Performance Analysis of the 802.11 Distributed Coordination Function under Sporadic Traffic

M. Garetto and C.-F. Chiasserini
Dipartimento di Elettronica, Politecnico di Torino, Italy
Email: {garetto, chiasserini}@polito.it

***Abstract*— We analyze the performance of the Distributed Coordination Function (DCF) for 802.11 WLANs. We consider a fixed number of contending stations within radio proximity, and we investigate the important case in which stations operate under non-saturated conditions. We assume that the MAC queues of wireless stations receive from the upper layers a stationary arrival process of packets. We identify the fundamental problems that arise in building an analytical model of the system, and we propose different approaches to overcome these difficulties. Finally, we apply our modelling technique to study several important issues in 802.11 networks, such as the impact of bursty traffic and the system performance in a multirate environment. The accuracy of the analytical results is verified by simulation with *ns-2*.**

***Index Terms*— Stochastic processes/Queueing theory**

I. Introduction

In the last few years, IEEE 802.11 Wireless Local Area Networks (WLANs) [1] have emerged as a prevailing technology for wireless access. Such a success is mainly due to the ability of 802.11 to provide a high-speed wireless environment, and to apply to infrastructured as well as ad hoc networking. Nevertheless, the 802.11 technology still presents some limitations, among others, the limited bandwidth of the communication channel and the lack of quality of service (QoS) support. High data-rates in 802.11 networks are possible but at the cost of a shorter transmission range, i.e., they can be employed only when the distance between transmitter and receiver is sufficiently small. It follows that an efficient sharing of the available bandwidth is of crucial importance. As for the QoS support, some enhancements to the current channel access schemes are under study. In particular, the IEEE 802.11e draft standard [2] aims at introducing service differentiation at the MAC layer so as to provide the desired QoS level for various classes of traffic.

These facts clearly indicate that the core element of the 802.11 technology is the MAC protocol, since it determines the efficiency of using the radio resources and the network performance. Therefore, it is of fundamental importance to develop a model of the 802.11 MAC function that accurately represent the system behavior under realistic assumptions and enable us to thoroughly investigate the system behavior.

The primary 802.11 MAC function is the so-called Distributed Coordination Function (DCF). The DCF is a random access scheme based on the Carrier Sense Multiple Access with Collision Avoidance protocol (CSMA/CA). The DCF has two operating modes: the basic channel access mode and the RTS/CTS (Request-to-Send/Clear-To-Send) mode. In the following, we assume the reader to be familiar with the access procedures of the DCF; for the necessary background see, for example, [1], [4], [5].

A. Our Contribution

In this work, we present an analytical model of the 802.11 DCF. As discussed in Section I-B, several analytical studies of the DCF have appeared in the literature [4]–[16]. Our model differs from previous work in that:
(i) it identifies the critical assumptions in the development of analytical models of 802.11 networks;
(ii) it presents a fairly simple as well as accurate model of the DCF in presence of non-saturated traffic sources;
(iii) it is general enough to account for different arrival processes and traffic patterns, in particular, it applies to the case of bursty traffic like that produced by the TCP protocol;
(iv) it evaluates the system performance in a multirate environment;
(v) it applies to the case where a station seizing the channel is entitled to transmit a burst of packets, as specified in the IEEE 802.11e [2];
(vi) it evaluates several metrics of interest, such as the network throughput, the packet loss probability, the distribution of the MAC queue length at the wireless stations, the average packet delay, and the round-trip-time (which, together with the packet loss probability, is the main variable affecting transport protocols such as TCP).

B. Related Work

There are several theoretical works investigating the performance of the DCF access scheme. In [3], the authors compute the theoretical upper limit for the IEEE 802.11 protocol capacity and optimize the system throughput by dynamically tuning the parameter settings of the DCF as traffic conditions vary. In [4], Bianchi proposes an analytical model, based on Markovian techniques, which well represents the behavior of wireless stations under saturated conditions. The same model has been subsequently refined in many different ways in [5]–[10]. In particular, the work in [9] accurately models the decrease of the stations' backoff counter, while [10] extends the Bianchi's model to compute the average service time and jitter experienced by a packet in a saturated, single-hop network.

More recently, research efforts have been devoted to analyzing the case of non-saturated stations [5], [11]–[16]. An easy way to extend the analysis to the case of non-saturated traffic sources is to consider a system in which the number of competing stations slowly varies over time, and locally apply results obtained under saturated conditions. The idea is proposed by Foh and Zukerman in [12]. In [13] the authors model exponential on-off sources using a product-form queuing network, while in [5] a processor sharing queueing model with state-dependent service rates is used to analyze non-persistent sources and compute the average flow transfer delay. In [15], the authors propose a $G/G/1$ model representing the queue of an 802.11 station, and derive results for the queue lengths and the channel access delay under general traffic arrival patterns. Note, however, that the model in [15] considers infinite retrials for a packet transmission, infinite buffers and, more importantly, that the behavior of individual stations are independent – an assumption that, as shown later in this paper, does not hold in the case of unsaturated sources. An analytical fluid model is described in [14], along with a scheduling scheme aimed at improving the protocol capacity.

Finally, relevant to our work is also the study presented in [17], where the authors present a methodology to estimate the number of competing stations using run-time measurements of collision probability.

C. Paper Organization

We present the basic model of the DCF for saturated sources in Section II. There, we first describe our assumptions and the simplified model of the channel occupation that we consider; then, we present our Markovian analysis and some numerical results. The case of non-saturated sources is studied in Section III, where two analytical models of increasing complexity are introduced. The results derived through our models are validated by using the *ns-2* simulator[1] [18]. In Section IV, we present some applications of our models. More specifically, we consider the cases where stations use variable frame sizes, employ multiple transmission rates, generate bursty traffic, and have the possibility to transmit bursts of frames, as in IEEE 802.11e. Finally, Section V concludes the paper and introduces some aspects that will be subject of future research.

II. Basic Model for Saturated Sources

In this section we present the basic model describing the behavior of saturated sources, following the approach of [4]. Our main assumptions are as follows:

- We consider a fixed number n of contending stations accessing the same wireless channel. These stations may either be associated with the same access point, or form an ad-hoc network. Stations operate under *saturation* conditions, that is, they always have data to send.
- All stations are within radio proximity, and there is no hidden terminal problem.
- Stations are equally likely to access the channel; in particular, there are no seizing effects [5].
- The communication channel is error-free.

A. Simplified Description of the Channel Occupation

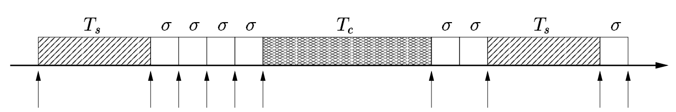

Fig. 1. Temporal evolution of the occupation of the wireless channel

A schematic representation of the occupation of the shared medium over time is depicted in Figure 1. The channel has three possible states: *(i)* busy channel due to a successful transmission; *(ii)* busy channel due to a collision caused by more than one station transmitting simultaneously; *(iii)* idle channel. The duration of the time intervals during which the channel remains in the three states are denoted by T_s, T_c, and σ, respectively. In Figure 1 the time instants at which the channel can change state are pointed to by arrows placed below the temporal axis. Note that, σ corresponds to one 802.11 slot time; at the end of an idle slot, stations can decrement their backoff time counters.

The internal structure of T_s and T_c is illustrated in Figure 2, for both the basic access and the RTS/CTS mechanisms. The duration of T_s and T_c vary depending

[1]We corrected some inconsistencies of the the current release of the *ns-2* simulator [18], namely *ns-2.27*, with respect to the IEEE 802.11 specifications [1].

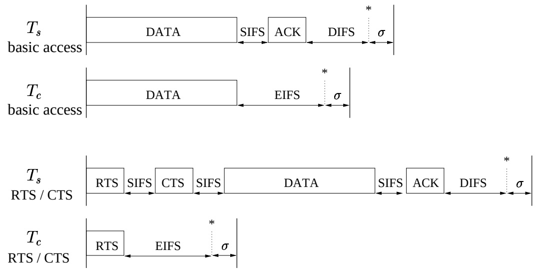

Fig. 2. Internal structure of a T_s and a T_c interval for the basic access and the RTS/CTS mechanisms

on the frame size and on the sending rate of the transmitting station(s). For simplicity, in this section we assume that all stations have a fixed frame size and transmit at the same rate, so that T_c and T_s are constant. This is not a limitation, as the analysis can be easily generalized to consider the case of variable duration (see Section IV).

Notice that, differently from [4], we include an additional slot time σ at the end of each interval as shown in Figure 2. The reason for that comes from the consideration that the backoff time counter is decremented by one only *after* an idle time slot has elapsed. As a consequence, at the time instants marked with an asterisk in Figure 2, no station can have zero backoff and attempt a new transmission (if the backoff counter were zero, the station should have started transmitting at the beginning of the current interval). An exception to this behavior occurs if a station, who has just transmitted in the current interval, extracts a value of zero for the new backoff counter. In this case, a station who has successfully sent a frame could send a new frame at the time instant marked with an asterisk without colliding with other stations, i.e., successfully with probability one. Recent versions of the standard [2] have eliminated the occurrence of this odd (and unfair) event forcing a station to wait for at least one time slot between two successive transmissions. Correspondingly, in our model, no station can start a transmission at the instants marked with an asterisk; if a station picks up zero as new backoff counter, its behavior is considered the same as if it had extracted a value of one. We observe that taking precisely into account the occurrence of a station extracting zero backoff would be excessively complicated in an analytical model, and the impact of this special case on global performance is marginal, as we have verified by simulation.

The above description of the occupation of the channel does not sacrifice any important detail of the protocol operation; indeed the main assumption is that all stations synchronize with the channel. Also, it provides a quite powerful way to analyze the system performance, since one can consider the evolution of the system only at the time instants at which the state of the channel changes (marked by arrows in Figure 1), This can be done by defining a *time step* as the time interval between two consecutive changes of the channel state; the duration of a time step will vary depending on whether a successful transmission, a collision, or an idle slot, has occurred. However, the internal structure of T_s and T_c (see Figure 2) imply that all stations with non-zero backoff always decrement their counter at the end of each time step (i.e., at the end of the idle interval of duration σ).

We have built a discrete time simulator based on the above description of the channel state, using the C language. The simulator basically maintains for each station only the value of the backoff counter and the current contention window size. Results match very well with *ns-2* simulations [18], while achieving an impressive speed-up with respect to the detailed event simulation as implemented in *ns-2*. Some numerical results are shown in Section II-C.

B. Markovian Analysis

In order to study the system analytically, a natural choice is to build a discrete Markov chain, embedded in the temporal evolution of the channel at the instants of a possible state change[2]. Moreover, relying on the fundamental assumption that the state of a station is independent of that of the others, we can reduce to considering the behavior of a single, tagged station. Following this approach, we propose the simple Markov chain depicted in Figure 3.

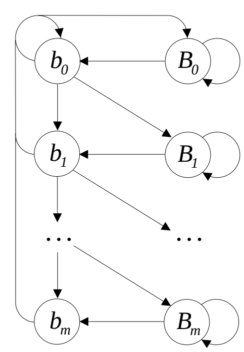

Fig. 3. Markov chain model for saturated sources

States labelled with b represent the station with backoff counter equal to zero, i.e., the case where the station actually transmits a frame in the current step. States labelled with B model the station while it decrements its backoff counter. States have an index in the range $\{0 \dots m\}$ representing the "backoff stage", where m is the *maximum retry limit* for the frame to be transmitted. Using the same notation as in [4], we denote by W_i the contention window size at backoff stage i. We have $W_i =$

[2]Note that the *time step* of the model is defined as in Section II-A.

TABLE I
TRANSITION PROBABILITIES OF THE BASIC MODEL
$\alpha_i = 2/W_i \quad \beta_i = 2/(W_i - 1)$

s_o	s_d	$P(s_o, s_d)$	Condition
b_i	b_0	$(1-p)\ \alpha_0$	$0 \le i < m$
	B_0	$(1-p)\ (1-\alpha_0)$	
	b_{i+1}	$p\ \alpha_{i+1}$	
	B_{i+1}	$p\ (1-\alpha_{i+1})$	
b_i	b_0	α_0	i = m
	B_0	$1-\alpha_0$	
B_i	b_i	β_i	$0 \le i \le m$
	B_i	$1-\beta_i$	

$\min(2^i\, \mathrm{CW}_{\min}, \mathrm{CW}_{\max})$, where $\mathrm{CW}_{\min}$ and $\mathrm{CW}_{\max}$ are constant parameters of the MAC protocol.

The transition probabilities $P(s_o, s_d)$ from state s_o to the successor state s_d are reported in Table I. The last column in the Table is a condition on the index of state s_o. To simplify the expression of the transition probabilities, we have put $\alpha_i = 2/W_i$ and $\beta_i = 2/(W_i - 1)$. Note that the factor 2 at the numerator of α_i and β_i is due to the fact that picking a value of 0 or 1 has the same effect of starting a new transmission at the beginning of the next interval (see our discussion in Section II-A). In the expression of the transition probabilities, the only unknown variable (to be computed) is p, which is the collision probability seen by a station transmitting on the channel.

Let us denote the stationary distribution of the Markov chain by $\boldsymbol{\pi} = \{\pi_s\}$, where s is a generic state of the model. Similarly to the model of Bianchi, we have that $\pi_{b_i} = p\ \pi_{b_{i-1}}$, for all $i > 0$.

The main differences with respect to previous models for saturated sources are as follows.

(i) In case of collision, a station with probability $\alpha_i = 2/W_i$ immediately retransmits the frame in the following step, being W_i the value of the updated contention window.

(ii) All states belonging to backoff stage i and having backoff counter greater than one, have been collapsed in a single state B_i. This has been done to reduce the number of states in view of the model extensions described later in the paper. The side-effect of this simplification is that the number of steps waited while decrementing the backoff counter is modelled as a geometrically distributed random variable, instead of a uniformly distributed variable. However, the transition probabilities $P(B_i, b_i)$'s have been chosen in such a way that the stationary probability of the collapsed states B_i are exactly the same that would result considering a uniformly distributed backoff, as done in [4]. The condition to be satisfied is given by,

$$\pi_{B_i} = \pi_{b_i}\ \frac{(W_i - 1)(W_i - 2)}{2\,W_i}$$

for all values of i. By doing so, all performance metrics derived with our simplified model coincide with those obtained with a more precise model of the backoff counter based on a uniform distribution.

Thanks to the particular structure of the Markov chain, the derivation of the stationary probabilities is straightforward. Indeed, all probabilities can be easily expressed as functions of π_{b_0}, and can be computed by normalizing the overall sum of probabilities to one.

After that, one can compute the probability τ that a station transmits during a time step, as $\tau = \sum_{i=0}^{m} \pi_{b_i}$. In order to derive the *conditional collision probability* p, we rely on the fundamental assumption that the state of the individual stations are independent. This yields the following expression,

$$p = 1 - (1 - \tau)^{n-1} \tag{1}$$

The two unknowns τ and p are computed by a simple iterative procedure, as in [4] where it is also shown that the fixed point approximation has a unique solution.

We compute the channel state probabilities Π_s, Π_c and Π_σ, which are the probabilities that a generic discrete time step is occupied by a successful transmission, a collision, or an idle slot, respectively, as follows:

$$\begin{aligned} \Pi_\sigma &= (1-\tau)^n \\ \Pi_s &= n\ \tau\ (1-\tau)^{n-1} \\ \Pi_c &= 1 - \Pi_\sigma - \Pi_c \end{aligned}$$

Finally, we calculate the *aggregate packet throughput* T_P, which is the main performance figure that can be derived from the model. This is defined as the rate (expressed in packets/s) at which data packets are successfully transferred by all stations over the wireless channel. It is given by,

$$T_P = \frac{\Pi_s}{\Pi_s\ T_s + \Pi_c\ T_c + \Pi_\sigma\ \sigma} \tag{2}$$

where the sum at the denominator computes the average duration of a time step.

C. Numerical Results

Here we present some results showing the accuracy of the simplified representation of the system behavior described in Section II-A, as well as of the Markovian analysis presented in Section II-B.

The parameters of the MAC and physical layers, unless otherwise specified, are set according to Table II and correspond to default values in the *ns-2* simulator.

TABLE II

PARAMETERS SETTING FOR THE MAC AND PHYSICAL LAYERS

SIFS	10 μs
DIFS	50 μs
EIFS	364 μs
σ	20 μs
BasicRate	2 Mbps
DataRate	11 Mbps
PLCP length	192 bits @ 1 Mbps
MAC header (RTS,CTS,ACK,DATA)	(20,14,14,28) bytes @ BasicRate
packet payload	1000 bytes
$(\mathrm{CW_{min}}, \mathrm{CW_{max}})$	(31,1023)
Short Retry Limit	7
Long Retry Limit	4

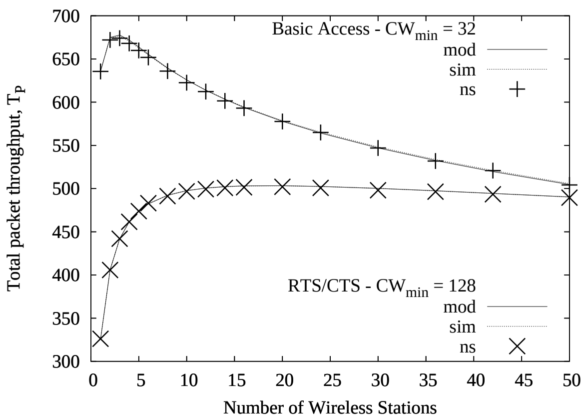

Fig. 4. Aggregate throughput as a function of the number of wireless stations, for basic channel access and RTS/CTS

Using the values in Table II, we obtain $T_s = 1283$ μs, $T_c = 1339$ μs for the basic channel access, and $T_s = 1823$ μs, $T_c = 656$ μs for the RTS/CTS scheme.

In Figure 4 we compare the packet throughput T_P obtained through *ns-2*, the discrete time simulator introduced in Section II-A (hereinafter called *sim*), and the Markovian analysis presented in Section II-B, in two different cases. In the first case, all stations employ the basic channel access scheme, and $\mathrm{CW_{min}}$ is set to the default value of 31. In the second case, the RTS/CTS access mechanism is used, and $\mathrm{CW_{min}}$ has been increased to 128. Figure 5 compares the probability of states b_i obtained by *sim* and the model, in case of the basic channel access.

Results perfectly match for any value of the number of wireless stations. This confirms that *(i)* the simplified description of the channel occupation (see Section II-A) represents correctly the protocol behavior; *(ii)* the assumptions on which the Markovian analysis is built are correct, especially the hypothesis on the independency of the stations behavior. Indeed, we have observed that not only aggregate metrics like T_P are correctly estimated, but also the detailed behavior of a station is accurately predicted.

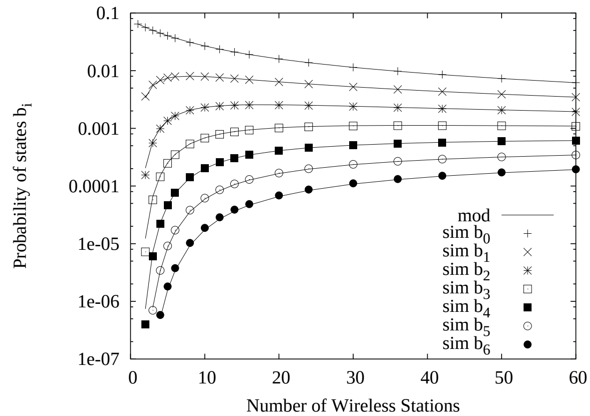

Fig. 5. Probability of states b_i obtained from the model and the simulator *sim*

III. MODELING NON-SATURATED SOURCES

In this section we describe two different approaches to deal with the more complicated case in which stations are not saturated. This means that the transmission queue of a station may become empty, thus the analysis in II-B does not apply anymore.

We assume that the MAC buffer at the wireless stations receive data packets from the upper layers according to some stationary, external arrival process with rate λ packets/s. For the sake of simplicity, we consider a Poisson arrival process, identical for all stations; however, our model is general enough to account for different arrival processes and traffic patterns. We denote by $\Lambda = \lambda\ n$ the total packet arrival rate at the stations' queues. The MAC buffer at each station is assumed to be of finite size K. Packets that cannot be stored in the buffer are immediately discarded upon arrival.

We represent the occupation of the wireless channel exactly as described before in Section II-A. Based on that, the discrete time simulator of the system already used under saturated conditions can be easily extended to consider non saturated sources: we simply need to keep track of the number of packets stored in the buffer of each station. During an interval of duration Δ, an average number $\lambda\Delta$ of new packets arrive at the MAC queue of each station. In the case of Poisson traffic, this number is distributed according to a Poisson distribution with mean $\lambda\Delta$, and we can exploit the fact that discrete time instants delimiting each interval form a renewal process.

The only problem to be faced is what happens when a new packet arrives at a previously empty queue, and the backoff counter of the station has been already decremented to zero. If the packet arrives when the channel is busy, or if the channel is not sensed as idle for a DIFS time, the station has to select a new backoff period. This happens most of the times unless we consider very underloaded conditions; thus we make the assumption that stations always select a new backoff period when a packet arrives at an empty queue of a station having zero backoff. Comparisons with *ns-2* simulations show that the error introduced by this assumption is negligible (see Figures 6 and 9), while it greatly simplifies the system analysis.

Although the extension of the discrete simulator is straightforward, the extension of the analysis turns out to be very difficult, as explained in the rest of the section where two different models of increasing complexity are presented.

A. Model A

As a first attempt, we incorporate in the Markov model of Section II-B the information about the number of packets currently stored in the queue of the tagged station. The resulting model, named model "A", comprises the states belonging to the set $\{b_{i,j}\ ,\ B_{i,j}\}$, where b and B have the same meaning as explained in Section II-B. The two indexes $0 \le i \le m$ and $0 \le j \le K$ stand for the backoff stage and the number of packets currently stored in the buffer, respectively. Transition probabilities can be easily derived from those reported in Table I, considering all possible variations in the number j of packets that can occur during a time step due to packet arrivals and departures at/from the queue.

We observe that a station transmits a packet in all states $b_{i,j}$, provided that $j \ge 1$. With probability $1-p$, the transmission is successful; thus $\Delta = T_s$ and one packet departs from the queue (the one that is successfully transmitted). With probability p we have a collision; thus $\Delta = T_c$ and no packet leaves the queue, except for states $b_{m,j}$ where the colliding packet is discarded because the maximum number of retransmission attempts has been reached. In states $b_{i,0}$[3] and $B_{i,j}$ the station does not transmit, thus the duration of the step depends on what the other stations are doing.

The average probability τ that a station transmits in an arbitrary time step is given by

$$\tau = \sum_{i=0}^{m} \sum_{j=1}^{K} \pi_{b_{i,j}}$$

We define the probabilities Π'_s, Π'_c and Π'_σ, that a time step is occupied by a successful transmission, a collision, or an idle slot, respectively, given that the tagged station does not transmit. Similarly to the basic model for saturated sources, we assume that the state of a station is independent of that of the others. Relying on this critical assumption, we can write

$$\begin{aligned}
\Pi'_\sigma &= (1-\tau)^{n-1} \\
\Pi'_s &= (n-1)\ \tau\ (1-\tau)^{n-2} \\
\Pi'_c &= 1 - \Pi'_\sigma - \Pi'_c
\end{aligned}$$

As a special case, if $n = 1$, we have $\Pi'_\sigma = 1$, $\Pi'_s = \Pi'_c = 0$.

The whole set of transition probabilities of the Markov chain is not reported here for the sake of brevity, however it can be easily derived from the discussion above. Notice that, based on the considerations done at the end of Section III, in state $b_{0,0}$ a new backoff value is selected upon arrival of a new packet.

The conditional collision probability p is still given by (1), if we rely on the independence assumption among stations. From the solution of the model, obtained with a fixed point approximation similar to the one used under saturated conditions, we obtain many significant performance measures. Besides the aggregate throughput T_P (see (2)), the model provides the entire distribution of the number of packets queued at a station. We denote by $\tilde{N}$ the number of packets in the queue at the discrete time instants of the embedded Markov chain, and by N the number of packets in the queue at any time. A key metric is the average buffer occupancy, $E[N]$, from which one can easily derive the average MAC queueing delay by applying Little's formula. Finally, the model provides the packet loss probability due to buffer overflow, as well as the discard probability due to the maximum retry limit.

We show detailed results when $n = 10$ stations, using the same parameters as in Table II and considering the basic channel access scheme. Similar results are obtained for the RTS/CTS mechanism, and for different values of n. The buffer size K is set to 20 packets.

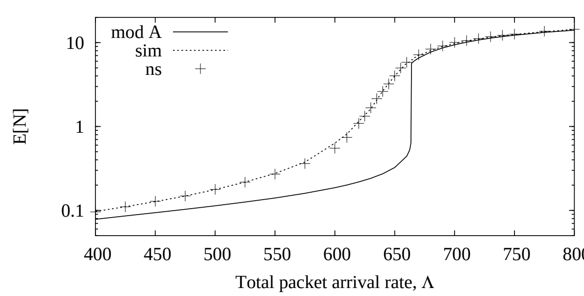

Fig. 6. Average buffer occupancy as a function of the total packet arrival rate, for $n = 10$ stations

In Figure 6 we show the average buffer occupancy as a function of Λ, comparing the results obtained with *mod A*, *sim* and *ns-2*. We observe a close match between *sim* and *ns-2*, while *mod A* significantly underestimates the number of packets in the queue, up to the point $\Lambda = 663$ packets/s, after which the prediction is accurate.

[3]Actually, there is only $b_{0,0}$ because it is not possible to have zero packets in states $b_{i,j}$, for all $i > 0$.

As a consequence, the model overestimates the aggregate throughput (see Figure 7), especially at the knee of the curve. Also, note that, at the knee of the curve, the model, as well as the *sim* and the *ns-2* simulators predict a higher value than the saturation throughput achieved under the same conditions (which is about 625 packets/s, as shown in Figure 4). This is because, under non-saturated conditions, the collision probability is smaller than with saturated sources, thus allowing for higher throughput. As expected, all curves converge to the saturation throughput when we further increase the arrival rate Λ.

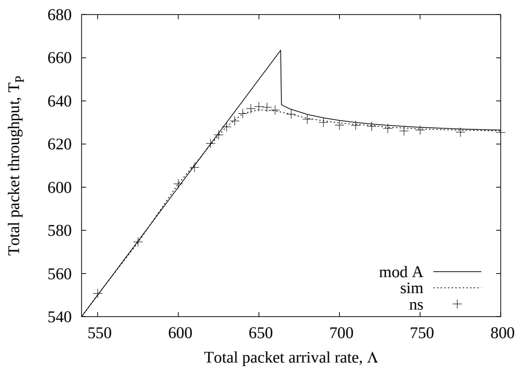

Fig. 7. Aggregate throughput of the total packet arrival rate, for $n = 10$ stations

The error that we encounter in the model is essentially due to the assumption that the state of individual stations are independent. This assumption was found to be correct under saturated conditions, but cannot be applied to non-saturated stations. To prove this important finding, we inspect the distribution of the number of competing stations, which are the stations having at least one packet in their queue at the beginning of a time step. We denote by $\tilde{C}$ the number of competing stations at the discrete time instants of the embedded Markov chain, and by C the number of competing stations at any time. The probability $\tilde{\gamma}$ that the buffer of a tagged station is not empty at the beginning of a step can be easily derived from the model as,

$$\tilde{\gamma} = \sum_{i=0}^{m} \sum_{j=1}^{K} (\pi_{b_{i,j}} + \pi_{B_{i,j}})$$

If stations were independent, we would expect to see a binomial distribution of the number of competing stations, with parameters $(\tilde{\gamma}, n)$. In Figure 8 we show the distribution of $\tilde{C}$ derived from *mod A* and *sim*, when we fix Λ to 640 packets/s.

It can be seen that the distribution obtained by *sim* deviates substantially from a binomial distribution; in particular, the average number of competing stations is dramatically underestimated by the model, as shown in Figure 9, where we compare the value of $E[C]$ obtained through *mod A*, *sim* and *ns-2*, as Λ varies.

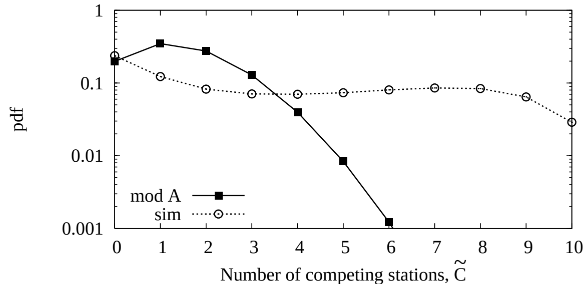

Fig. 8. Distribution of the number $\tilde{C}$ of competing stations, for $\Lambda = 640$ and $n = 10$ stations

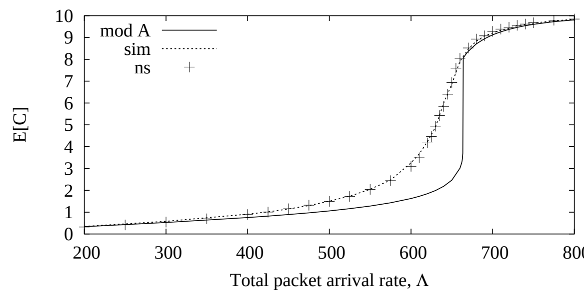

Fig. 9. Average number $E[C]$ of competing stations as a function of the total packet arrival rate, for $n = 10$ stations

Our conclusion is that *an 802.11 network under non-saturated conditions cannot be correctly analyzed relying on the independence assumption among stations*.

Interestingly, this conclusion would not apply if we had a hypothetical system in which: $T_c = T_s = \sigma = 1$ ms. In this case we have found that the results derived through *mod A* and *sim* coincide perfectly for all values of Λ, i.e., the assumption of independence among stations holds. This scenario, however, is quite unrealistic: small values of σ relatively to T_c and T_s are needed to obtain small delays and an efficient utilization of the radio resources.

To address the problem above it is necessary to model the evolution of the number of competing stations; we do this in the following section.

B. Model B

The description of the state of the system can be extended by keeping track of the the number of stations having non-empty transmission queue. Following this approach, we describe in detail the behavior of a tagged station using the same states of *mod A*, and we add to each state of the tagged station an indication of the number k of stations (excluding the tagged one) having at least one packet in the queue.

The state space of the resulting model, named model "B", is the set $\{b_{i,j,k}\ , B_{i,j,k}\}$, where the newly introduced index k takes values in the range $[0 \cdots n-1]$.

The transition probabilities of the model extend those of model "A" considering possible variations in the number k of competing stations during a time step. Such a number may vary due to the following events: i) one or more of the stations having an empty buffer receive new data to send, thus increasing k; ii) one of the stations having non-empty buffer successfully transmits a packet leaving an empty queue, thus decreasing k by one. Notice that these two events can occur simultaneously during the successful transmission of a packet.

The number of stations that join the competing set during a time step depends on the duration Δ of the step and on the current value of k. During an interval Δ, a station's queue receives at least one packet with probability $q = 1 - e^{-\lambda\Delta}$. The number of stations that join the set of competing stations is thus distributed according to a binomial distribution with parameters $(q, N-1-k)$.

The possible durations Δ of a time step are related to the probabilities that the channel is occupied by a successful transmission, a collision, or an idle slot. To compute such probabilities, we need to specify the probability $\tau(\tilde{C})$ that one of the stations belonging to the competing set transmits in a given time step. This set may include also the tagged station, thus $0 \le \tilde{C} \le n$.

In a generic state of the model, we have $\tilde{C} = k$ if $j = 0$, whereas $\tilde{C} = k+1$ if $j > 0$.

The probability $\tau(\tilde{C})$, $\tilde{C} \ge 1$, is given by the formula:

$$\tau(\tilde{C}) = \frac{\sum_{i=0}^{m}\sum_{j=1}^{K}\pi_{b_{i,j,\tilde{C}-1}}}{\pi_{b_{0,0,\tilde{C}}} + \pi_{B_{0,0,\tilde{C}}} + \sum_{i=0}^{m}\sum_{j=1}^{K}(\pi_{b_{i,j,\tilde{C}-1}} + \pi_{B_{i,j,\tilde{C}-1}})}$$

If $\tilde{C} = 0$, we obviously have $\tau(0) = 0$. Note that, differently from the models previously described in the paper, the probability $\tau(\tilde{C})$ may vary from one state to another, as it depends on $\tilde{C}$.

The tagged station transmits a packet in all states $b_{i,j,k}$, provided that $j \ge 1$. The *conditional collision probability* is given by

$$p(\tilde{C}) = 1 - [1 - \tau(\tilde{C})]^{\tilde{C}-1}$$

With probability $1 - p(\tilde{C})$, the transmission is successful.

In states $b_{i,0,k}$ and $B_{i,j,k}$, the station does not transmit, and the probabilities $\Pi'_s(\tilde{C}, k)$, $\Pi'_c(\tilde{C}, k)$ and $\Pi'_\sigma(\tilde{C}, k)$ that the channel is occupied by a successful transmission, a collision, or an idle slot, respectively, are functions of k and of the total number of competing stations $\tilde{C}$:[4]

$$\begin{aligned}
\Pi'_\sigma(\tilde{C}, k) &= [1 - \tau(\tilde{C})]^k \\
\Pi'_s(\tilde{C}, k) &= k\ \tau(\tilde{C})\ [1 - \tau(\tilde{C})]^{k-1} \\
\Pi'_c(\tilde{C}, k) &= 1 - \bar{\Pi}_\sigma(\tilde{C}, k) - \bar{\Pi}_c(\tilde{C}, k)
\end{aligned}$$

As a special case, if $k = 0$, we have $\bar{\Pi}_\sigma(\tilde{C}, 0) = 1$, $\bar{\Pi}_s(\tilde{C}, 0) = \bar{\Pi}_c(\tilde{C}, 0) = 0$.

The last parameter that we need to specify is the probability P_E that, upon successful transmission of a packet, a station other than the tagged one finds itself with an empty buffer, thus decreasing the number of competing stations. This quantity turns out to be the most critical to estimate by our model, as we do not maintain state information about the buffer occupancy at each station. A first approach is to assume that P_E depends only on the number $\tilde{C}$ of competing stations, and derive this probability from the buffer occupancy probabilities of the tagged station when attempting a transmission. This approach is motivated by the fact that a given station is likely to behave similarly to all other competing stations. Thus, we write

$$P_E(\tilde{C}) = \frac{\sum_{i=0}^{m}\pi_{b_{i,1,\tilde{C}-1}}}{\sum_{i=0}^{m}\sum_{j=1}^{K}\pi_{b_{i,j,\tilde{C}-1}}} e^{-\lambda T_S} \qquad (3)$$

Notice that, besides having a single packet in the queue at the beginning of a transmission interval, no packets must arrive during the corresponding step in order to leave an empty queue (this explains the factor $e^{-\lambda T_S}$). We have found that the model based on this estimate of P_E, referred to as *mod B'*, tends to underestimate the congestion level of the network, resulting in an average buffer occupancy smaller than what is obtained with *ns-2* (or *sim*). The error is shown in Figure 10 for the same case considered in Figure 6, and it is due to additional, neglected correlations among the states of competing stations.

A refined estimate of P_E considers this probability dependent on both the number $\tilde{C}$ of competing stations and the backoff stage i. In particular, a good approximation is obtained assuming that, whenever the tagged station is at backoff stage i, the other competing stations are at a backoff stage h that differs from i at most by one, i.e., $|h - i| \le 1$. The intuition behind this assumption is that the backoff stage of the tagged station will not differ *significantly* from the backoff stage of all other competing stations. Under this assumption, we obtain the refined estimate of P_E,

$$P_E(\tilde{C}, i) = \frac{\sum_{h:|h-i|\le 1}\pi_{b_{h,1,\tilde{C}-1}}}{\sum_{h:|h-i|\le 1}\sum_{j=1}^{K}\pi_{b_{h,j,\tilde{C}-1}}} e^{-\lambda T_S} \qquad (4)$$

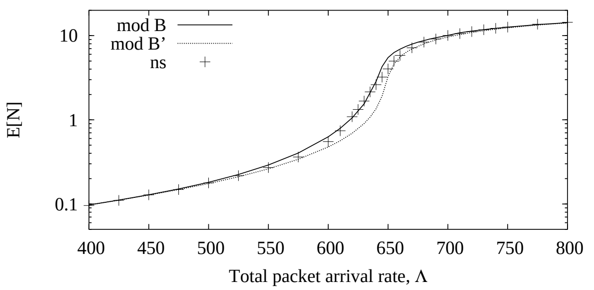

Fig. 10. Average buffer occupancy $E[N]$ as a function of the total packet arrival rate, for $n = 10$ stations

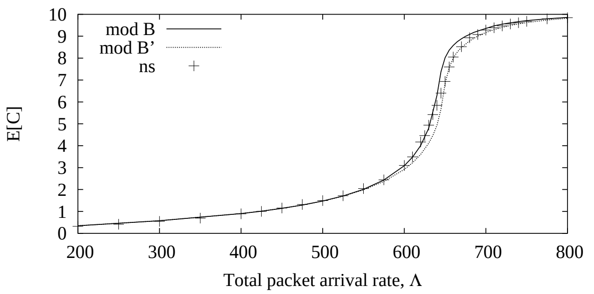

Fig. 11. Average number $E[C]$ of competing stations as a function of the total packet arrival rate, for $n = 10$ stations

The corresponding model, referred to as *mod B*, provides a better prediction of the system behavior, especially for values of Λ smaller than the saturation throughput. In Figures 10 and 11 we show the close match with *ns-2* simulations for the average buffer occupancy and the average number of competing stations, respectively, as well as the differences with respect to *mod B'*. The model is able to compute not only average values, but the entire distributions of the number of packets in the queue and of competing stations. Figures 12 and 13 show the good agreement among the distributions obtained with *mod B*, *sim* and *ns-2* for various values of Λ, thus proving the accuracy of our approach.

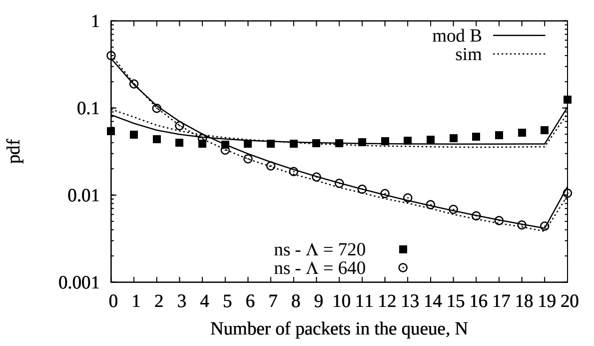

Fig. 12. Queue length distributions obtained for $n = 10$ stations, $\Lambda = 640$ and 720 packets/s

[4]Or, equivalently, they are functions of k and j.

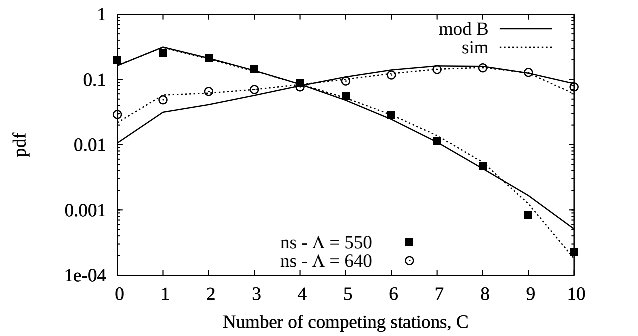

Fig. 13. Distributions of the number C of competing stations for $n = 10$ stations, $\Lambda = 550$ and 640 packets/s

C. Models Complexity

In this section we briefly discuss the computational complexity of our models. The number of states of *mod A* is $O(mK)$, where m is the retransmission limit and K is the maximum buffer size. The number of states of *mod B* (or *mod B'*) is $O(mKn)$, where n is the total number of stations. Note that, while *mod A* is not able to provide satisfactory results, we have shown that $O(mKN)$ states are sufficient to obtain very accurate results of detailed distributions. It is still an open question whether it would be possible to obtain satisfactory performance predictions with a reduced number of states.

Both *mod A* and *mod B* (or *mod B'*) require a fixed point approximation to converge. In terms of solution complexity, *mod B* can be very efficient (a few seconds) in case of limited numbers stations and small buffer sizes, such that the product Kn is smaller than, say, one hundred. For larger numbers of stations in the network and/or buffer sizes, the analytical solution is usually faster than a simulation with *ns-2*, but can be slower than the discrete simulator *sim*, which would be, in this case, the most efficient way to explore the parameter space.

IV. APPLICATIONS

Here we present some possible extensions and application examples of our analytical models for the study of some interesting performance issues in 802.11 WLANs. These examples are intended to show the effectiveness and flexibility of our modelling approach, not to provide an exhaustive study of the considered issues.

A. Variable Frame Sizes

The extension of the model to the case of variable frame size is rather straightforward, provided that we know the distribution of the payload size of the packets received from the upper layers. Indeed, we simply have to compute the mean value of the duration of a successful interval, $\bar{T}_s$, and of a collision interval, $\bar{T}_c$, and use these values in place of the constant values T_s and T_c introduced in Section II-A. The only difficulty is in the computation of $\bar{T}_c$, where we need to account for the longest packet payload involved in a collision. As suggested in [4], one can neglect the probability of three or more packets simultaneously colliding, with marginal impact on results, and derive a simple expression of $\bar{T}_c$.

B. Impact of Bursty Traffic

So far we have considered only the case of Poisson packet arrivals at the stations' MAC buffers. However, our model can account for a general arrival process of packets, provided that we can characterize the distribution of the number of packets that arrive during each channel occupation interval.

While a Poisson process is suitable to describing streaming traffic carried by UDP, such as voice and video, it is not adequate to represent bursty traffic like the one produced by TCP, or sporadic on-off traffic sources. One simple way to extend the model to more general traffic patterns is to assume a Poisson batched arrival process, in which groups of packets arrive simultaneously at the buffer. This process has been found to characterize very well the burstiness produced by window-based protocols such as TCP [19]. For simplicity, here we assume a constant batch size X, and we vary X to see what is the impact of different levels of burstiness on the performance of the 802.11 DCF. The case of $X = 1$ corresponds to a standard Poisson process. The arrival rate of batches is $\lambda_b = \lambda/X$, and the number of batches that arrive during a generic interval Δ has a Poisson distribution with mean $\lambda_b\Delta$.

We consider the basic access scheme with the usual parameters as in Table II, and we assume $n = 5$ stations equipped with MAC buffers of size $K = 40$. Under these conditions, the saturation throughput of the network is about 663 packets/s.

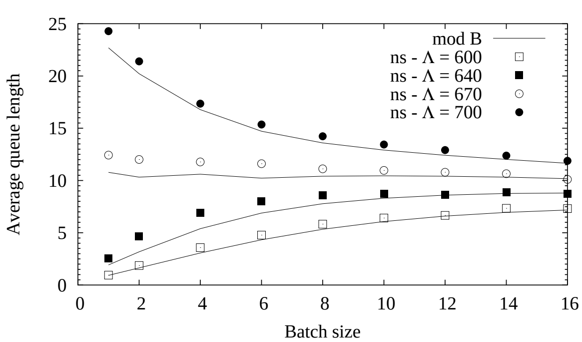

Fig. 14. Average queue length as a function of the batch size, when $n = 5$, $K = 40$, and for different values of Λ

In Figure 14 we show the average buffer occupancy $E[N]$ as a function of the batch size X, for different values of the aggregate arrival rate Λ, comparing results of *mod B* and *ns-2*. We observe that, for the same aggregate arrival rate of packets, the impact of the batch size can be huge, especially if Λ is far away from the saturation throughput. Below saturation, the burstiness dramatically increases the average queue length, while the opposite is true above saturation, as typically happens in most queuing systems. The agreement of analytical and simulation results is excellent.

The model is also able to compute the average packet loss probability. In our systems, packets are dropped either because of buffer overflow at the MAC layer, or because of exceeded retransmission limit. In our scenario, the impact of the second event is negligible with respect to the first one. In Figure 15 we show the total packet loss probability as a function of the batch size X, for the same values of Λ considered in Figure 14, comparing results of *mod B* and *ns-2*. Again, our analytical predictions are quite accurate.

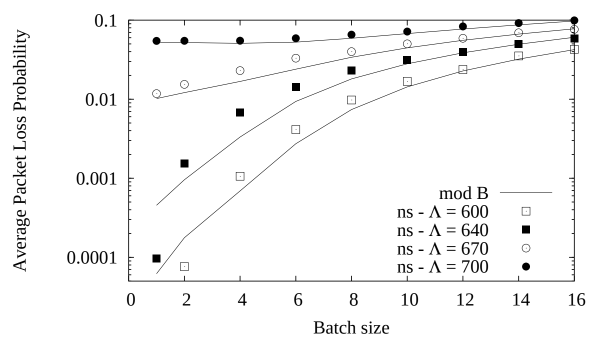

Fig. 15. Total packet loss probability as a function of the batch size, when $n = 5$, $K = 40$, and for different values of Λ

We remark that our model is able to compute the fundamental parameters that drive the performance of TCP: the packet loss probability and the RTT. In fact, the variable part of the RTT is essentially due to queueing delay at the MAC level, which can be immediately computed from the already obtained average queue length. Therefore (assuming that the wireless network is the bottleneck), our model could be coupled in closed loop with an analytical model of TCP to compute the arrival rate of packets and modified traffic burstiness in reaction to losses and delays.

C. Impact of Batch Services

Through our approach, we can also evaluate the effectiveness of some techniques that have been proposed to reduce channel access delay in 802.11 networks. In particular, according to the IEEE 802.11e draft standard, a station successfully accessing the channel is allowed to send up to M frames, with $M \geq 1$, thereby reducing the delay arising from the channel access and improving the efficiency of resource utilization. This behavior can be regarded as batch services in the queue of a station, i.e., the dual of the batch arrival process examined in the previous section. Our model has all information to account for batch services, as it maintains the number of packets waiting for transmission in the queue. Notice however that, since more than one frame can be sent during a successful transmission, it is necessary to recompute the average duration of T_c. Finally, we could consider the joint impact of batch arrivals and batch services.

D. Multirate Environments

The 802.11b standard allows stations to send the payload of data frames at four different bit rates, namely, 1, 2, 5.5, and 11 Mb/s. In this section we explore the impact on overall network performance of a mixture of n_H stations operating at *high* (11 Mb/s) rate, and n_L stations transmitting at *low* (2 Mb/s) rate. (A similar network scenario was studied in [20], in the case of an infrastructured network.) The number n still denotes the total number of stations, while $f = n_L/n$ is the fraction of low-rate stations.

We extend our model to such multirate environment by computing the average duration of a successful or collision interval, similarly to the case of variable frame sizes (Section IV-A). We start analyzing saturated conditions, fixing $n = 20$ and varying the number of low-rate stations. Parameters are set as in Table II, but we consider different values of payload size (the same for all stations), as this parameter has an important impact on results. Figure 16 reports the aggregate data throughput T_b in bit/s for both the basic access and RTS/CTS scheme. We observe the throughput degradation with the increase of the number of low-rate stations, the more significant the larger the size of data packets used.

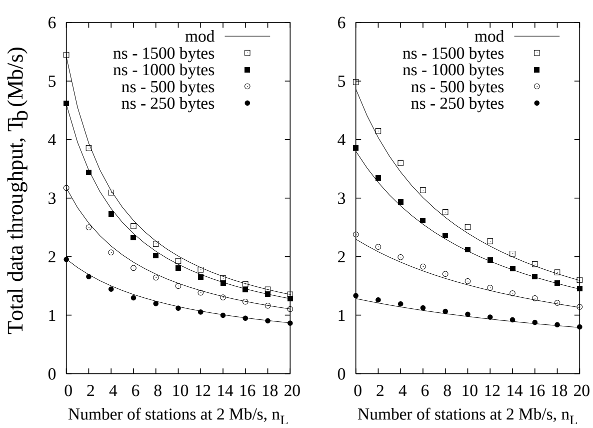

Fig. 16. Total data throughput as a function of the number of stations operating at 2 Mb/s, in case of $n = 20$, using different payload sizes, for the basic access (left plot) and RTS/CTS scheme (right plot)

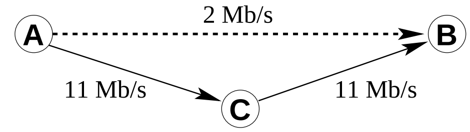

Fig. 17. Station **A** can send directly to **B** using 2 Mb/s data rate, or pass traffic through station **C** by two hops at 11 Mb/s data rate

These results allow us also to address an interesting issue that emerges in multirate environments where stations can relay traffic on behalf of others. In 802.11 networks, stations transmitting at higher rate have a shorter radio range. Thus, the following question arises [21]: Is it preferable (in terms of overall network performance) that a station reaches its destination by a single hop at *low rate* or by two hops at *high rate*? This dilemma is illustrated in Figure 17, and can appear in both infrastructure-based [22] and pure ad-hoc wireless networks [21].

The problem is that, while high-data rates improve network utilization, more traffic has to be sent in case of multihop communications. Recall that we assume that only one successful transmission can happen at a time throughout the network. If distant stations decide to operate at low rate, the average goodput (traffic delivered to the destination) per station is T_b/n, and it is fairly the same for all stations. If distant stations decide to use two hops, all stations send at 11 Mb/s achieving an overall throughput $T_b' > T_b$, but the average goodput per station is $T_b'(1-f)/n$, because we have to discount the additional traffic sent by distant stations in their first hop. Looking at Figure 16, we observe that the best choice jointly depends on payload size, fraction of low-rate stations, and access scheme. For example, in case of 500 bytes payload, two-hops are better up to $n_L = 10$ for the basic access scheme, but only for $n_L = 1$ in case of the RTS/CTS scheme.

Similar problems can be studied also under non-saturated conditions. In this case, we assume that the network has spare capacity to accommodate the additional traffic due to multihops, and we want to minimize the average delay to reach the destination. As the general problem is quite complex, we look at preliminary results obtained in one particular scenario. We consider n_H fast stations, which can transmit at 11 Mb/s because either they are close enough to the intended destination or they can use a nearby relay to forward packets toward the destination, and n_L slow stations which opt for sending data at low-rate, directly to their destination. We set the payload size to 500 bytes and the buffer size K to 20; furthermore, we assume that all stations generate their own traffic at 50 packets/s, according to a Poisson process.

Figure 18 reports the mean channel access delay of a station, computed as the average queueing delay of a packet in the MAC buffer, as a function of the number n_H of high-rate stations, and for different numbers n_L of (additional) low-rate stations. Note that packets generated by sources that use a relay (i.e., a two-hop path) have to access the channel twice, thus experiencing (on average) double the delay predicted in a single-hop network where all stations operate at 11 Mb/s. Nevertheless, Figure 18 shows that, in general, the use of multihop communications is to be preferred. Indeed, the presence of just a few low-rate stations degrades performance dramatically, because delays increase considerably, and fewer high-rate stations can coexist before buffers saturate and start dropping packets.

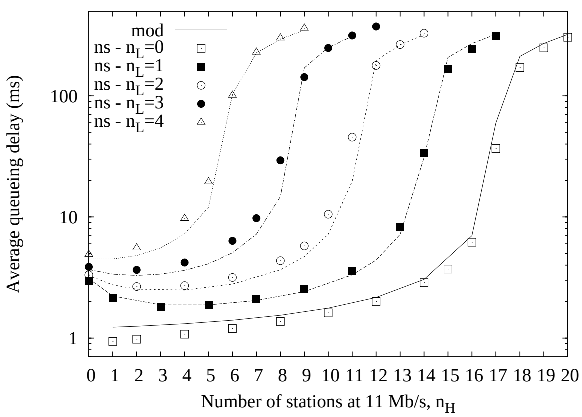

Fig. 18. Average queueing delay as a function of the number of stations operating at 11 Mb/s and different numbers of (additional) stations at 2 Mb/s, in the case of 500 bytes payload and basic access scheme

V. Conclusions and Future Work

In this paper we identified the critical assumptions to develop analytical models for 802.11 WLANs, both under saturated and non-saturated conditions. Our models enabled us to compute several metrics of interest, and are general enough to account for several aspects of 802.11 networks, among which, different traffic patterns and multirate environments.

Future work will address the case where not all stations are in each other's radio proximity, and will further investigate the performance of multihop networks.

References

[1] Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specifications ISO/IEC 8802-11:1999(E); ANSI/IEEE Std 802.11, 1999 Edition, USA.

[2] IEEE 802.11 WG, Draft Supplement to Part 11: Wireless Medium Access Control (MAC) and Physical Layer (PHY) specifications: Medium Access Control Enhancements for Quality of Service, IEEE 802.11e/Draft 3.0, Aug. 2002.

[3] F. Calì, M. Conti, and E. Gregori, "Dynamic Tuning of the IEEE 802.11 Protocol to Achieve a Theoretical Throughput Limit," *IEEE/ACM Transactions on Networking*, Vol. 8 , No. 6, pp. 785–799, Dec. 2000.

[4] G. Bianchi, "Performance Analysis of the IEEE 802.11 Distributed Coordination Function," *IEEE Journal on Selected Areas in Communications*, Vol. 18, No. 3, pp. 535–547, Mar. 2000.

[5] R. Litjens, F. Roijers, J.L. van den Berg, R.J. Boucherie, and M. Fleuren, "Performance Analysis of wireless LANs: an Integrated Packet/Flow Level Approach," *ITC Conference*, Berlin, Germany, Aug. 2003.

[6] V.M. Vishnevskii and A.I. Lyakhov, "IEEE 802.11 Wireless LAN: Saturation Throughput Analysis with Seizing Effect Consideration," *Cluster Computing*, Vol. 5, No. 2, Kluwer Academic Publishers, pp. 133–144, 2002.

[7] P. Chatzimisios , A.C. Boucouvalas, and V. Vitsas, "IEEE 802.11: Packet Delay - A Finite Retry Limit Analysis," *IEEE GLOBECOM*, San Francisco, CA, Dec. 2003.

[8] P. Chatzimisios , A.C. Boucouvalas, and V. Vitsas, "Performance Analysis of IEEE 802.11 DCF in the Presence of Transmission Errors," *IEEE ICC*, Paris, France, June 2004.

[9] E. Ziouva and T. Antonakopoulos, "CSMA/CA Performance under High Traffic Conditions: Throughput and Delay Analysis," *Computer Communications*, Vol. 25, No. 3, pp. 313–321, 2002.

[10] M.M. Carvalho and J.J. Garcia-Luna-Aceves, "Delay Analysis of IEEE 802.11 in Single-hop Networks," *IEEE ICNP*, Atlanta, Georgia, Nov. 2003.

[11] W. Yue and Y. Matsumoto, "An Exact Analysis for CSMA/CA Protocol in Integrated Voice/Data Wireless LANs," *IEEE GLOBECOM*, pp. 1805–10, Nov. 2000.

[12] C.H. Foh and M. Zukerman, "Performance Analysis of the IEEE 802.11 MAC Protocol," *EWC 2002*, Florence, Italy, pp. 184–190, Feb. 2002.

[13] E.M.M. Winands, T.J.J. Denteneer, J.A.C. Resing, and R. Rietman, "A Finite-Source Feedback Queueing Network as a Model for the IEEE 802.11 Distributed Coordination Function," *5th European Wireless Conference: Mobile and Wireless Systems beyond 3G*, Barcelona, Spain, pp. 551–557, Feb. 2004.

[14] H. Kim and J.C. Hou, "Improving Protocol Capacity with Model-based Frame Scheduling in IEEE 802.11-operated WLANs," *IEEE/ACM MOBICOM*, San Diego, CA, pp. 190–204, Sep. 2003.

[15] O. Tickoo and B. Sikdar, "Queueing Analysis and Delay Mitigation in IEEE 802.11 Random Access MAC based Wireless Networks," *IEEE INFOCOM*, Hong Kong, Mar. 2004.

[16] G.R. Cantieni, Q. Ni, C. Barakat, and T. Turletti, "Performance Analysis of Finite Load Sources in 802.11b Multirate Environments," *INRIA Research Report RR-4881*.

[17] G. Bianchi and I. Tinnirello, "Kalman Filter Estimation of the Number of Competing Terminals in an IEEE 802.11 Network," *IEEE INFOCOM*, San Francisco, CA, Apr. 2003.

[18] ns-2, The Network Simulator, `http://www.isi.edu/nsnam/ns/`

[19] M. Garetto and D. Towsley, "Modeling, Simulation and Measurements of Queuing Delay Under Long-Tail Internet Traffic," *ACM Sigmetrics*, San Diego, CA, June 2003.

[20] M. Heusse, F. Rousseau, G. Berger-Sabbatel, and A. Duda, "Performance Anomaly of 802.11b," *IEEE INFOCOM*, San Francisco, CA, Apr. 2003.

[21] B. Awerbuch, D. Holmer, and H. Rubens, "High Throughput Route Selection in Multi-rate Ad Hoc Wireless Networks," *1st IFIP WONS 2004*, Madonna di Campiglio, Italy, Jan. 2004.

[22] H. Zhu and G. Cao, "On improving the performance of IEEE 802.11 with multi-hop concepts," *IEEE ICCCN*, Dallas, TX, Oct. 2003.