s})}\pi_{\mathbf{s}-\mathbf{e}_i,q_i^+}(\mathbf{a},\tilde{\mathbf{r}}_i)f_{T_i}(r_i)\\
&\qquad+\sum_{i\in B(\mathbf{s})}\pi_{\mathbf{s}+\mathbf{e}_i}(\mathbf{a},\tilde{\mathbf{r}}_i)f_{B_i}(r_i).
\end{aligned}\tag{34}$$

Noting that

$$-\int_0^\infty \frac{\partial \pi_{\mathbf{s}}(\mathbf{a},\mathbf{r})}{\partial a_i}da_i = \pi_{\mathbf{s}}(\tilde{\mathbf{a}}_i,\mathbf{r})\tag{35}$$

and integrating (34) for $\mathbf{a}\in\mathbb{R}_+^n$ results in the simpler equation

$$\begin{aligned}
&-\sum_{i\in T(\mathbf{s})}\frac{\partial \pi_{\mathbf{s}}(\mathbf{r})}{\partial r_i}-\sum_{i\in B(\mathbf{s})}\frac{\partial \pi_{\mathbf{s},q_i^+}(\mathbf{r})}{\partial r_i}\\
&\quad=\sum_{i\in T(\mathbf{s})}\pi_{\mathbf{s}-\mathbf{e}_i,q_i^+}(\tilde{\mathbf{r}}_i)f_{T_i}(r_i)+\sum_{i\in B(\mathbf{s})}\pi_{\mathbf{s}+\mathbf{e}_i}(\tilde{\mathbf{r}}_i)f_{B_i}(r_i)
\end{aligned}\tag{36}$$

where $\pi_{\mathbf{s}}(\mathbf{r}) = \int \pi_{\mathbf{s}}(\mathbf{a},\mathbf{r})d\mathbf{a}$ and $\pi_{\mathbf{s},q_i^+}(\mathbf{r}) = \int \pi_{\mathbf{s},q_i^+}(\mathbf{a},\mathbf{r})d\mathbf{a}$.

By realizing that $\pi_{\mathbf{s},q_i^+}(\mathbf{r}) \triangleq P[\mathbf{q}\in Q(\mathbf{s}+\mathbf{e}_i)|\mathbf{s},\mathbf{r}]\pi_{\mathbf{s}}(\mathbf{r})$, we claim that, for any $i\in B(\mathbf{s})$, the conditional probability $P[\mathbf{q}\in Q(\mathbf{s}+\mathbf{e}_i)|\mathbf{s},\mathbf{r}] = P[q_i>0|\mathbf{s},\mathbf{r}]= P[q_i>0|\mathbf{s},r_i]$. Given that the active link set is $\mathbf{s}$ and $i\in B(\mathbf{s})$, the first equality holds from the definition of $Q(\mathbf{s}+\mathbf{e}_i)$, and the second equality holds because, by freezing the arrival process, the node state (s_i,q_i,a_i,r_i) becomes independent of the state of other nodes. In fact, when the node is unfrozen, the state (s_i,q_i,a_i,r_i) evolves as if there was nobody else in the network. If we now define $f_{\widehat{B}_i}(r_i) = P[B_i>r_i]/E[B_i]$ as the density of the residual backoff time of node i and $\rho_i(\mathbf{s}) = P[q_i>0|\mathbf{s}]$ as the conditional probability of the queue of node i being nonempty given that $\mathbf{s}$ is active, then

$$\begin{aligned}
&P[q_i>0|\mathbf{s},r_i]\\
&\quad=\frac{\pi(r_i|\mathbf{s},q_i>0)P[q_i>0|\mathbf{s}]}{\pi(r_i|\mathbf{s},q_i>0)P[q_i>0|\mathbf{s}]+\pi(r_i|\mathbf{s},q_i=0)P[q_i=0|\mathbf{s}]}\\
&\quad=\frac{\rho_i f_{\widehat{B}_i}(r_i)}{\rho_i f_{\widehat{B}_i}(r_i)+(1-\rho_i)f_{B_i}(r_i)},
\end{aligned}\tag{37}$$

where $\pi(r_i|\mathbf{s},q_i>0) = f_{\widehat{B}_i}(r_i)$, $\pi(r_i|\mathbf{s},q_i=0) = f_{B_i}(r_i)$, and $\rho_i = \rho_i(\mathbf{s})$ are all independent from the particular set $\mathbf{s}$, given that $i\in B(\mathbf{s})$. In this case, the density $\pi_{\mathbf{s},q_i^+}(\mathbf{r})$ is then

$$\pi_{\mathbf{s},q_i^+}(\mathbf{r}) = \frac{\rho_i f_{\widehat{B}_i}(r_i)}{\rho_i f_{\widehat{B}_i}(r_i)+(1-\rho_i)f_{B_i}(r_i)}\pi_{\mathbf{s}}(\mathbf{r}).\tag{38}$$

Finally, by defining $\phi_{\mathbf{s}}$ as

$$\phi_{\mathbf{s}} = \frac{\prod_{i:s_i=1}\rho_i\theta_i}{\sum_{\mathbf{k}}\prod_{i:k_i=1}\rho_i\theta_i}\tag{39}$$

the solution to (36) and (38) is

$$\pi_{\mathbf{s}}(\mathbf{r}) = \phi_{\mathbf{s}}\prod_{i\in T(\mathbf{s})}f_{\widehat{T}_i}(r_i)\prod_{i\notin T(\mathbf{s})}\left[\rho_i f_{\widehat{B}_i}(r_i)+(1-\rho_i)f_{B_i}(r_i)\right]\tag{40}$$

where $f_{\widehat{T}_i}(r_i) = P[T_i>r_i]/E[T_i]$ is the residual transmission time of node i. In fact, the terms of the left-hand and right-hand sides of (36) match on a one-to-one basis under this density. Integrating $\pi_{\mathbf{s}}(\mathbf{r})$ for $\mathbf{r}\in\mathbb{R}_+^n$, we have $\pi_{\mathbf{s}} = \phi_{\mathbf{s}}$. ■

Theorem 2: There is a one-to-one correspondence between a feasible throughput vector $\boldsymbol{\lambda}$ and $\boldsymbol{\rho}$ via $\boldsymbol{\pi}$, i.e., $\boldsymbol{\lambda}\leftrightarrow\boldsymbol{\pi}\leftrightarrow\boldsymbol{\rho}$.

Proof: This proof is derived in two steps. First, we prove the one-to-one correspondence between $\boldsymbol{\lambda}$ and $\boldsymbol{\pi}$, i.e., $\boldsymbol{\lambda}\leftrightarrow\boldsymbol{\pi}$, and later we do the same for $\boldsymbol{\pi}$ and $\boldsymbol{\rho}$, i.e., $\boldsymbol{\pi}\leftrightarrow\boldsymbol{\rho}$.

$\boldsymbol{\lambda}\leftrightarrow\boldsymbol{\pi}$: For the purpose of contradiction, let $\boldsymbol{\pi}$ and $\boldsymbol{\pi}'$ be two different distributions in the form of (11) that generate the same per-node throughput, i.e., $\boldsymbol{\lambda} = \mathbf{S}^T\boldsymbol{\pi} = \mathbf{S}^T\boldsymbol{\pi}'$. In addition, define the variables $\nu_i = \log(\rho_i\theta_i)$ and $\nu_i' = \log(\rho_i'\theta_i')$ such that $\boldsymbol{\pi}$ and $\boldsymbol{\pi}'$ can be expressed from (11) as

$$\pi_S = \frac{\exp\sum_{i\in S}\nu_i}{\sum_K\exp\sum_{k\in K}\nu_k}\quad \pi_S' = \frac{\exp\sum_{i\in S}\nu_i'}{\sum_K\exp\sum_{k\in K}\nu_k'}.\tag{41}$$

The Kullback-Leibler (KL) divergence is a measure of difference between two distributions $\boldsymbol{\pi}$ and $\boldsymbol{\pi}'$. In our case, it is defined as $D_{\mathrm{KL}}(\boldsymbol{\pi}\|\boldsymbol{\pi}') = \sum_S \pi_S\log(\pi_S/\pi_S')$. This measure is not necessarily symmetric, i.e., $D_{\mathrm{KL}}(\boldsymbol{\pi}\|\boldsymbol{\pi}')\neq D_{\mathrm{KL}}(\boldsymbol{\pi}'\|\boldsymbol{\pi})$, but it is always non-negative and it is zero if and only if $\boldsymbol{\pi} = \boldsymbol{\pi}'$. From (41), $D_{\mathrm{KL}}(\boldsymbol{\pi}\|\boldsymbol{\pi}')$ is computed as

$$\begin{aligned}
&D_{\mathrm{KL}}(\boldsymbol{\pi}\|\boldsymbol{\pi}')\\
&\quad=\sum_S\pi_S\left[\sum_{i\in S}\nu_i-\sum_{i\in S}\nu_i'+\log\frac{\sum_K\exp\sum_{k\in K}\nu_k'}{\sum_K\exp\sum_{k\in K}\nu_k}\right]\\
&\quad=\sum_{i=1}^n(\nu_i-\nu_i')\left(\sum_{S:i\in S}\pi_S\right)+\log\frac{\sum_K\exp\sum_{k\in K}\nu_k'}{\sum_K\exp\sum_{k\in K}\nu_k}\\
&\quad=\sum_{i=1}^n(\nu_i-\nu_i')\,\lambda_i+\log\frac{\sum_K\exp\sum_{k\in K}\nu_k'}{\sum_K\exp\sum_{k\in K}\nu_k}.
\end{aligned}\tag{42}$$

From the assumption that $\lambda_i = \sum_{i\in S}\pi_S = \sum_{i\in S}\pi_S'$, the reverse measure $D_{\mathrm{KL}}(\boldsymbol{\pi}'\|\boldsymbol{\pi})$ can be similarly computed as

$$D_{\mathrm{KL}}(\boldsymbol{\pi}'\|\boldsymbol{\pi}) = \sum_{i=1}^n(\nu_i'-\nu_i)\,\lambda_i-\log\frac{\sum_K\exp\sum_{k\in K}\nu_k'}{\sum_K\exp\sum_{k\in K}\nu_k},\tag{43}$$

and thus $D_{\mathrm{KL}}(\boldsymbol{\pi}\|\boldsymbol{\pi}') = -D_{\mathrm{KL}}(\boldsymbol{\pi}'\|\boldsymbol{\pi})$. Since the KL divergence is non-negative, then $D_{\mathrm{KL}}(\boldsymbol{\pi}\|\boldsymbol{\pi}') = D_{\mathrm{KL}}(\boldsymbol{\pi}'\|\boldsymbol{\pi}) = 0$. As a result, both distributions $\boldsymbol{\pi}$ and $\boldsymbol{\pi}'$ must be equal, which contradicts our initial assumption that $\boldsymbol{\pi}\neq\boldsymbol{\pi}'$. Each feasible $\boldsymbol{\lambda}$ is then generated by a unique distribution $\boldsymbol{\pi}$ and, since each distribution $\boldsymbol{\pi}$ cannot generate more than one vector $\boldsymbol{\lambda}$, there is a one-to-one correspondence between $\boldsymbol{\lambda}$ and $\boldsymbol{\pi}$.

$\boldsymbol{\pi}\leftrightarrow\boldsymbol{\rho}$: Let $\boldsymbol{\rho}$ and $\boldsymbol{\rho}'$ be two non-negative vectors that generate the steady-state solutions $\boldsymbol{\pi}$ and $\boldsymbol{\pi}'$, respectively. For the purpose of contradiction, assume that $\boldsymbol{\rho}\neq\boldsymbol{\rho}'$ but $\boldsymbol{\pi} = \boldsymbol{\pi}'$. If both solutions are identical, then $\pi_\emptyset = \pi'_\emptyset$ and thus the two normalizing constants $\sum_S \prod_{i\in S} \rho_i\theta_i = \sum_S \prod_{i\in S} \rho'_i\theta_i$ must also be equal. Now consider the sets $S = \{i\}$ where only a single transmitter is active. Since $\pi_i = \pi'_i$ and the normalizing constants are equal, then $\rho_i\theta_i = \rho'_i\theta_i$ for $i = 1, 2, \ldots, n$ and thus $\boldsymbol{\rho} = \boldsymbol{\rho}'$, which contradicts the assumption that $\boldsymbol{\rho} \neq \boldsymbol{\rho}'$. As a result, there is a unique vector $\boldsymbol{\rho}$ capable of generating $\boldsymbol{\pi}$. Since each vector $\boldsymbol{\rho}$ cannot generate more than one solution $\boldsymbol{\pi}$, there is a one-to-one correspondence between $\boldsymbol{\pi}$ and $\boldsymbol{\rho}$.

Therefore, $\boldsymbol{\lambda} \leftrightarrow \boldsymbol{\pi} \leftrightarrow \boldsymbol{\rho}$ and, from transitivity, there is a one-to-one correspondence between $\boldsymbol{\lambda}$ and $\boldsymbol{\rho}$, i.e., $\boldsymbol{\lambda} \leftrightarrow \boldsymbol{\rho}$. ■

Theorem 3: The optimal $\boldsymbol{\nu}^\star = \arg\max_{\boldsymbol{\nu}\in\mathcal{D}} g(\boldsymbol{\nu}, \boldsymbol{\lambda}')$ *is unique and generates the network response* $\boldsymbol{\lambda}$ *to any input rate* $\boldsymbol{\lambda}'$.

Proof: First, we prove that $h(\boldsymbol{\nu}) = \log \sum_S \exp \sum_{i\in S} \nu_i$ is strictly convex. From the strict concavity of the logarithmic function, we know that $a^\alpha b^{1-\alpha} < \alpha a + (1-\alpha)b$ for any positive a and b, with $a \neq b$, and $0 < \alpha < 1$. Let $\boldsymbol{\pi}$ and $\boldsymbol{\pi}'$ be two different distributions in the form of (41) generated by $\boldsymbol{\nu}$ and $\boldsymbol{\nu}'$, respectively. Then, for at least one link set S, we have $\pi_S \neq \pi'_S$ and thus $(\pi_S)^\alpha(\pi'_S)^{1-\alpha} < \alpha\pi_S + (1-\alpha)\pi'_S$. For other link sets, the inequality does not need to be strict. Summing over all sets, $\sum_S (\pi_S)^\alpha(\pi'_S)^{1-\alpha} < 1$. From (41),

$$\sum_S \left(\exp \sum_{i\in S} \nu_i\right)^\alpha \left(\exp \sum_{i\in S} \nu'_i\right)^{1-\alpha} < \left(\sum_K \exp \sum_{i\in K} \nu_i\right)^\alpha \left(\sum_K \exp \sum_{i\in K} \nu'_i\right)^{1-\alpha}. \quad (44)$$

From Theorem 2, we know that, since $\boldsymbol{\pi} \neq \boldsymbol{\pi}'$, then $\boldsymbol{\nu} \neq \boldsymbol{\nu}'$. Therefore, taking the logarithm for both sides of (44) results in $h(\alpha\boldsymbol{\nu} + (1-\alpha)\boldsymbol{\nu}') < \alpha h(\boldsymbol{\nu}) + (1-\alpha)h(\boldsymbol{\nu}')$ for any $\boldsymbol{\nu} \neq \boldsymbol{\nu}'$ and $0 < \alpha < 1$. By definition, function $h(\boldsymbol{\nu})$ is then strictly convex and $g(\boldsymbol{\nu}, \boldsymbol{\lambda}') = \boldsymbol{\nu}^T\boldsymbol{\lambda}' - h(\boldsymbol{\nu})$ is strictly concave in $\boldsymbol{\nu}$. As a result, the optimal $\boldsymbol{\nu}^\star = \arg\max_{\boldsymbol{\nu}\in\mathcal{D}} g(\boldsymbol{\nu}, \boldsymbol{\lambda}')$ is unique.

Let $\boldsymbol{\lambda}^\star$ be the normalized throughput generated by $\boldsymbol{\nu}^\star$, i.e.,

$$\lambda^\star_i = \sum_{S:i\in S} \frac{\exp \sum_{j\in S} \nu^\star_j}{\sum_K \exp \sum_{j\in K} \nu^\star_j}, \qquad \text{for } i = 1, 2, \ldots, n. \quad (45)$$

From the monotonicity of the logarithmic function and from Theorem 2, $\boldsymbol{\nu}^\star \leftrightarrow \boldsymbol{\rho}^\star \leftrightarrow \boldsymbol{\lambda}^\star$ is a one-to-one correspondence, and therefore $\boldsymbol{\lambda}^\star$ is also unique.

We now prove that $\boldsymbol{\lambda}^\star$ is the network response $\boldsymbol{\lambda}$. First, Slater's condition is always satisfied, since the feasible set $\mathcal{D}$ only has linear inequalities and is never empty, resulting in zero duality gap. The optimal $\boldsymbol{\nu}^\star = \arg\max_{\boldsymbol{\nu}\in\mathcal{D}} g(\boldsymbol{\nu}, \boldsymbol{\lambda}')$ must then satisfy the Karush–Kuhn–Tucker (KKT) conditions

$$\nu^\star_i \leq \log\theta_i \quad (46a)$$

$$\lambda^\star_i \leq \lambda'_i \quad (46b)$$

$$(\lambda'_i - \lambda^\star_i)(\nu^\star_i - \log\theta_i) = 0 \quad (46c)$$

for $i = 1, 2, \ldots, n$. The network response $\boldsymbol{\lambda}$ satisfies all conditions in (46). Conditions (46a) and (46b) are satisfied by any vector in Λ and, since $\boldsymbol{\lambda} \in \Lambda$, we focus only on the complementary slackness condition in (46c). If at input rate $\boldsymbol{\lambda}'$ the network response of τ_i is $\lambda_i < \lambda'_i$, then τ_i must be saturated, i.e., $\nu_i = \log\theta_i$; otherwise, if τ_i is not saturated, i.e., $\nu_i < \log\theta_i$, then it must sustain the input rate $\lambda_i = \lambda'_i$. In both cases, (46c) is satisfied. Since $\boldsymbol{\lambda}^\star$ is unique and $\boldsymbol{\lambda}$ also satisfies the KKT conditions in (46), then $\boldsymbol{\lambda}^\star = \boldsymbol{\lambda}$. ■

Acknowledgment

The authors are greatly in debt to B. Krishnamachari for suggesting that their preliminary theory could characterize the CSMA/CA capacity region and to S. Borst for realizing that the product-form solution does not hold for general arrival processes. The authors would also like to thank P. Polakos and S. Liew for helpful discussions.

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. Commun.*, vol. COM-23, no. 12, pp. 1400–1416, Dec. 1975.

[2] S. Liew, C. Kai, H. Leung, and P. Wong, "Back-of-the-envelope computation of throughput distributions in CSMA wireless networks," *IEEE Trans. Mobile Comput.*, vol. 9, no. 9, pp. 1319–1331, Sep. 2010.

[3] P. M. van de Ven, S. C. Borst, J. S. H. van Leeuwaarden, and A. Proutière, "Insensitivity and stability of random-access networks," *Performance Eval.*, vol. 67, no. 11, pp. 1230–1242, Nov. 2010.

[4] M. Garetto, T. Salonidis, and E. W. Knightly, "Modeling per-flow throughput and capturing starvation in CSMA multi-hop wireless networks," *IEEE/ACM Trans. Netw.*, vol. 16, no. 4, pp. 864–877, Aug. 2008.

[5] 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.

[6] 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.

[7] J. M. Brázio, "Capacity analysis of multihop packet radio networks under a general class of channel access protocols and capture models," Ph.D. dissertation, Stanford Univ., Stanford, CA, USA, Mar. 1987.

[8] M. Durvy, O. Dousse, and P. Thiran, "On the fairness of large CSMA networks," *IEEE J. Sel. Areas Commun.*, vol. 27, no. 7, pp. 1093–1104, Sep. 2009.

[9] B. Nardelli and E. Knightly, "Closed-form throughput expressions for CSMA networks with collisions and hidden terminals," in *Proc. IEEE INFOCOM*, Mar. 2012, pp. 2309–2317.

[10] 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, pp. 39–50.

[11] P. C. Ng and S. C. Liew, "Throughput analysis of IEEE802.11 multi-hop ad hoc networks," *IEEE/ACM Trans. Netw.*, vol. 15, no. 2, pp. 309–322, Apr. 2007.

[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] L. Jiang and J. Walrand, "A distributed CSMA algorithm for throughput and utility maximization in wireless networks," *IEEE/ACM Trans. Netw.*, vol. 18, no. 3, pp. 960–972, Jun. 2010.

[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, pp. 1–12.

[15] A. Jindal and K. Psounis, "The achievable rate region of 802.11-scheduled multihop networks," *IEEE/ACM Trans. Netw.*, vol. 17, no. 4, pp. 1118–1131, Aug. 2009.

[16] C. Kai and S. Zhang, "Throughput analysis of CSMA wireless networks with finite offered-load," in *Proc. IEEE ICC*, Jun. 2013, pp. 6101–6106.

[17] P. Gupta and P. R. Kumar, "The capacity of wireless networks," *IEEE Trans. Inf. Theory*, vol. 46, no. 2, pp. 388–404, Mar. 2000.

[18] S. Toumpis and A. J. Goldsmith, "Capacity regions for wireless ad hoc networks," *IEEE Trans. Wireless Commun.*, vol. 2, no. 4, pp. 736–748, Jul. 2003.

[19] K. Jain, J. Padhye, V. Padmanabhan, and L. Qiu, "Impact of interference on multi-hop wireless network performance," in *Proc. ACM MobiCom*, San Diego, CA, USA, Sep. 2003, pp. 66–80.

[20] M. Kodialam and T. Nandagopal, "Characterizing the capacity region in multi-radio multi-channel wireless mesh networks," in *Proc. ACM MobiCom*, Cologne, Germany, Aug. 2005, pp. 73–87.

[21] C.-K. Chau, M. Chen, and S. C. Liew, "Capacity of large-scale CSMA wireless networks," *IEEE/ACM Trans. Netw.*, vol. 19, no. 3, pp. 893–906, Jun. 2011.

Rafael Laufer (S'09–M'11) received the B.Sc. and M.Sc. degrees in electrical engineering from Universidade Federal do Rio de Janeiro (UFRJ), Rio de Janeiro, Brazil, in 2003 and 2005, respectively, and the Ph.D. degree in computer science from the University of California, Los Angeles, CA, USA, in 2011.

He is currently a Member of Technical Staff at Bell Laboratories, Alcatel-Lucent in Holmdel, NJ. During his graduate studies, he worked at Bell Labs, Cisco, Technicolor, and EPFL.

Dr. Laufer was the recipient of the Marconi Society's Young Scholar Award in 2008 in recognition of outstanding academic achievement.

Leonard Kleinrock (M'64–SM'71–F'73–LF'97) received the B.E.E. degree from the City College of New York, New York, NY, USA, in 1957, and the M.S. and Ph.D. degrees from the Massachusetts Institute of Technology (MIT), Cambridge, in 1959 and 1963, respectively.

He is a Distinguished Professor of computer science with the University of California, Los Angeles, where he has been since 1963, serving as a Chairman of the department from 1991 to 1995. He has published over 250 papers and authored six books on a wide array of subjects. During his tenure at UCLA, he supervised the research for 48 Ph.D. students and numerous M.S. students.

Dr. Kleinrock is a member of the National Academy of Engineering, the American Academy of Arts and Sciences, an INFORMS fellow, an IEC fellow, a Guggenheim fellow, and a founding member of the Computer Science and Telecommunications Board of the National Research Council. His work was recently recognized when he received the 2007 National Medal of Science, the highest honor for achievement in science bestowed by the President of the United States. Among his many other honors, he is the recipient of the L.M. Ericsson Prize, the NAE Charles Stark Draper Prize, the Marconi International Fellowship Award, the Dan David Prize, and the Okawa Prize.