2011 50th IEEE Conference on Decision and Control and
European Control Conference (CDC-ECC)
Orlando, FL, USA, December 12-15, 2011

Performance bounds for CSMA-based Medium Access Control

Nikolaos M. Freris

Abstract— **In recent work, it was shown that the class of distributed random access MAC schemes leveraging Carrier-Sense Multiple Access (CSMA) is throughput-optimal. To fully assess the potential of such schemes, it is challenging to study their performance in terms of mean delays and compare it against that of centralized scheduling.**

In this paper, we present upper and lower bounds on the performance of CSMA-based random access MAC. We modify the ideal CSMA model to obtain one that further incorporates queue length information and admits a product-form stationary distribution. We analytically calculate its mean delay at the steady-state, and show that it yields an upper bound on the delay of ideal CSMA. We also derive a lower bound which is independent of the scheduling algorithm. The derived bounds coincide with the best known bounds for the popular max-weight scheduling, whence the performance of such distributed low-complexity schemes lies in the same regime as that of the centralized, generally NP-hard max-weight scheduling. Our results extend to slotted systems, as well as to a wide range of arrival-service processes. Finally, we develop a method for deriving upper and lower bounds on the performance of MAC algorithms by use of Linear Programming (LP) and present comparative simulation results.

Keywords: Wireless Networks, Medium Access Control (MAC), Carrier-Sense Multiple Access (CSMA), Random-access MAC, Resource allocation, Distributed algorithms

I. INTRODUCTION

In a wireless ad hoc network nodes forward packets in a multi-hop fashion in order to deliver data to the desired destinations. The decentralized nature of such systems requires developing efficient distributed algorithms for resource allocation: power control, rate control, and Medium Access Control (MAC), so as to provide high throughput, low latencies and service differentiation to competing data flows. This in turn requires the cooperation of the different network layers; the transport layer is responsible for injecting traffic within the capacity region [1] of the network based on measured congestion; the MAC layer is responsible for scheduling transmissions in order to serve the injected traffic by minimizing collisions among concurrent transmissions. In the realm of wireless, the traditional network stack which separates routing, flow control, scheduling, and power control is not optimal. A cross-layer design featuring the coupling of the transport layer with the Network and MAC layers through queue-length information-sharing was shown to be optimal [2-7] via the Network Utility Maximization (NUM) framework.

The bottleneck of the cross-layer design lies at the MAC layer [6] which requires fast and efficient control of transmissions. There are two main approaches for MAC: a) *deterministic scheduling*, when the medium access controller selects the set of links to be activated at each time slot, typically in a centralized fashion by taking into account information about the queue lengths corresponding to different flows [3-6], and b) *random access*, where decisions are "soft," i.e., each link is activated with a certain probability defined by the medium access controller; such schemes are typically distributed and again, the controller considers queue-length information in assigning probabilities [7], [8], [9].

An increasingly popular approach to multiple access is based on *Carrier-Sense Multiple Access (CSMA)* random access algorithms. In CSMA, a sending node can sense whether the channel is busy before it initiates a packet transmission; if it detects the channel to be busy, it waits for a random amount of time (called *back-off* time) before it attempts to reserve the channel again, otherwise it makes the transmission. CSMA schemes are readily implementable in a decentralized fashion and are widely used in practice (e.g., the four-phase handshake reservation mechanism of the IEEE 802.11 protocol [7]). It was recently established [8], [9] that the class of CSMA-based random-access MAC protocols is throughput-optimal.

In this work, we use a modification of the ideal CSMA model [8], [10] to obtain a positive recurrent, reversible Markov chain model that incorporates the queue lengths as well as the independent set information and admits a product-form stationary distribution. We show that the mean delay of the new model yields an upper bound on the mean delay of the original model and analytically derive an upper bound on the mean system delay at the steady-state. We also derive a fundamental lower bound which is independent of the scheduling algorithm. The main result is that the derived bounds coincide with those for the popular max-weight scheduling algorithm [12]. We also develop another method for obtaining upper and lower bounds on the mean total delay of MAC protocols via solving linear programs (LPs). Our results naturally extend to the slotted discrete-time Markov model of [9], as well as to a wide range of arrival-service processes through the analysis of [10]. Finally, we briefly present simulation results for two different topologies and two arrival rates in each case.

II. RELATED LITERATURE

It is well known that *Maximum-Weight Scheduling (MWS)* [1] is *throughput-optimal*, in the sense that it can stabilize the network queues for all arrival rates in the capacity region, without knowledge of their values. However, MWS amounts to solving a complex combinatorial optimization

Nikolaos M. Freris is with IBM Research - Zürich, Säumerstrasse 4, 8803 Rüschlikon, Switzerland. E-mail: nif@zurich.ibm.com

problem in each time slot; the problem is by nature centralized and additionally NP-hard [6] for general interference models. As a consequence, MWS is impractical for many actual implementations, which has motivated research for efficient decentralized MAC protocols.

A low-complexity alternative to MWS is *Maximal Scheduling (MS)*, which can typically support only a fraction of the capacity region [13]. Yet another low-complexity alternative is *Greedy Maximal Scheduling (GMS)*, which is throughput-optimal if the network satisfies the *local-pooling* condition [14]. A distributed greedy protocol similar to IEEE 802.11 was shown to support only a fraction (which depends on the network topology and the interference model) of the capacity region even if collisions are ignored [15].

Jiang and Walrand [8] established that the ideal distributed CSMA random access scheme [10] is throughput-optimal; they considered a continuous-time Markov chain model which ignores collisions, and is also controllable by the mean values of the back-off times. They developed a distributed iterative scheme for adjusting the control parameters so that the steady-state service rate can support any arrival rate vector in the capacity region. Ni and Srikant [9] proposed a slotted protocol with minimal communication overhead to avoid collisions, and showed that it attains the same steady-state distribution as ideal CSMA, hence is throughput-optimal. Last but not least, an upper bound on the mean delay of CSMA was derived by Ni et al. [11] who considered the effect of the mixing time of the underlying Markov chains in calculating mean total delays.

III. Network Model

We model a single-channel wireless network as an (undirected) *interference graph* $G = (V, E)$, where V represents the set of wireless links and $(k, l) \in E$ if links k and l cannot be simultaneously active. We denote the total number of links by $K := |V|$, and consider a link enumeration $k = 1, \dots, K$. A *feasible schedule* is a subset of links that can be simultaneously active without any two links interfering with one another, i.e., an *independent set* [16] of G. Let $\mathcal{I}$ denote the set of independent sets, and let $N := |\mathcal{I}|$. We will consider the enumeration $\mathcal{I} = \{\mathbf{x}^i\}_{i=1}^N$, where each independent set is represented by a vector $\mathbf{x}^i \in \{0,1\}^K$ with the $k-$th entry, x_k^i, equal to 1 if the $k-$th link is active, and equal to 0 else. With some abuse of notation, we also treat $\mathbf{x}^i$ as a set and write $k \in \mathbf{x}^i$ if $x_k^i = 1$; by convention we set $\mathbf{x}^1 := \varnothing$. A scheduling algorithm is a method of deciding which independent set to activate at each time instant.

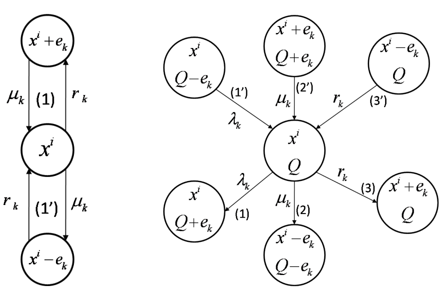
We consider K single-hop flows, one at each link k, and assume that the arrivals at each link are independent Poisson processes with mean arrival rate vector $\boldsymbol{\lambda} \in \mathbb{R}_+^K$, while the service times are independent exponentially distributed random variables with mean service rate vector $\boldsymbol{\mu} \in \mathbb{R}_+^K$. [1] The unsent packets over a link k are stored in a queue whose length is denoted by Q_k; the vector of all queue lengths is $\boldsymbol{Q} := [Q_k]_{k=1}^K \in \mathbb{N}^K$. The *capacity region* of the network is defined as the set of all arrival rates for which there exists a scheduling algorithm such that the evolution of queue lengths is described by a positive recurrent Markov chain (MC). This is the set Λ defined by [1]:

$$\Lambda \quad := \quad \{\boldsymbol{\lambda} \in \mathbb{R}_+^K : \boldsymbol{\lambda} < \sum_{i=1}^N p_i \mathbf{x}^i \cdot \boldsymbol{\mu}\}, \tag{1}$$

for some $\boldsymbol{p} \in \mathbb{R}_+^N : \sum_{i=1}^N p_i \leq 1$, where $\cdot$ denotes entry-by-entry multiplication of two vectors and the inequality is also interpreted element-wise. We call a vector in Λ *strictly feasible* and one in the closure of the capacity region, $\bar{\Lambda}$, as *feasible*. Note that $\Lambda, \bar{\Lambda}$ are convex sets that can be fully characterized by using only *maximal* independent sets in (1).

A. Model for ideal CSMA

The ideal CSMA model [8], [10] assumes that a link can be active *only if* its corresponding queue is non-empty. In this model, the *activation set*, i.e., the set of links that are activated by MAC at a given time, is always a feasible schedule and there are no collisions. For a given time, let the activation set be $\mathbf{x}^i$; completion of service at a link $k \in \mathbf{x}^i$ renders the link inactive and the system moves to $\mathbf{x}^i - \mathbf{e}_k$, where the vector $\mathbf{e}_k \in \{0,1\}^K$ has all entries but the $k-th$ one equal to 0, and its $k-th$ entry is equal to 1. In the case that a) link $k \notin \mathbf{x}^i$ (k is inactive), b) its corresponding queue is non-empty, and c) $\mathbf{x}^i + \mathbf{e}_k$ is a feasible schedule, link k will be activated after a *back-off* time exponentially distributed with mean $\frac{1}{r_k}$; this corresponds to a transmitter sensing the channel for interfering ongoing transmissions and backing-off for an exponential waiting time before initiating transmission, only when it senses the channel idle. If we further assume that the system is *fully backlogged* [8] (i.e., that all queues are non-empty, or equivalently lower-bounded by 1) and define the activation set as the system state, then the system evolution is captured by a continuous-time finite Markov chain shown in Figure 1(a).



(a) Activation set Markov chain

(b) Joint queue-length and activation set Markov chain

Fig. 1. Idealized CSMA model.

IV. Performance Bounds for CSMA

For the purpose of performance analysis, we define Markov chains that incorporate queue-lengths in addition

[1]Our results generalize to multi-hop traffic by means of the analysis in [8] as well as to non-Markovian models through the analysis of [10]

to the activation set information. We develop a modified protocol which admits a product-form distribution, and show that its mean delay yields an upper bound on the mean delay of CSMA. We analytically calculate the joint stationary distribution for queue-lengths and activation set of the modified protocol, and use it to obtain an upper bound on the mean delay of ideal CSMA at the steady-state. We also derive a fundamental lower bound based on the fact that for each clique of the interference graph at most one link can be activated at a given time. Finally, we derive linear programs (LPs) to obtain upper and lower performance bounds for CSMA-based random-access schemes, as well as for MWS.

A. Ideal CSMA

It is not hard to verify that the Markov chain (MC) described in Figure 1(a) is irreducible. The detailed balance equations [17] are satisfied and its stationary distribution is a Markov random field (MRF) given by:

$$P(\mathbf{x}^i) = \frac{1}{C}\prod_{k:k\in\mathbf{x}^i}\bar{r}_k, \tag{2}$$

where $\bar{r}_k := \frac{r_k}{\mu_k}$ and $C := \sum_{\mathbf{x}^i\in\mathcal{I}}\prod_{k:k\in\mathbf{x}^i}\bar{r}_k$. The main result in [8] is that the map $\boldsymbol{r}\in\mathbb{R}_+^K \to \Lambda$ is *onto*, hence the ideal CSMA model is throughput-optimal.

Let us consider the composite state comprising the activation set $\mathbf{x}^i$ and the queue-length vector $\mathbf{Q}\in\mathbb{N}^K$. Under the Poisson-Exponential assumption for the arrival, service and back-off time distributions, the state evolution is captured by the MC shown in Figure 1(b). The set of feasible states is $\{(\mathbf{x}^i,\mathbf{Q}) : \mathbf{x}^i\in\mathcal{I}, \mathbf{Q}\in\mathbb{N}^K, \mathbf{x}_k^i = 1 \Rightarrow Q_k > 0\}$.

From state $(\mathbf{x}^i,\mathbf{Q})$, the system can make a transition to:

- state $(\mathbf{x}^i,\mathbf{Q}+\mathbf{e}_k)$, for any link $k\in V$, if there is an arrival at link k; this happens with rate λ_k (cf. (1) in Figure 1(b)).
- state $(\mathbf{x}^i-\mathbf{e}_k,\mathbf{Q}-\mathbf{e}_k)$, for any link $k\in\mathbf{x}^i$, if there is a service completion at link k; this happens with rate μ_k (cf. (2) in Figure 1(b)).
- state $(\mathbf{x}^i+\mathbf{e}_k,\mathbf{Q})$, for any link $k\notin\mathbf{x}^i$ such that $\mathbf{x}^i+\mathbf{e}_k\in\mathcal{I}$, if the wait period (as defined by the random back-off value) is concluded; this happens with rate r_k (cf. (3) in Figure 1(b)).

In Figure 1(b) we also show all possible transitions to a feasible state $(\mathbf{x}^i,\mathbf{Q})$; these are simply the counterparts of the three cases above (respectively (1'), (2'), (3')). We summarize some properties of the"ideal CSMA MC", in the next lemma.

Lemma 1 (Properties of the ideal CSMA MC): The ideal CSMA Markov chain is irreducible but admits no product-form stationary distribution such that the queue lengths are independent from one another and from the activation set; this is true whether the system is assumed fully backlogged or not.

Proof: To establish irreducibility, consider two feasible states $(\mathbf{x}^i,\mathbf{Q}^1)$ and $(\mathbf{x}^j,\mathbf{Q}^2)$; we will construct a path of feasible transitions connecting the two states. First, consider a series of arrivals (if necessary) such that the new state is $(\mathbf{x}^i,\mathbf{Q}^3)$ where $Q_k^3 = Q_k^2$ for all $k\in V$ s.t. $Q_k^2 < Q_k^1$, and $Q_k^3 = Q_k^1$, otherwise, whence we have $\mathbf{Q}^3 \geq \mathbf{Q}^2$. There is a path comprising of arrivals to followed by departures from all links $k\in\mathbf{x}^i$ that moves the system to state $(\varnothing,\mathbf{Q}^3)$. Then, via a series of successive transitions to single-link independent sets $\{k\}$ for all $k : Q_k^3 > Q_k^2$ followed by departures from these links, the system can move to $(\varnothing,\mathbf{Q}^2)$. The system can move to $(\mathbf{x}^j,\mathbf{Q}^2)$ via $\sum_{k=1}^K \mathbf{x}_k^j$ transitions corresponding to wait period completions.

The queue-length process $\mathbf{Q}$ is not independent from the activation set process $\mathbf{x}$ since $P(\mathbf{x}^i|\mathbf{Q}=\mathbf{0}) = 0 < P(\mathbf{x}^i)$, for any $\mathbf{x}^i \neq \varnothing$, where $\mathbf{0}$ denotes the all-zero vector. For the case where the system is fully backlogged, let us suppose by contradiction that $P(\mathbf{x}=\mathbf{x}^i,\mathbf{Q}) = P(\mathbf{x}=\mathbf{x}^i)P(\mathbf{Q}) = P(\mathbf{x}=\mathbf{x}^i)\prod_{k\in V}P(Q_k=q_k)$. The global balance equations yield

$$\begin{aligned}\sum_{k\in K_1}\lambda_k P(\mathbf{Q}-\mathbf{e}_k)P(\mathbf{x}^i) + \sum_{k\in K_2}\mu_k P(\mathbf{Q}+\mathbf{e}_k)P(\mathbf{x}^i+\mathbf{e}_k)\\ + \sum_{k\in\mathbf{x}^i} r_k P(\mathbf{Q})P(\mathbf{x}^i-\mathbf{e}_k)\\ = P(\mathbf{x}^i)P(\mathbf{Q})\Big(\sum_{k\in V}\lambda_k + \sum_{k\in\mathbf{x}^i\cap K_1}\mu_k + \sum_{k\in K_2} r_k\Big),\end{aligned} \tag{3}$$

where $K_1 := \{k : Q_k > 1\}, K_2 := \{k : k\notin\mathbf{x}^i, \mathbf{x}^i+\mathbf{e}_k\in\mathcal{I}\}$. Considering a state $(\mathbf{x}^i,\mathbf{Q})$ with $\mathbf{Q}=\mathbf{1}$ (the all-one vector) and $\mathbf{x}^i$ maximal schedule (whence $K_1 = K_2 = \varnothing$) and applying (2), we get in the light of (2):

$$\sum_{k\in V}\lambda_k = \sum_{k\in\mathbf{x}^i}\mu_k, \tag{4}$$

which does not hold true for each $\boldsymbol{\lambda}\in\Lambda$. ■

B. Modified CSMA model

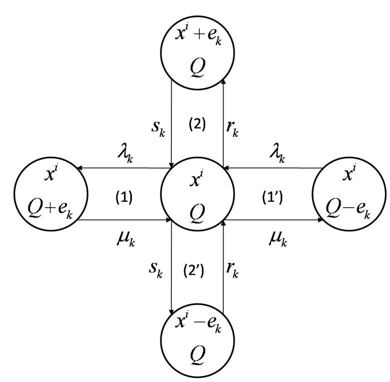


Fig. 2. Modified CSMA: Joint queue-length and activation set Markov chain.

For performance analysis purposes, we introduce a modified CSMA-based scheme, which we call *Performance-CSMA (PCSMA)*.[2] PCSMA has three main differences from the ideal CSMA. First, upon service completion at a given link, the link is not "released" but remains active. Second, an active link k is released after a random amount of time which is exponentially distributed (independent of all other processes) with mean $\frac{1}{s_k}$, whether service at the link has

[2]Note that this scheme does not necessarily constitute a proposal for actual implementation, but rather a tool for deriving bounds on the performance of ideal CSMA.

been completed or not. Third, an arrival at link k is only admitted into the system if link k is currently active, and is discarded otherwise. PCSMA is described by a Markov chain with state-space $\{(\mathbf{x},\mathbf{Q}) : \mathbf{x} \in \mathcal{I}, \mathbf{Q} \in \mathbb{N}^K\}$ (see also Figure 2).

From state $(\mathbf{x}^i,\mathbf{Q})$, the system can make a transition to:

- state $(\mathbf{x}^i,\mathbf{Q}+\mathbf{e}_k)$, for any link $k \in \mathbf{x}^i$, if there is an arrival at link k; this happens with rate λ_k (cf. (1) in Figure 2).
- state $(\mathbf{x}^i,\mathbf{Q}-\mathbf{e}_k)$, for any link $k \in \mathbf{x}^i : Q_k > 0$, if there is a service completion at link k; this happens with rate μ_k (cf. (1') in Figure 2).
- state $(\mathbf{x}^i+\mathbf{e}_k,\mathbf{Q})$, for any link $k \notin \mathbf{x}^i, \mathbf{x}^i+\mathbf{e}_k \in \mathcal{I}$, if the wait period (as defined by its random back-off value) is concluded; this happens with rate r_k (cf. (2) in Figure 2).
- state $(\mathbf{x}^i-\mathbf{e}_k,\mathbf{Q})$, for any link $k \in \mathbf{x}^i$, if the release time (as defined by its random release value) is reached; this happens with rate s_k (cf. (2') in Figure 2).

In PCSMA the queue lengths and independent sets are, by construction, decoupled. Note that both PCSMA and ideal CSMA feature *idling*, in the sense that they spend a non-zero fraction of time in non-maximal independent sets. However, unlike ideal CSMA, PCSMA has the additional property of *wasted resources*, since a link can be active even if its queue is empty. Finally, PCSMA has the additional feature of discarding arrivals at non-active links.

Theorem 1 (Properties of PCSMA):

1) The Markov chain of PCSMA is irreducible.
2) Let $\boldsymbol{\lambda} < \boldsymbol{\mu}$, then the Markov chain is positive-recurrent and reversible. Its stationary distribution is

$$P(\mathbf{x}^i,\mathbf{Q}) = \frac{1}{C}\prod_{k:k\in\mathbf{x}^i}\bar{r}_k\prod_{k\in V}(1-\rho_k)\rho_k^{Q_k}, \quad (5)$$

where $\bar{r}_k := \frac{r_k}{s_k}, C := \sum_{\mathbf{x}^i\in\mathcal{I}}\prod_{k:k\in x^i}\bar{r}_k$, and for each link $k \in V$, $\rho_k := \frac{\lambda_k}{\mu_k}$. In particular, the queues are mutually independent M/M/1 queues, independent from the activation set process whose distribution is the same as that of ideal CSMA.

3) If we consider the *effective* mean arrival rate into the system, by ignoring discarded arrivals, then PCSMA is throughput-optimal.

Proof: Consider two feasible states $(\mathbf{x}^i,\mathbf{Q}^1)$ and $(\mathbf{x}^j,\mathbf{Q}^2)$. We construct a path of feasible transitions connecting the two states. First, consider a series of release events moving the system to $(\varnothing,\mathbf{Q}^1)$. Then, by successively capturing each link k with $Q_k^1 < Q_k^2$ and considering an arrival followed by a link release, and by successively capturing each link l with $Q_l^1 > Q_l^2$ and considering a departure followed by link release, we can move the system to $(\varnothing,\mathbf{Q}^2)$. Finally, a series of $\sum_{k=1}^K x_k^j$ transitions corresponding to wait period completions moves the system to $(\mathbf{x}^j,\mathbf{Q}^2)$. This proves the first part. The distribution (5) satisfies the detailed balance equations whence the second part follows.

The *effective* mean arrival rate $\bar{\boldsymbol{\lambda}}$ into the system (after ignoring discarded arrivals) is given by

$$\bar{\boldsymbol{\lambda}} := \sum_{\mathbf{x}^i\in\mathcal{I}} p_i\boldsymbol{\lambda}\cdot\mathbf{x}^i = \boldsymbol{\lambda}\cdot\sum_{\mathbf{x}^i\in\mathcal{I}} p_i\mathbf{x}^i, \quad (6)$$

where $p_i := P(\mathbf{x}^i)$ is the activation set stationary distribution (2). Consider $\bar{\boldsymbol{\lambda}} \in \Lambda$ arbitrary. By the definition of Λ, there exists $\epsilon > 0$ sufficiently small such that $(1+\epsilon)\bar{\boldsymbol{\lambda}} \in \Lambda$. From the result in [8], there exists $\bar{\mathbf{r}} \in \mathbb{R}_+^K$, such that the stationary distribution satisfies $(1+\epsilon)\bar{\boldsymbol{\lambda}} = \boldsymbol{\mu}\cdot\sum_{\mathbf{x}^i\in\mathcal{I}} p_i(\bar{\mathbf{r}})\mathbf{x}^i$. Picking $\boldsymbol{\lambda} = \frac{1}{1+\epsilon}\boldsymbol{\mu}$ and $\mathbf{r},\mathbf{s}$ such that $\frac{r_k}{s_k} = \bar{r}_k$ for all $k \in V$ proves the third part. ∎

C. Upper bound on delays

Let us represent an ideal CSMA setup by the tuple CSMA$(\boldsymbol{\lambda},\boldsymbol{\mu},\boldsymbol{r})$ and a PCSMA setup as PCSMA$(\boldsymbol{\lambda},\boldsymbol{\mu},\boldsymbol{r},\boldsymbol{s})$. We will show that the latter gives an upper bound on the delay of the former, which can be analytically calculated through the product-form distribution (5).

Theorem 2 (Upper bound on mean delay of ideal CSMA): The mean delay of PCSMA yields an upper bound on the mean delay of ideal CSMA. In particular, for each link $k \in V$, an upper bound on the mean delay $\mathbb{E}[D_k]$ is given by:

$$\mathbb{E}[D_k] \leq \frac{1}{\mu_k - \lambda_k^{in}}, \quad (7)$$

where

$$\lambda_k^{in} := \frac{\lambda_k}{\sum_{i\in\mathcal{I}} p_i(\boldsymbol{r})x_k^i}, \quad (8)$$

is the *injected* arrival rate used to attain an *effective* arrival rate $\boldsymbol{\lambda} \in \Lambda$. An upper bound on mean system delay of CSMA $\mathbb{E}[D]$ is given by:

$$\mathbb{E}[D] \leq \frac{1}{\sum_{k\in V}\lambda_k}\sum_{k\in V}\frac{\rho_k^{in}}{1-\rho_k^{in}}, \quad (9)$$

where $\rho_k^{in} := \frac{\lambda_k^{in}}{\mu_k}$.

Proof: The proof is by a stochastic coupling argument similar to [10]. First consider a variant of the ideal CSMA protocol, where $(\mathbf{x}^i,\mathbf{Q})$ is feasible even if $Q_k = 0$, for some $k \in \mathbf{x}^i$, that is to say the system has the "wasted resources" property of PCSMA. It is clear that the mean delay of this variant (which we call CSMAv) yields an upper bound on the mean delay of ideal CSMA. Now let us consider PCSMA$(\boldsymbol{\lambda}^{in},\boldsymbol{\mu},\boldsymbol{r},\boldsymbol{\mu})$ and CSMAv$(\boldsymbol{\lambda},\boldsymbol{\mu},\boldsymbol{r})$. For a sufficiently large interval $[0,T]$, the two systems spend the same amount of time, modulo an $o(T)$ term (where $\lim_{T\to\infty}\frac{o(T)}{T} = 0$), on each independent set. They also make the same number of transitions between adjacent (i.e., differing by a single link) independent sets modulo an $o(T)$ term. The number of arrivals at link k is $\lambda_k T + o(T)$ in both systems, where for PCSMA we consider the number of *accepted* arrivals. Finally, the number of service completions from a particular independent set is the same for both systems, modulo an $o(T)$ term. Combining these arguments and Little's law, we get that the mean delay experienced at each link is the same for both systems. The mean number of customers in queue for link k is given by $\frac{\rho_k^{in}}{1-\rho_k^{in}}$ [17] , and the upper bounds on mean delays (7), (9) are a simple application of Little's law. ■

Remark 1: It was shown in [10] that the independent set stationary distribution (2) is insensitive to the inter-arrival and inter-service time distributions, as long as the arrival and service processes are modeled as renewal processes. Based on the natural decoupling of the independent set process from the queue-length process, we can extend the upper bound (7) to more general queueing models, e.g., consider G/G/1 queues. In particular, consider independent *injected* arrival, service and back-off renewal processes with mean values $(\boldsymbol{\lambda}^{in}, \boldsymbol{\mu}, \boldsymbol{r})$. The stationary distribution for the PCSMA variant is

$$P(x^i, Q) = \frac{1}{C} \prod_{k:k\in x^i} \bar{r}_k \prod_{k\in V} P_k(Q_k, \lambda_k^{in}, \mu_k), \qquad (10)$$

where $\bar{r}_k := \frac{r_k}{\mu_k}, C := \sum_{\mathbf{x}^i\in\mathcal{I}} \prod_{k:k\in x^i} \bar{r}_k$, and $P_k(Q_k, \lambda_k^{in}, \mu_k)$ denotes the stationary distribution corresponding to the particular queueing model used at the $k-$th link. The mean delay for CSMA is upper bounded by

$$\mathbb{E}[D_k] \le D_k(\lambda_k^{in}, \mu_k), \qquad (11)$$

where $D_k(\lambda_k^{in}, \mu_k)$ is the mean delay of the $k-$th link as it can be calculated by its corresponding queueing model, and similarly for the mean system delay $\mathbb{E}[D]$.

Remark 2 (Upper bound for slotted system): Let us consider a G/D/1 queueing model [17] for arrival/service at each link; this corresponds to a *slotted* system with random arrivals and deterministic service times. We further assume that at each slot one customer is served at each active queue, i.e., $\boldsymbol{\mu} = \mathbf{1}$. Let us denote the number of arrivals (at a given slot) at link k by A_k with mean λ_k and variance $Var(A_k)$. Then for $\boldsymbol{\lambda} \in \Lambda$, the upper bound for PCSMA corresponding to $G/D/1$ is [17]:

$$\mathbb{E}[D] \le \quad \sum_{k\in V} \frac{\lambda_k + Var(A_k) - \lambda_k^2}{2(\bar{\mu}_k - \lambda_k)(\sum_{k\in V}\lambda_k)}, \qquad (12)$$

where $\bar{\boldsymbol{\mu}}$ is an arbitrary vector in the capacity region that can be achieved by a proper selection of $\boldsymbol{r}, \boldsymbol{s}$. Note also that for the M/D/1 (as well as M/M/1) queueing model, the bound becomes (9). This bound is tight, in the sense that it goes to infinity only if $\boldsymbol{\lambda}$ converges to the boundary of the capacity region. The important observation is that this upper bound coincides with the upper bound derived in [12] for MWS. This implies that the performance of the distributed, efficient random access CSMA MAC lies in the same regime as that of the centralized, generally NP-hard MWS. This is a promising result on the performance of such schemes.

D. Lower bound on delays

We derive a fundamental lower bound on the performance of CSMA-based MAC based solely on the collision-free assumption; this is independent of the particular scheduling algorithm (deterministic or random access).

Let us first consider the case of a complete interference graph, where any two links mutually interfere. For simplicity, we assume $\mu_k = 1$, for each $k \in V$. Since at most one link can be active at each time instant, the network, as a whole, behaves like an M/M/1 queue with arrival rate $\lambda := \sum_{k\in V} \lambda_k$ and service rate equal to $\mu = 1$, whence the mean number of customers in the network is given by $\mathbb{E}[\sum_{k\in V} Q_k] = \frac{\lambda}{1-\lambda}$.

Now let us consider a partition of the links into cliques of the interference graph. By ignoring the interference between links that belong in different cliques, we can obtain a lower bound on mean delay. Let us denote the finite set of all *clique-partitions*, i.e., partitions where each subset is a clique, by $\mathcal{P}$; an atom of this set is a partition $\mathbf{P} = \{P_j\}_{j=1}^{|\mathbf{P}|}$ where each P_j is a clique. The lower bound on mean delay is then given by:

$$\mathbb{E}[D] \ge \max_{\mathbf{P}\in\mathcal{P}} \sum_{P_j\in\mathbf{P}} \frac{\sum_{k\in P_j}\lambda_k}{(1-\sum_{k\in P_j}\lambda_k)(\sum_{k\in V}\lambda_k)}. \qquad (13)$$

This lower bound is *tight*, e.g., it is exact for a network that is a union of non-interfering cliques. Following the same principle, we can derive fundamental lower bounds for other queueing models e.g., the G/D/1 as in [12].

E. Linear programs for performance analysis

An alternative approach for obtaining upper and lower bounds on the performance of CSMA-based queueing systems can be derived following the theory developed in [18], [19]; We derive LPs for the performance analysis of CSMA-based MAC schemes, both deterministic and random access. For a given time instant, let $w_k = 1$ if link k is active, and 0 else. Let $v_i = 1$ if the system is at the $i-$th independent set, and 0 else. Clearly, $\sum_{i=1}^N v_i = 1$, and we also have

$$w_k = \sum_{i:k\in\mathbf{x}^i} v_i. \qquad (14)$$

If we assume that the scheduling algorithm is *non-idling*, i.e., that the system cannot be at an independent set $\mathbf{x}^i$ if there exists some $\mathbf{x}^j \supset \mathbf{x}^i$ and $k \in \mathbf{x}^j\setminus\mathbf{x}^i$ with $Q_k > 0$, we obtain the following *maximality* constraint:

$$v_i \sum_{k\in\mathbf{x}^j\setminus\mathbf{x}^i} Q_k = 0 \;\; \forall \mathbf{x}^j \supset \mathbf{x}^i, \mathbf{x}^j \ne \mathbf{x}^i. \qquad (15)$$

In particular, the activation set cannot be null when $\sum_{k\in V} Q_k > 0$, so we get

$$Q_k = \sum_{i=2}^N Q_k v_i. \qquad (16)$$

By assuming that a stationary distribution exists (see [19] for how to proceed in a self-contained way without making assumptions) and considering the second moments of the queue-length process, we can get the following linear constraints that the system needs to satisfy at steady-state:

$$\lambda_k \sum_{i=2}^N y_{ik} + \lambda_k - \mu_k \sum_{i:k\in\mathbf{x}^i} y_{ik} = 0, \qquad (17)$$

$$\lambda_k \sum_{i=2}^N y_{il} - \mu_k \sum_{i:k\in\mathbf{x}^i} y_{il} + \lambda_l \sum_{i=2}^N y_{ik} - \mu_l \sum_{i:l\in\mathbf{x}^i} y_{ik} = 0, \qquad (18)$$

$$\sum_{k\in\mathbf{x}^j\setminus\mathbf{x}^i} y_{ik} = 0, \qquad (19)$$

where $y_{ik} := \mathbb{E}[v_i Q_k]$, The constraints (17), (18) hold for $k \neq l$ and are derived via Lippman's uniformization [18]. The constraints (19) hold for all $\mathbf{x}^j \supset \mathbf{x}^i \in \mathcal{I}$, and correspond to the non-idling property (15).

If we consider the MWS algorithm we get the additional constraints

$$\sum_{k\in\mathbf{x}^i} y_{ik} \geq \sum_{k\in\mathbf{x}^j} y_{ik} \quad \forall \mathbf{x}^j \neq \mathbf{x}^i. \tag{20}$$

It follows from (16) that $\mathbb{E}[Q_k] = \sum_{i=2}^{N} y_{ik}$, so we can find lower/upper bounds on the mean number of customers by numerically solving linear programs with objective function $\sum_{k=1}^{K}\sum_{i=2}^{N} y_{ik}$ subject to nonnegativity constraints $y_{ik} \geq 0$, along with (17), (18), (19); the LPs are feasible if the arrival rate vector lies in the capacity region [19]. For MWS, we need to further consider (20).

There is no ordering relation between the upper (lower) bound derived in the previous section and the ones acquired by LPs. Consequently, both approaches can be exploited for performance analysis; denoting the upper and lower delay bounds derived via solving LPs by UB_{lp}, LB_{lp}, respectively, we have that the mean system delay lies in:

$$\max(LB, LB_{lp}) \leq \mathbb{E}[D] \leq \min(UB, UB_{lp}). \tag{21}$$

V. Simulations

We have performed numerous experiments, with a wide range of network topologies and arrival rates. For length considerations, we present results corresponding to a network of $n = 5$ nodes with $\boldsymbol{\mu} = \mathbf{1}$, and two different interference topologies: a) cycle graph b) star graph. For each topology, we consider two different feasible arrival rate vectors, and summarize the results in Table I. For the cycle graph, UB_{lp} is the lowest upper bound for both cases, while the highest lower bound is equal to LB in the first case, and is LB_{lp} in the second. For the star graph, note that the lower bounds coincide, while UB is the lowest upper bound in both cases. These results illustrate that none of the two lower (upper) bounds derived above dominates the other in general.

TABLE I
Delays bounds

Topology	Arrivals	LB_{lp}	UB_{lp}	LB	UB
Cycle	(0.2,0.3,0.2,0.3,0.2)'	1.97	3.684	2.25	6.187
	(0.1,0.2,0.4,0.2,0.1)'	2.25	3.571	2.04	4.985
Star	(0.8,0.8,0.8,0.8,0.1)'	6.364	16.667	6.364	13.428
	(0.3,0.5,0.6,0.8,0.15)'	9.331	19.441	9.331	13.128

VI. Conclusion

We have introduced a modification of the ideal CSMA model [10] which is described by a positive recurrent, reversible Markov chain model that incorporates the queue lengths as well as the activation set information. The mean delay of such scheme is an upper bound on the mean delay of ideal CSMA and can be calculated analytically. We have also derived a fundamental lower bound that holds for any CSMA-based MAC protocol. These bounds coincide with those for max-weight scheduling, which is a promising result on the performance of the low complexity, distributed, random access MAC. Finally, we have developed a method for obtaining upper and lower performance bounds via solving linear programs (LPs), and have presented comparative simulation results.

References

[1] L. Tassiulas and A. Ephremides, "Stability properties of constrained queueing systems and scheduling policies for maximum throughput in multihop radio networks," *IEEE Transactions on Automatic Control*, vol. 37(12) Dec. 1992, pp. 1936–1948.

[2] F. P. Kelly, A. Maulloo, and D. Tan, "Rate control in communication networks: shadow prices, proportional fairness and stability," *Journal of the Operational Research Society*, vol. 49, 1998, pp. 237–252.

[3] X. Lin and N. B. Shroff, "Joint rate control and scheduling in multihop wireless networks," in *Proceedings of the IEEE Conference on Decision and Control*, Dec. 2004, pp. 237–252.

[4] M. J. Neely, E. Modiano, and C. Rohrs, "Dynamic power allocation and routing for time varying wireless networks," *IEEE Journal on Selected Areas in Communications*, vol. 23(1) Jan. 2005, pp. 89–103.

[5] A. Eryilmaz and R. Srikant, "Joint congestion control, routing and MAC for stability and fairness in wireless networks," *IEEE Journal on Selected Areas in Communications*, vol. 24, Aug. 2006, pp. 1514–1524.

[6] X. Lin, N. Shroff, and R. Srikant, "A tutorial on cross-layer optimization in wireless networks," *IEEE Selected Areas of Communications*, vol. 14, 2006, pp. 1452–1463.

[7] R. McCabe, N. Freris, and P. R. Kumar, "Controlled random access MAC for network utility maximization in wireless networks," in *Proceedings of the 47th IEEE Conference on Decision and Control*, Dec. 2008, pp. 2350–2355.

[8] L. Jiang and J. Walrand, "A Distributed CSMA Algorithm for Throughput and Utility Maximization in Wireless Networks," in *IEEE/ACM Transactions on Networking*, vol. 18(3), Jun. 2010, pp. 960–972.

[9] J. Ni and R. Srikant, "Distributed CSMA/CA algorithms for achieving maximum throughput in wireless networks," *Information Theory and Applications Workshop (ITA)*, Feb. 2009.

[10] S. C. Liew, C. Kai, J. Leung, and B. Wong, "Back-of-the-envelope computation of throughput distributions in CSMA wireless networks," in *IEEE Transactions on Mobile Computing*, vol. 9(9), Sep. 2010, pp. 1319–1331.

[11] L. Jiang, J. Ni, R. Srikant, and J. Walrand, "Performance bounds of distributed CSMA scheduling," *Information Theory and Applications Workshop (ITA)*, Feb. 2010.

[12] G. R. Gupta and N. B. Shroff, "Delay Analysis for Wireless Networks with Single Hop Traffic and General Interference Constraints," in *IEEE/ACM Transactions on Networking*, vol. 18(2), Apr. 2010, pp. 393–405.

[13] X. Wu, R. Srikant, and J. R. Perkins, "Queue-length stability of maximal greedy schedules in wireless networks," in *IEEE Transactions on Mobile Computing*, Jun. 2007, pp. 595–605.

[14] A. Dimakis and J. Walrand, "Sufficient conditions for stability of longest queue- first scheduling: Second-order properties using fluid limits," in *Advances in Applied Probabilities*, vol. 38(2), 2006, pp. 505–521.

[15] X. Wu, R. Srikant and R. Perkins "Scheduling efficiency of distributed greedy scheduling algorithms in wireless networks," in *IEEE Transactions on Mobile Computing*, vol. 6(6), Jun. 2007, pp. 595–605.

[16] D. West, *Introduction to Graph Theory.* Englewood Cliffs, NJ: Prentice Hall, 1996.

[17] D. Gross, J. Shortle, J. Thompson, and C. Harris *Fundamentals of queueing Theory.* Wiley Series in Probability and Statistics, Fourth Edition, 2008.

[18] S. Kumar and P. R. Kumar, "Performance bounds for queueing networks and scheduling policies," in *IEEE Transactions on Automatic Control*, vol. 39(8), Aug. 1994, pp. 1600–1611.

[19] P. R. Kumar and Sean P. Meyn, "Duality and linear programs for stability and performance analysis of queueing networks and scheduling policies," *IEEE Transactions on Automatic Control*, vol. 41(1), Jan. 1996, pp. 4–17.