Service-Martingales: Theory and Applications to the Delay Analysis of Random Access Protocols

Felix Poloczek
University of Warwick / TU Berlin

Florin Ciucu
University of Warwick

***Abstract*—This paper proposes a martingale extension of effective-capacity, a concept which has been instrumental in teletraffic theory to model the link-layer wireless channel and analyze QoS metrics. Together with a recently developed concept of an arrival-martingale, the proposed *service-martingale* concept enables the queueing analysis of a bursty source sharing a MAC channel. In particular, the paper derives the first rigorous and accurate stochastic delay bounds for a Markovian source sharing either an Aloha or CSMA/CA channel, and further considers two extended scenarios accounting for 1) in-source scheduling and 2) spatial multiplexing MIMO. By leveraging the powerful martingale methodology, the obtained bounds are remarkably tight and improve state-of-the-art bounds by several orders of magnitude. Moreover, the obtained bounds indicate that MIMO spatial multiplexing is subject to the fundamental power-of-two phenomena.**

I. Introduction

Classical works concerned with the throughput and delay analysis of random access protocols (e.g., Aloha or CSMA) rely on strong assumptions. One is that the point process comprising of both newly generated and retransmitted (due to collisions) packets is a Poisson process (Abramson [1], Kleinrock and Tobagi [23], and more recently Yang and Yum [32]). A related assumption is that, at each source, packets arrive as a *blocked* Poisson process, in the sense that *at most* one packet can be backlogged at any source (Tobagi [29] or Beuerman and Coyle [3]); this model is related to the infinite source model in which each source generates a single packet during its lifetime (Lam [24]). Another related and simplifying assumption is to discard the buffered packets at the beginning of a transmission period for a source (Takagi and Kleinrock [27]).

Such conceivably unnatural assumptions enable a tractable analysis but preclude the analysis of realistic bursty sources, i.e., non-Poisson. In particular, the obtained results only capture the *access delay*, and not the other component of the actual delay, i.e., the *queueing delay*. For an elaborate discussion on fundamental drawbacks of ignoring data burstiness in the context of the multiaccess channel, in connection to information theory, see Gallager [17] and Ephremides and Hajek [14].

More recent literature addresses the throughput or delay analysis of the prevalent 802.11 CSMA/CA protocol. Some influential works include Bianchi [4], Cali *et al.* [5], Carvalho and Garcia-Luna-Aceves [6], which share the common assumption of saturated sources (i.e., ignoring burstiness). An approximate queueing analysis accounting for random arrivals is undertaken in Tickoo and Sikdar [28], by approximating the probability of non-empty queues as if the system behaved as an M/M/1 queue. A related approximation of the probability that a source finds itself empty upon a successful transmission is considered by Garetto and Chiasserini [18]. Another work addressing non-saturated arrivals is Alizadeh-Shabdiz and Subramaniam [2]; in addition to enforcing a technical independence assumption from [4], the analysis crucially relies on an M/G/1 approximation of the network.

While such existing results clearly provide valuable insights into the behavior of the notoriously difficult CSMA/CA protocol, the state-of-the-art literature lacks a mathematically rigorous (and also accurate) analysis under random arrivals, especially non-Poisson/bursty. The goal of this paper is to fill this gap by providing the first rigorous and accurate delay analysis in single-hop Aloha and CSMA/CA networks, subject to Markovian arrivals. A crucial feature of the proposed analysis is that it rigorously accounts for buffering and consequently it captures the total (i.e., access plus queueing) delay experienced by a tagged Markovian source.

The starting point of this paper is a recent system theoretic approach to analyze CSMA/CA networks by Ciucu *et al.* [9]. In that work the authors adopt a simplified CSMA/CA model proposed by Durvy et al. [13], which was argued to retain the key features of CSMA/CA. Unfortunately, the analysis relies on the queueing methodology of the stochastic network calculus (see Chang [7], Fidler [15], or Jiang and Liu [20]), which was convincingly shown to lead to very inaccurate results, especially in the case of bursty arrivals (see Ciucu *et al.* [10], [26]); using an alternative martingale queueing methodology, it was further shown that the available inaccurate results could be drastically improved in the case of Markovian arrivals.

In this paper we extend the martingale methodology from Poloczek and Ciucu [26] in order to fit the delay analysis of Aloha and CSMA/CA networks with Markovian arrivals. The novel element of the proposed extension is the concept of a *service-martingale* which models the Markovian service, characteristic to a multiaccess channel such as CSMA/CA, in the martingale domain. By combining service-martingales with the arrival-martingales defined in [26] to model Markovian arrivals, we obtain sharp stochastic bounds on the backlog and delay distributions of a Markovian source over Aloha and CSMA/CA multiaccess channels.

The proposed combination of arrival- and service-martingales parallels the combination of effective bandwidth and effective capacity, which provides an elegant methodology to analyze the queueing behavior in wireless networks. The *effective bandwidth* is an arrival model, defined in terms of moment generating functions, and which is particularly suitable to analyze queues with Markovian and long range dependent arrivals (see Kelly [22]). In turn, the *effective capacity* is an ingenious service model defined in terms of Laplace transforms, and which is particularly suitable to model the channel capacity in wireless scenarios (see Wu and Negi [30], [31]). Alike the conventional network calculus, the elegant queueing methodology based on effective bandwidth/capacity suffers in terms of the accuracy of the produced queueing results; this drawback was convincingly illustrated through simulations by Choudhury *et al.* [8] and analytically by Ciucu *et al.* [10], [26] (e.g., existing bounds can be loose by several orders of magnitude, in the case of Markovian arrivals).

A key benefit of our proposed methodology integrating arrival- and service-martingales is its *modularity*: Indeed, we provide three conceivably straightforward applications to both simple and complex MAC scenarios. The first (simple) scenario is standard and involves the analysis of a tagged bursty source sharing a MAC channel. We then consider two complex extensions by additionally accounting for 1) in-source scheduling, i.e., the tagged source consists of multiple flows scheduled according to a SP (Static Priority) policy before being transmitted over the shared channel, and 2) spatial multiplexing MIMO (multiple-input multiple-output), i.e., the tagged source is transmitted over multiple shared MAC channels. A qualitative insight of the obtained stochastic bounds is that MIMO reduces the delays of bursty sources exponentially (in the number of channels), and, more interestingly, that it is subject to a fundamental power-of-two phenomena.

The rest of the paper is organized as follows. In Section II we introduce the concept of service-martingales, and derive general performance metrics (backlog and delay) for a source modelled by arrival-martingales. In Section III we apply these results to a Markovian tagged source transmitting over Aloha and CSMA/CA channels; numerical results illustrate the remarkable tightness of the obtained stochastic bounds. In Section IV we provide further applications to scenarios with in-source SP scheduling and spatial multiplexing MIMO. Finally, we conclude the paper in Section V.

II. Theory

Consider the single server scenario from Figure 1. A flow A, defined in terms of the bivariate arrival process

$$A(m,n)=\sum_{k=m+1}^{n} a_k \ ,$$

arrives at a server characterized by a service process S, which is defined in bivariate form $S(m,n)$; the corresponding departure process D is also defined in bivariate form. We assume $S(m,n)$ to be driven by a stochastic process $(s_n)_n$, i.e., $S(m,n)$ is $\sigma(s_{m+1},\ldots,s_n)$-measurable. Further, $(a_k)_k$ and $(s_k)_k$ are stationary, ergodic, reversible, and statistically

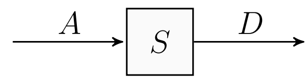

Fig. 1. A server with an arrival process A, service process S, and departure process D

independent. The reversibility assumption is mild, in the sense that Markov processes (i.e., the type of bursty sources we are interested in) can be a-priorily reversed; the concrete Markov processes from this paper are however reversible. For brevity we write $X(n):=X(0,n)$ for any bivariate process X.

The key role of the service process S is to model the service received by A in either a queueing system, or at the data link layer corresponding to some MAC protocol (e.g., Aloha or CSMA/CA); the second case particularly suits this paper. Formally, S couples the arrival and departure processes A and D, respectively, in terms of a $(\min,+)$ convolution, i.e.,

$$D(n)\geq A*S(n):=\min_{0\leq m\leq n}\left\{A(m)+S(m,n)\right\} \ , \quad (1)$$

for all arrival processes A and $n\geq 0$. In other words, the service process plays as similar role as that of an impulse-response in linear and time invariant (LTI) systems, except for the use of an inequality, and not equality, in Eq. (1), due to the lack of $(\min,+)$ linearity; for a related discussion see Ciucu and Schmitt [11].

Next, we give the two central definitions in this paper concerning arrival and service modelling. The first is a slight modification of Definition 3 from [26].

Definition 1 (Arrival-Martingales)**.** *The flow A admits arrival-martingales if for every $\theta>0$ there is a $K_a\geq 0$ and a function $h_a:\mathrm{rng}(a)\rightarrow\mathbb{R}^+$ such that the process*

$$h_a(a_n)e^{\theta(A(n)-nK_a)} \ , \ n\geq 0 \ , \quad (2)$$

is a supermartingale.

In the definition, 'rng' stands for the range operator. The parameters K_a and h_a implicitly depend on θ; the augmented notation $K_a(\theta)$ and $h_a(\theta)$ is omitted for brevity, when clear from the context.

Definition 2 (Service-Martingales)**.** *The service process S admits service-martingales if for every $\theta>0$ there is a $K_s\geq 0$ and a function $h_s:\mathrm{rng}(s)\rightarrow\mathbb{R}^+$ such that the process*

$$h_s(s_n)e^{\theta(nK_s-S(n))} \ , \ n\geq 0 \ , \quad (3)$$

is a supermartingale.

Arrival- and service-martingales relate to each other by a sign change of θ, and closely resemble with the concepts of effective bandwidth and capacity, respectively. The crucial difference is that while the effective bandwidth and capacity are defined in terms of the moment generating function (MGF) and Laplace transform of $A(n)$ and $S(n)$, respectively, the arrival- and service-martingales are defined as stochastic processes and not as (deterministic) numbers, albeit in terms of similar exponential transforms. We point out that retaining the random structure of the arrivals and service is in fact instrumental to a sharp analysis of queueing metrics, as convincingly shown in Ciucu *et al.* [10], [26].

Let us now state an auxiliary definition which will become important in the general proofs of the performance metrics (see Theorems 9 and 10).

Definition 3 (Threshold)**.** *For h_a and h_s as in Definitions 2 and 3 define the threshold*

$$H := \min\{h_a(x)h_s(y) : x - y > 0\} \ .$$

Intuitively, H is the smallest value of $h_a(x)h_s(y)$ such that the instantaneous arrival (i.e., x) is larger than any value of the stochastic process driving the service process (i.e., y).

In the following we make two technical remarks and then give three illustrative examples of arrival- and service-martingales.

Remark 4. *If (2) is a supermartingale, then by stationarity the "time-shifted" process*

$$h_a(a_{n+k})e^{\theta(A(k,n+k)-nK_a)}$$

is also a supermartingale, for some fixed $k \geq 0$.

Remark 5. *If (3) is a supermartingale, then by the reversibility assumption*

$$h_s(s_m)e^{\theta((n-m)K_s-S(m,n))}$$

is also a supermartingale, for every $n \geq 0$, now in the "reversed" time-domain $m \in \{n, n-1, \ldots, 1, 0\}$ (aka a backward-supermartingale).

Example 6 (Constant Arrivals/Service)**.** *Let $x_n \equiv C$ for some constant $C > 0$, and denote generically $A(n) := S(n) := \sum_{k=1}^{n} x_k$. Then A admits arrival-martingales and S admits service-martingales.*

Proof. For $\theta > 0$ and $\theta < 0$, one can choose the corresponding K and h as $K > C$ and $K < C$, respectively, and $h = 1$ (or any other constant). □

Although trivial, these constructions are important as they can model a saturated node (arrival-martingales) and a constant-rate server (service-martingales). More sophisticated examples are given next for i.i.d. (identically and independently distributed) and then for more general Markov-modulated arrivals/service.

Example 7 (i.i.d. Arrivals/Service)**.** *Let $x_1, x_2, \ldots$ be i.i.d. random variables with nonnegative distribution and denote generically $A(n) := S(n) := \sum_{k=1}^{n} x_k$. Then A admits arrival-martingales (for $\theta > 0$ such that $\mathbb{E}[e^{\theta x_1}] < \infty$), and also S admits service-martingales.*

Proof. For $\theta > 0$ let h_a, h_s be constant (without loss of generality take $h_a = h_s = 1$) and K_a and K_s such that

$$\mathbb{E}[e^{\theta x_1}] = e^{\theta K_a} \ \text{ and } \ \mathbb{E}[e^{-\theta x_1}] = e^{-\theta K_s} \ ,$$

respectively. In particular, we have according to the i.i.d. assumption:

$$\begin{aligned}
&\mathbb{E}\left[h(x_{n+1})e^{\theta((n+1)K_s - S(n+1))} \mid x_1, \ldots, x_n\right] \\
&\quad = h(x_{n+1})e^{\theta(nK_s - S(n))} E\left[e^{-\theta x_{n+1}}\right] e^{\theta K_s} \\
&\quad = h(x_n)e^{\theta(nK_s - S(n))} \ ,
\end{aligned}$$

and thus $S(n)$ admits service-martingales. The proof for the arrival-martingales proceeds similarly by a sign change. □

Example 8 (Markov-modulated Arrivals/Service)**.** *Let $(x_n)_n$ be a Markov chain with finite state space $\mathcal{S} = \{0, 1, \ldots, N_{max}\}$ and $f : \mathcal{S} \to \mathbb{R}^+$ be a deterministic function. Then both a Markov-modulated arrival flow and service process, denoted generically by $A(n) := S(n) := \sum_{k=1}^{n} f(x_k)$, admit arrival- and service-martingales, respectively.*

Proof. Let $\theta \in \mathbb{R} \setminus \{0\}$ and let T denote the transition matrix of x_n, i.e.,

$$T_{i,j} = \mathbb{P}(x_{n+1} = j \mid x_n = i) \ .$$

Define the *exponential column-transform* of T^θ by

$$T^\theta_{i,j} := T_{i,j}e^{\theta f(x_j)} \ ,$$

let $sp(T^\theta)$ be its spectral radius, and $h : \mathcal{S} \to \mathbb{R}^+$ be a corresponding right-eigenvector (note that by the Perron-Frobenius Theorem $sp(T^\theta)$ is positive and h can be chosen to be positive).

For arbitrary $K > 0$ we can write:

$$\begin{aligned}
&\mathbb{E}\left[h(x_{n+1})e^{\theta(A(n+1)-(n+1)K)} \mid x_1, \ldots, x_n\right] \\
&\quad = e^{\theta(A(n)-nK)}\mathbb{E}[h(x_{n+1})e^{\theta x_{n+1}} \mid x_n]e^{-\theta K} \\
&\quad = e^{\theta(A(n)-nK)} \left(T^\theta h\right)(x_n)\, e^{-\theta K} \\
&\quad = h\,(x_n)\, e^{\theta(A(n)-nK)} sp(T^\theta)e^{-\theta K} \ ,
\end{aligned} \tag{4}$$

where $\left(T^\theta h\right)(x_n)$ denotes the x_n^{th} component of the vector $T^\theta h$.

In the case when $\theta > 0$ we have from the Perron-Frobenius Theorem:

$$1 < \min_i \sum_j T^\theta_{i,j} \leq sp(T^\theta) \leq \max_i \sum_j T^\theta_{i,j} \leq e^{\theta \max_i f(x_i)} < \infty \ ,$$

implying that we can choose the 'K' (for the arrival-martingales) as any value satisfying $K \geq \frac{\log sp(T^\theta)}{\theta}$, proving thus the supermartingale property in Eq. (4) for the arrival-martingales, i.e.,

$$\begin{aligned}
&\mathbb{E}\left[h(x_{n+1})e^{\theta(A(n+1)-(n+1)K)} \mid x_1, \ldots, x_n\right] \\
&\quad \leq h\,(x_n)\, e^{\theta(A(n)-nK)} \ .
\end{aligned}$$

In turn, when $\theta < 0$, the Perron-Frobenius Theorem yields:

$$1 > \max_i \sum_j T^\theta_{i,j} \geq sp(T^\theta) \geq \max_i \sum_j T^\theta_{i,j} \geq e^{\theta \max f(x_i)} > 0 \ .$$

Moreover, by continuity, we have:

$$\lim_{K \to 0} e^{\theta K} = 1 \ \text{ and } \ \lim_{K \to \infty} e^{\theta K} = \infty \ .$$

We can thus choose the 'K' (for the service-martingales) as any value satisfying $0 < K \leq \frac{\log sp(T^\theta)}{\theta}$, and by a sign change in Eq. (4) the supermartingale property for the service-martingales follows, i.e.,

$$\mathbb{E}\left[h(x_{n+1})e^{-\theta((n+1)K-S(n+1))} \mid x_1,\ldots,x_n\right] \leq h\left(x_n\right)e^{-\theta(nK-S(n))} .$$

□

We note that the core argument from the proof of Example 8, i.e., the exponential column-transform, is due to Duffield [12]; the obtained constructions for arrival- and service-martingales will be used in the Applications sections.

For the rest of this section we assume that the arrival flow A and the service process S admit arrival- and service-martingales, respectively. The corresponding parameters are denoted by K_a and h_a for the arrival-, and by K_s and h_s for the service-martingales. Recall that these parameters implicitly depend on the value of θ.

The performance metrics of interest are the (stationary) backlog distribution, which has the representation

$$Q =_{\mathcal{D}} \sup_{n\geq 0}\{A(n)-S(n)\} ,$$

and the virtual delay at time n, defined by

$$W(n) := \min\{k \geq 0 \mid A(n-k) \leq D(n)\} .$$

Theorem 9 (Backlog)**.** *Assume that the statistically independent processes A and S admit arrival- and service-martingales, respectively. Further, as as stability condition, assume that*

$$\theta^* := \sup\{\theta > 0 : K_a \leq K_s\}$$

and let H as in Definition 3. Then the following backlog bound holds for any $\sigma \geq 0$

$$\mathbb{P}(Q \geq \sigma) \leq \frac{\mathbb{E}[h_a(a_0)]\mathbb{E}[h_s(s_0)]}{H}e^{-\theta^*\sigma} .$$

Proof. Let θ^* as defined, and the corresponding parameters K_a, h_a, K_s, and h_s (all depending on θ^*). By the independence assumption, the process

$$h_a(a_n)h_s(s_n)e^{\theta^*(A(n)-nK_a+nK_s-S(n))}$$

is a supermartingale. As by definition (of θ^*) $K_s - K_a \leq 0$,

$$M(n) := h_a(a_n)h_s(s_n)e^{\theta^*(A(n)-S(n))}$$

is a supermartingale as well. Now define the stopping time N as the first time when $A(n) - S(n)$ exceeds σ, i.e.,

$$N := \min\{n : A(n) - S(n) \geq \sigma\} .$$

Note that $N = \infty$ is possible and $\mathbb{P}(Q \geq \sigma) = \mathbb{P}(N < \infty)$. By the optional stopping theorem applied to the stopping time $N \wedge n := \min\{N, n\}$ (for $n \geq 0$) we have

$$\begin{aligned}\mathbb{E}[h_a(a_0)]\mathbb{E}[h_s(s_0)] &= \mathbb{E}[M(0)] = \mathbb{E}[M(N\wedge n)]\\ &\geq \mathbb{E}[M(N\wedge n)1_{\{N\leq n\}}]\\ &= \mathbb{E}[h_a(a_N)h_s(s_N)e^{\theta^*(A(N)-S(N))}1_{\{N\leq n\}}]\\ &\geq He^{\theta^*\sigma}\mathbb{P}(N \leq n) .\end{aligned}$$

For the last step note that by the minimality of N, $a_N > s_N$ and so with Definition 3: $h_a(a_N)h_s(s_N) \geq H$. The proof completes by letting $n \to \infty$. □

Theorem 10 (Delay)**.** *In the situation of Theorem 9, the following stochastic bound holds for the virtual delay*

$$\mathbb{P}(W(n) \geq k) \leq \frac{\mathbb{E}[h_a(a_0)]\mathbb{E}[h_s(s_0)]}{H}e^{-\theta^*K_sk} .$$

Proof. Let θ^* as defined, and the corresponding parameters K_a, h_a, K_s, and h_s (again, all depending on θ^*). Given the service-process representation from Eq. (1) and the reversibility assumption, we can write:

$$\begin{aligned}&\mathbb{P}\left(W(n) \geq k\right) = \mathbb{P}\left(A(0,n-k) \geq D(n)\right)\\ &\leq \mathbb{P}\left(A(n-k) \geq \min_{0\leq m\leq n}\{A(m)+S(m,n)\}\right)\\ &\leq \mathbb{P}\left(\max_{n\geq k}\{A(k,n)-S(n)\} \geq 0\right)\\ &\leq \mathbb{P}\left(\max_{n\geq k}\{A(k,n)-(n-k)K_a+nK_s-S(n)\} \geq kK_s\right) .\end{aligned}$$

Using Remark 4 and the independence assumption, it follows that

$$h_a(a_n)h_s(s_n)e^{\theta(A(k,n)-(n-k)K_a+nK_s-S(n))}$$

is also a supermartingale (in the time-domain $\{k, k+1, \ldots\}$). Therefore, by invoking the same arguments as in the proof of Theorem 9, the above inequalities continue to:

$$\begin{aligned}\mathbb{P}\left(W(n) \geq k\right) &\leq \frac{\mathbb{E}[h_a(a_k)]\mathbb{E}[h_s(s_k)e^{\theta^*(kK_s-S(0,k))}]}{H}e^{-\theta^*K_sk}\\ &\leq \frac{\mathbb{E}[h_a(a_0)]\mathbb{E}[h_s(s_0)]}{H}e^{-\theta^*K_sk} ,\end{aligned}$$

where we lastly used the stationarity of $(a_n)_n$ and the property that the expectation of supermartingales is non-increasing. □

III. Applications. Backlog and Delay

In this section we apply the previous theoretical results to analyze the queueing performance of a bursty source, denoted by L, and transmitting over an Aloha and CSMA/CA shared channel together with $L-1$ other (saturated) sources (see Figure 2).

In both cases we consider a bursty source L being modelled by a Markov-Modulated On-Off (MMOO) process as in Figure 3: $(a_n)_n$ is a Markov chain with state space $\mathcal{S} = \{0, 1\}$ and associated transition probabilities p_a and q_a.

The transition matrix of the Markov chain is given by

$$T_a = \begin{pmatrix}1-p_a & p_a\\ q_a & 1-q_a\end{pmatrix} ,$$

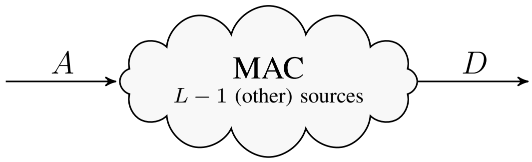

Fig. 2. A tagged source L, comprising of arrival and departure processes A and D, respectively, competing on a MAC shared channel (Aloha or CSMA/CA) with $L-1$ other sources

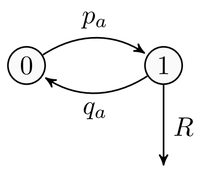

Fig. 3. The arrival process for source L, modelled in terms of a Markov-Modulated On-Off (MMOO) process

whereas a's steady state distribution is given by

$$\pi_{\mathbf{a}} = \left(\frac{q_a}{p_a + q_a}, \frac{p_a}{p_a + q_a} \right) \ .$$

The cumulative arrival process can be represented as

$$A(n) = \sum_{k=1}^{n} f(a_k) \ , \tag{5}$$

where $f(0) = 0$, $f(1) = R$, and $R > 0$ is the peak rate transmitted while the source is in state '1' (i.e., the "On" state).

In the following we consider the two cases when the source L shares an Aloha or CSMA/CA channel with $L-1$ other (saturated) sources denoted by $\{1, 2, \ldots, L-1\}$.

A. Aloha

With the (slotted) Aloha protocol, in each time slot a source transmits with a fixed probability $p_{tr} > 0$, independently from the other sources and also from previous transmissions. Thus, the probability of a successful transmission is given by

$$p_{suc} := p_{tr} \left(1 - p_{tr}\right)^{L-1} \ .$$

During the interval of a successful transmission the link provides an ideal capacity $C > 0$. In any other interval, due to a successful transmission of another source or a collision, no capacity is provided (for source L). The service process for the source L is thus given by

$$S(m, n) := \sum_{k=m+1}^{n} s_k \ ,$$

with the (instantaneous) service rates

$$s_k := \begin{cases} C & \mathbb{P} = p_{tr} \left(1 - p_{tr}\right)^{L-1} \\ 0 & \mathbb{P} = 1 - p_{tr} \left(1 - p_{tr}\right)^{L-1} \end{cases}$$

(see Figure 4 and also Ciucu *et al.* [9]).

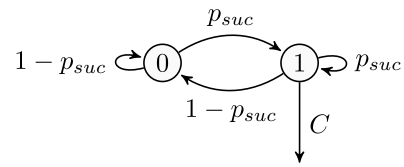

Fig. 4. The service process for source L, modelled in terms of a process with independent increments correspondign to a Aloha link

To compute stochastic bounds on the backlog and delay of source L, we observe that A and S admit arrival- and service-martingales, according to Examples 8 and 7, respectively. We consider the corresponding parameters, i.e., $h_a(\cdot)$ and K_a for the arrival-martingales, and K_s for the service martingales (recall that $h_s(\cdot) = 1$ for the i.i.d. service-martingales).

We now state the main result for the Aloha scenario.

Corollary 11 (Backlog and Delay for Aloha)**.** *Assume the stability condition $E[a_1] < E[s_1]$ and let*

$$\theta^* := \sup\{\theta > 0 : sp(T_a^\theta) = \mathcal{L}_s(\theta)^{-1}\} \ ,$$

where $sp(\cdot)$ denotes the maximal positive eigenvalue, and

$$\mathcal{L}_s(\theta) := 1 - p_{tr} \left(1 - p_{tr}\right)^{L-1} + p_{tr} \left(1 - p_{tr}\right)^{L-1} e^{-\theta C}$$

is the Laplace transform of s_k. Let further h_a be a (positive) eigenvector of $T_a^{\theta^*}$. Then the following bounds hold for the backlog and delay of source L:

$$\mathbb{P}(Q \geq \sigma) \leq \frac{\mathbb{E}[h_a(a_0)]}{H} e^{-\theta^* \sigma}$$

$$\mathbb{P}(W(n) \geq k) \leq \frac{\mathbb{E}[h_a(a_0)]}{H} e^{-\theta^* K_s k} \ ,$$

where H is defined as in Definition 3.

Proof. Note first that θ^* is well-defined (i.e., the supremum is taken over a non-empty set) because

$$\frac{d}{d\theta} sp(T_a^\theta) \Big|_{\theta=0} = \mathbb{E}[a_1] < \mathbb{E}[s_1] = \frac{d}{d\theta} \mathcal{L}_s(\theta)^{-1} \Big|_{\theta=0} \ .$$

Note also that the more explicit definition of θ^* follows from Theorem 9, whereby the values K_a and K_s are from Examples 8 and 7, respectively, i.e.,

$$\theta^* := \sup \left\{ \theta > 0 : \frac{\log sp(T_a^\theta)}{\theta} \leq \frac{\log E\left[e^{-\theta s_1}\right]}{-\theta} \right\} \ .$$

The replacement of the inequality by an equality is possible due to the continuity of the eigenvalues and the Laplace transform. The rest of the proof follows from Theorems 9 and 10 using the constructions from Examples 7 (for the service-martingales) and 8 (for the arrival-martingales). □

To illustrate the accuracy of the obtained delay bounds, we quickly provide several numerical results in Figures 5 and 6, by varying both the utilization and also the number of sources. The bounds are shown as continuous lines and the simulation results are shown as box-plots.

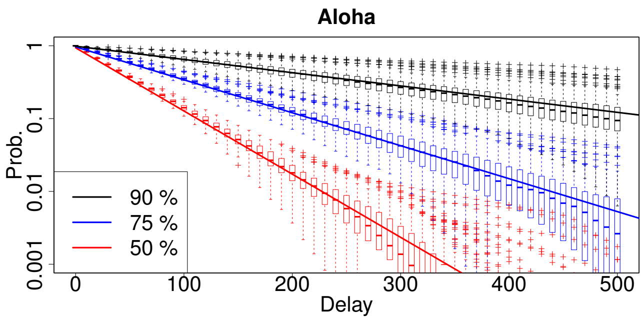

Fig. 5. CCDF of the virtual delay of source L with probabilities $p_a = 0.1$, $q_a = 0.5$, $p_{tr} = 0.2$, $L = 10$ sources, and utilizations $\rho = 0.5, 0.75, 0.9$ (bottom to top), respectively

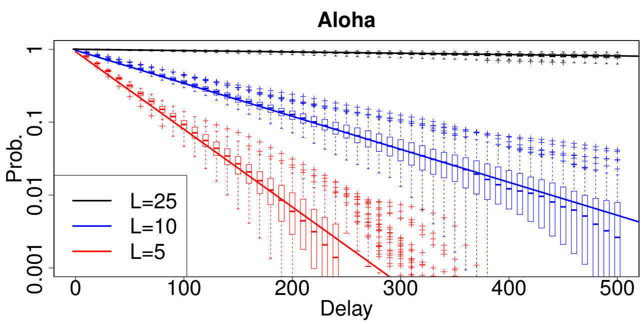

Fig. 6. CCDF of the virtual delay of source L with probabilities $p_a = 0.1$, $q_a = 0.5$, $p_{tr} = 0.2$, $\rho = 0.75$, and number of sources $L = 5, 10, 25$ (bottom to top), respectively

B. CSMA/CA

We adopt the CSMA/CA model from Durvy *et al.* [13] in terms of a Markov chain $(s_n)_n$, as depicted in Figure 7. Due to its tree structure, the Markov chain is reversible Kelly [21]). The source L can transmit (subject to current buffer occupancy) at some peak rate $C > 0$ (i.e., ideal channel's capacity) while in state L, whereas all sources are in backoff mode while in state 0.

The transition matrix is given by

$$T_s = \begin{pmatrix} 1-p_s & \frac{p_s}{L} & \dots & \frac{p_s}{L} \\ q_s & 1-q_s & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ q_s & 0 & \dots & 1-q_s \end{pmatrix} ,$$

whereas the steady-state distribution of s is given by

$$\pi_{\mathbf{s}} = \left(\frac{q_s}{p_s+q_s}, \frac{p_s}{L(p_s+q_s)}, \dots, \frac{p_s}{L(p_s+q_s)} \right) .$$

Using the methodology from [9], the service process $S(m,n)$ of link L can be represented by

$$S(m,n) := \sum_{k=m+1}^{n} C 1_{\{s_k = L\}} , \quad (6)$$

where $1_{\{\cdot\}}$ denotes the indicator function.

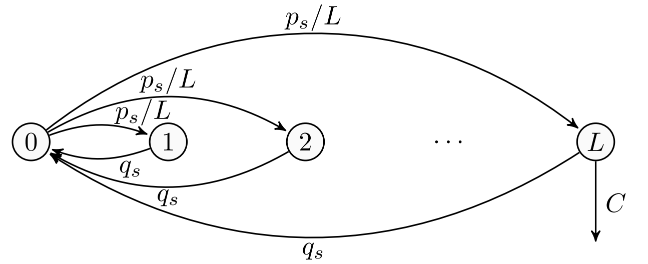

Fig. 7. The service process for source L, modelled in terms of a Markov process corresponding to a CSMA/CA link

To compute stochastic bounds on the backlog and delay experienced by the source L, Example 8 implies that both the arrival and service processes A and S from Eqs. (5) and (6) admit arrival- and service-martingales, respectively. Consider the corresponding parameters, i.e., $h_a(\cdot)$ and K_a for the arrival-martingales and $h_s(\cdot)$ and K_s for the service martingales; let also T_s^θ be the corresponding exponential column-transform for the service process.

We now state the main result for the CSMA/CA scenario.

Corollary 12 (Backlog and Delay for CSMA/CA)**.** *Assume the stability condition $E[a_1] < E[s_1]$ and let*

$$\theta^* := \sup\{\theta > 0 : sp(T_a^\theta) = (sp(T_s^{-\theta}))^{-1}\} ,$$

where $sp(\cdot)$ denotes the maximal positive eigenvalue. Let also h_a and h_s be corresponding (positive) eigenvectors of $T_a^{\theta^*}$ and $T_s^{\theta^*}$, respectively. Then the following bounds hold for the backlog and delay of source L:

$$\mathbb{P}(Q \geq \sigma) \leq \frac{\mathbb{E}[h_a(a_0)]\mathbb{E}[h_s(s_0)]}{H} e^{-\theta^*\sigma}$$
$$\mathbb{P}(W(n) \geq k) \leq \frac{\mathbb{E}[h_a(a_0)]\mathbb{E}[h_s(s_0)]}{H} e^{-\theta^* K_s k} ,$$

where H is defined as in Definition 3.

Proof. Note first that θ^* is well-defined (i.e., the supremum is taken over a non-empty set) because

$$\frac{d}{d\theta} sp(T_a^\theta)\Big|_{\theta=0} = \mathbb{E}[a_1] < \mathbb{E}[s_1] = \frac{d}{d\theta}\left(sp(T_s^{-\theta})\right)^{-1}\Big|_{\theta=0} .$$

For the rest of the proof simply apply Theorems 9 and 10 for the constructions of arrival- and service-martingales from Example 8. According to these constructions, the definition of θ^* from Theorem 9 reduces to the more explicit form

$$\theta^* := \sup\left\{\theta > 0 : \frac{\log sp(T_a^\theta)}{\theta} \leq \frac{\log sp(T_s^{-\theta})}{-\theta}\right\} .$$

The further replacement of the inequality by an equality is due to the same argument as in the proof of Corollary 11. □

As for Aloha, we quickly provide several numerical results in Figures 8 and 9; the figures confirm that the stochastic delay bounds are very accurate for a broad range of scenarios (note that at large values of the tail delay, the box plots widen due to the availability of fewer data points in the simulations).

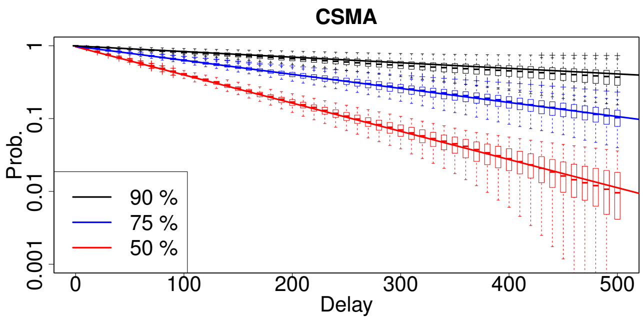

Fig. 8. CCDF of the virtual delay of source L with probabilities $p_a = 0.1$, $q_a = 0.5$, $p_s = 0.8$, $q_s = 0.2$, $L = 10$ sources, and utilizations $\rho = 0.5, 0.75, 0.9$ (bottom to top), respectively

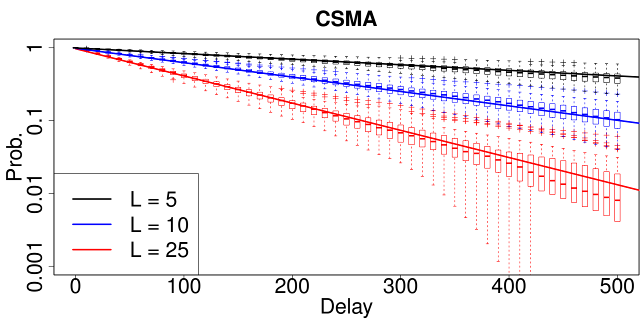

Fig. 9. CCDF of the virtual delay of source L with probabilities $p_a = 0.1$, $q_a = 0.5$, $p_s = 0.8$, $q_s = 0.2$, utilization $\rho = 0.75$, and number of sources $L = 5, 10, 25$ (bottom to top) flows, respectively

Finally, we note that for both Aloha and CSMA/CA, the arrival and service processes are independent. That is due to the fact that the construction of the service process is oblivious to the arrival process, and in particular it holds for saturated arrivals; such constructions are conservative since the network nodes do not rely on backlog state information from neighborhood nodes, and thus the channel may be underutilized.

IV. Further Applications. Scheduling and MIMO

In this section we present more complex applications of the general results from Section II. Concretely, we extend the CSMA/CA scenario from Section III-B in two directions: 1) accounting for in-source scheduling and 2) accounting for spatial multiplexing MIMO.

A. In-Source Scheduling

We generalize the basic scenario from Section III-B by assuming that the tagged source L comprises of multiple flows, whose transmissions are first scheduled before being sent over the CSMA/CA channel. Without loss of generality we assume only two flows, whose arrivals and departures are denoted by A and D, and A' and D', respectively, and a Static Priority (SP) scheduling policy (see Figure 10).

The arrival processes A and A' of the source L are statistically independent, and are assumed for simplicity to have the same parameters as in Section III for the arrival-martingales,

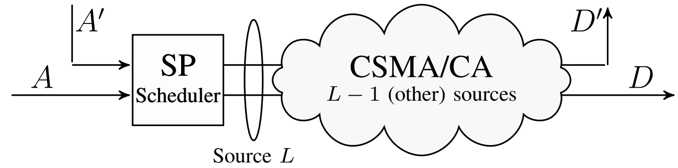

Fig. 10. A tagged source L, comprising of two arrival flows A and A', which are scheduled according to an SP policy before being transmitted over the channel

i.e., $K_a = K'_a$ and $h_a(\cdot) = h'_a(\cdot)$ obtained from Example 8; let also T_a^θ be the corresponding exponential column-transform (of a single flow).

In this scheduled system, we are interested in the performance metrics (i.e., backlog and delay) for the flow A. Because the channel offers an *exact* service process (i.e., $S(m,n)$ defined in Eq. (6)), in the sense that Eq. (1) is in fact satisfied with equality (see [9]), it follows that the overall service process available to the flow A is given by

$$S_A(m,n) = S(m,n) - A'(m,n) \ .$$

This service process is known in the (stochastic) network calculus literature as the leftover service process (see also Chang [7] and Fidler [16]).

Concerning the service process $S(n)$, recall from Example 8 that it admits service-martingales with parameters $h_s(\cdot)$ and K_s; let T_s^θ be the corresponding column-transform.

Corollary 13 (Backlog and Delay for SP + CSMA/CA). *Assume the stability condition $2E[a_1] < E[s_1]$ and let*

$$\theta^* := \sup\{\theta > 0 : \left(sp(T_a^\theta)\right)^2 = \left(sp(T_s^{-\theta})\right)^{-1}\} \ ,$$

where $sp(\cdot)$ denotes the maximal positive eigenvalue. Let also h_a and h_s be corresponding (positive) eigenvectors of $T_a^{\theta^*}$ and $T_s^{\theta^*}$, respectively. Then the following bounds hold for the backlog and delay of the (sub-)arrival flow A of source L:

$$\mathbb{P}(Q \geq \sigma) \leq \frac{\mathbb{E}[h_a(a_0)]^2 \mathbb{E}[h_s(s_0)]}{H} e^{-\theta^* \sigma}$$
$$\mathbb{P}(W(n) \geq k) \leq \frac{\mathbb{E}[h_a(a_0)]^2 \mathbb{E}[h_s(s_0)]}{H} e^{-\theta^*(K_s - K'_a)k} \ ,$$

where

$$H := \min\{h_a(x) h'_a(x') h_s(y) : x + x' - y > 0\} \ .$$

Proof. Note first that θ^* is well-defined using the same argument from Corollary 12. Next we slightly adapt Theorems 9 and 10 for the constructions from Example 8. The key observation (in the case of the delay) is that by the independence assumption of A, A', and S, the product

$$h_a(a_n) h_a(a'_n) h_s(s_n) e^{\theta\left(A(k,n) - (n-k)K_a + A'(n) - nK'_a + nK_s - S(n)\right)}$$

is a supermartingale. Note also that $A'(k,n)$ is shifted with respect to both $A'(n)$ and $S(n)$, whence the term $K_s - K_a$ in the asymptotic decay rate of the delay. Finally, the definition of θ^* from Theorem 9 becomes

$$\theta^* := \sup\{\theta > 0 : 2K_a \leq K_s\} \ ,$$

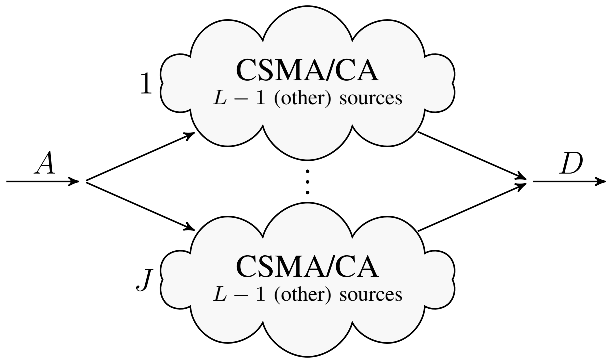

Fig. 11. Spatial multiplexing MIMO: the tagged source L is transmitted over J independent MAC channels

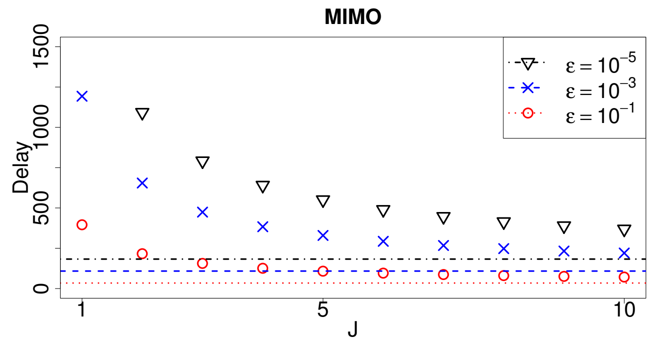

Fig. 12. The tail delays from Corollary 14 as a function of the number of channels J ($p_a = 0.1$, $q_a = 0.5$, $p_s = 0.8$, $q_s = 0.2$, utilization $\rho = 0.75$, and $\varepsilon = 10^{-5}, 10^{-3}, 10^{-1}$); the bottom horizontal lines correspond to the tail delays under deterministic service (the corresponding bounds are computed with Theorem 10 and Examples 6 and 8)

which completes the proof. □

Corollary 13 generalizes the SP delay bounds from [26], available for a constant-rate server; similar generalizations are immediate in the case of other scheduling policies such as FIFO and EDF. Corollary 13 reveals the *modularity* feature of the proposed methodology, in the sense of jointly analyzing interconnected systems such as in-source scheduling and MAC protocols; a further convincing example is provided next.

B. MIMO

Here we generalize the basic scenario from Section III-B by considering a spatial multiplexing MIMO (multiple-input multiple-output) scenario (see, e.g., Heath and Paulraj [19]), in which the source L is served by J CSMA/CA channels (see Figure 11). To keep the analysis tractable, we assume the independence of the channels and disregard fading effects.

The source L has the same arrival process as in Section III, in particular with the parameters K_a and $h_a(\cdot)$ for the corresponding arrival-martingales. Furthermore, by extending the notations from Section III-B, we assume that the service on each channel $j = 1, 2, \ldots, J$ is modulated by i.i.d. Markov processes $(s_{j,n})_n$ (with the same parameters as in Section III-B). For the particular case of MIMO spatial multiplexing, the *overall* service process $S_j(m, n)$ of link L can be represented by

$$S(m,n) := \sum_{j=1}^{J} S_j(m,n) := \sum_{j=1}^{J} \sum_{k=m+1}^{n} C1_{\{s_{j,k}=L\}} \ , \quad (7)$$

where $S_j(m,n)$ is the service process for channel j.

Using Example 8, each service process $S_j(n)$ admits service-martingales with parameters $h_s(\cdot)$ and K_s (due to the i.i.d. assumption across the modulated Markov processes). Let also T_a^θ and T_s^θ be the corresponding exponential column-transforms for the arrival and service processes.

Corollary 14 (Backlog and Delay for MIMO)**.** *Assume the stability condition* $E[a_1] < JE[s_1]$ *and let*

$$\theta^* := \sup\{\theta > 0 : sp(T_a^\theta) = (sp(T_s^{-\theta}))^{-J}\} \ ,$$

where $sp(\cdot)$ *denotes the maximal positive eigenvalue. Let also* h_a *and* h_s *be corresponding (positive) eigenvectors of* $T_a^{\theta^*}$ *and* $T_s^{\theta^*}$*, respectively. Then the following bounds hold for the backlog and delay of source* L*:*

$$\mathbb{P}(Q \geq \sigma) \leq \frac{\mathbb{E}[h_a(a_0)]\mathbb{E}[h_s(s_0)]^J}{H_J} e^{-\theta^*\sigma}$$

$$\mathbb{P}(W(n) \geq k) \leq \frac{\mathbb{E}[h_a(a_0)]\mathbb{E}[h_s(s_0)]^J}{H_J} e^{-\theta^* K_s k} \ ,$$

where

$$H_J := \min\{h_a(x) \prod_{j=1}^{J} h_s(y_j) : x - \sum_{j=1}^{J} y_j > 0\} \ .$$

Proof. As in Corollary 12, θ^* is well-defined. We make the key observation that by the independence assumption on the Markov processes $(s_{j,n})_n$, the product

$$\prod_{j=1}^{J} h_j(s_{j,n}) e^{\theta(JK_s - S(n))}$$

is a service-martingale for the overall service process S. Consequently, the definition of θ^* from Theorem 9 becomes

$$\theta^* := \sup\{\theta > 0 : K_a \leq JK_s\} \ .$$

The rest proceeds as in Corollary 12. □

Let us now analyze the impact of the number of channels J, in particular on the probabilistic delay of source L. Due to the implicit definition of θ^* from Corollary 14 in terms of eigenvalues/vectors, a quantitative result is conceivably difficult to be obtained. We thus resort to a numerical experiment, using the same numerical values as in Section III-B. Concretely, in Figure 12, we illustrate the tail delay for three violation probabilities (i.e., $\varepsilon = 10^{-5}, 10^{-3}, 10^{-1}$) as a function of the number of channels J, and for a normalized utilization $\rho = 0.75$ (for each J). The key observation is the exponential decay of the delay, an effect which is more pronounced for smaller (and thus more practical) values of ε.

The figure also includes the corresponding delays in a scenario with deterministic (and normalized) service, for the three values of ε (i.e., the three horizontal bottom lines, which are invariant to J). As expected, for each ε, the tail delays converge to the horizontal line corresponding to a deterministic service; especially for small values of ε, the convergence is however very slow and not visible in the current plot. While we limit J to 10 for both practical considerations and the readability of the plot, we point out that for $\varepsilon = 10^{-5}$ the convergence is still not visible at $J = 100$, but only around $J = 1000$ (i.e., an impractical regime).

Overall, the figure convincingly indicates that MIMO spatial multiplexing is subject to the power-of-two phenomena. Concretely, for realistic small values of ε, there is a dramatic decrease in delay when increasing the number of channels from $J = 1$ to $J = 2$. The delays continue to decrease by further increasing J, but at much smaller rates. See also Mitzenmacher [25] for a related discussion of the power-of-two phenomena in the context of randomized load balancing.

V. Conclusions

In this paper we have developed the first rigorous and accurate methodology to compute queueing performance metrics (i.e., backlog and delay) for bursty sources sharing a MAC (bursty) channel: the sources are modelled using an existing arrival-martingale model, whereas the available service for the source at the shared channel is modelled using a novel service-martingale model. By leveraging the modelling power of the proposed martingale methodology we have shown that the obtained stochastic bounds are remarkably tight in the case of Markov-modulated sources, and Aloha and CSMA/CA channels. We have also shown that our methodology offers an attractive modularity feature, in the sense that we could extend basic results to much more complex scenarios accounting for in-source SP scheduling or MIMO spatial multiplexing.

Acknowledgement

This work was partially funded by the DFG grant Ci 195/1-1.

References

[1] N. Abramson. The Aloha system: another alternative for computer communications. In *Proceedings of AFIPS Joint Computer Conferences*, pages 281–285, 1970.

[2] F. Alizadeh-Shabdiz and S. Subramaniam. Analytical models for single-hop and multi-hop ad hoc networks. In *ACM Broadnets*, pages 449–458, Oct 2004.

[3] S. Beuerman and E. Coyle. The delay characteristics of CSMA/CD networks. *IEEE Transactions on Communications*, 36(5):553–563, May 1988.

[4] G. Bianchi. Performance analysis of the IEEE 802.11 distributed coordination function. *IEEE Journal on Selected Areas in Communications*, 18(3):535–547, Mar. 2000.

[5] F. Cali, M. Conti, and E. Gregori. Dynamic tuning of the IEEE 802.11 protocol to achieve a theoretical throughput limit. *IEEE/ACM Transactions on Networking*, 8(6):785–799, Dec 2000.

[6] M. Carvalho and J. Garcia-Luna-Aceves. Delay analysis of IEEE 802.11 in single-hop networks. In *IEEE International Conference on Network Protocols (ICNP)*, pages 146–155, Nov 2003.

[7] C.-S. Chang. *Performance Guarantees in Communication Networks*. Springer Verlag, 2000.

[8] G. Choudhury, D. Lucantoni, and W. Whitt. Squeezing the most out of ATM. *IEEE Transactions on Communications*, 44(2):203–217, Feb. 1996.

[9] F. Ciucu, R. Khalili, Y. Jiang, L. Yang, and Y. Cui. Towards a system theoretic approach to wireless network capacity in finite time and space. In *IEEE Infocom*, 2014.

[10] F. Ciucu, F. Poloczek, and J. Schmitt. Sharp per-flow delay bounds for bursty arrivals: The case of FIFO, SP, and EDF scheduling. In *IEEE Infocom*, 2014.

[11] F. Ciucu and J. Schmitt. Perspectives on network calculus - No free lunch but still good value. In *ACM Sigcomm*, 2012.

[12] N. G. Duffield. Exponential bounds for queues with Markovian arrivals. *Queueing Systems*, 17(3-4):413–430, Sept. 1994.

[13] M. Durvy, O. Dousse, and P. Thiran. Self-organization properties of CSMA/CA systems and their consequences on fairness. *IEEE Transactions on Information Theory*, 55(3):931–943, Mar. 2009.

[14] A. Ephremides and B. E. Hajek. Information theory and communication networks: An unconsummated union. *IEEE Transactions on Information Theory*, 44(6):2416–2434, Oct. 1998.

[15] M. Fidler. An end-to-end probabilistic network calculus with moment generating functions. In *IEEE International Workshop on Quality of Service (IWQoS)*, pages 261–270, 2006.

[16] M. Fidler. A survey of deterministic and stochastic service curve models in the network calculus. *IEEE Communications Surveys & Tutorials*, 12(1):59–86, Feb. 2010.

[17] R. G. Gallager. A perspective on multiaccess channels. *IEEE Transactions on Information Theory*, 31(2):124–142, Mar. 1985.

[18] M. Garetto and C.-F. Chiasserini. Performance analysis of 802.11 WLANs under sporadic traffic. In *IFIP Networking*, pages 1343–1347, 2005.

[19] R. Heath and A. Paulraj. Switching between diversity and multiplexing in MIMO systems. *IEEE Transactions on Communications*, 53(6):962–968, June 2005.

[20] Y. Jiang and Y. Liu. *Stochastic Network Calculus*. Springer, 2008.

[21] F. P. Kelly. *Reversibility and Stochastic Networks*. Wiley, Chichester, 1979.

[22] F. P. Kelly. Notes on effective bandwidths. In *Stochastic Networks: Theory and Applications. (Editors: F.P. Kelly, S. Zachary and I.B. Ziedins) Royal Statistical Society Lecture Notes Series, 4*, pages 141–168. Oxford University Press, 1996.

[23] L. Kleinrock and F. A. Tobagi. Packet switching in radio channels: Part I–carrier sense multiple-access modes and their throughput-delay characteristics. *IEEE Transactions on Communications*, 23(12):1400–1416, Dec. 1975.

[24] S. Lam. *Packet Switching in a Multi-Access Broadcast Channel with Application to Satellite Communication in a Computer Network*. PhD thesis, University of California at Los Angeles, Los Angeles, CA, 1974.

[25] M. Mitzenmacher. The Power of Two Choices in Randomized Load Balancing. *IEEE Transactions on Parallel and Distributed Systems*, 12(10):1094–1104, Oct. 2001.

[26] F. Poloczek and F. Ciucu. Scheduling analysis with martingales. *Performance Evaluation*, 79:56 – 72, 2014.

[27] H. Takagi and L. Kleinrock. Throughput analysis for persistent CSMA systems. *IEEE Transactions on Communications*, COM-33(7):627–638, July 1985.

[28] O. Tickoo and B. Sikdar. Queueing analysis and delay mitigation in IEEE 802.11 random access MAC based wireless networks. In *IEEE Infocom*, pages 1404–1413, March 2004.

[29] F. A. Tobagi. Distributions of packet delay and interdeparture time in slotted Aloha and carrier sense multiple access. *Journal of the ACM*, 29(4):907–927, Oct. 1982.

[30] D. Wu and R. Negi. Effective capacity: A wireless link model for support of quality of service. *IEEE Transactions on Wireless Communication*, 2(4):630–643, July 2003.

[31] D. Wu and R. Negi. Effective capacity-based quality of service measures for wireless networks. *ACM Mobile Networks and Applications*, 11(1):91–99, Feb. 2006.

[32] Y. Yang and T.-S. P. Yum. Delay distributions of slotted Aloha and CSMA. *IEEE Transactions on Communications*, 51(11):1846–1857, Nov. 2003.