Modeling the 802.11 Distributed Coordination Function with Heterogenous Finite Load

David Malone[1], Ken Duffy[1], Douglas J. Leith

***Abstract*— Analysis of the 802.11 CSMA/CA mechanism has received considerable attention recently. Bianchi [1] presents an analytic model under a saturated traffic assumption. Bianchi's model is accurate, but typical network conditions are non-saturated and heterogenous. We present an extension of his model to a non-saturated environment. The model's predictions, validated against simulation, accurately capture many interesting features of non-saturated operation. For example, the model predicts that peak throughput occurs prior to saturation. Our model allows stations to have different traffic arrival rates, enabling us to address the question of fairness between competing flows. Although we use a specific arrival process, it encompasses a wide range interesting traffic types including, in particular, VoIP.**

***Index Terms*— 802.11, CSMA/CA, non-saturated traffic, heterogenous network.**

I. Introduction

The 802.11 wireless LAN standard has been widely deployed during recent years and has received considerable research attention. The 802.11 MAC layer uses a CSMA/CA algorithm with binary exponential back-off to regulate access to the shared wireless channel. While this CSMA/CA algorithm has been the subject of numerous empirical studies, an analytic framework for reasoning about its properties remains notably lacking. Developing analysis tools is desirable not only because of the wide deployment of 802.11 equipment but also because the CSMA/CA mechanism continues to play a central role in new standards proposals such as 802.11e. A key difficulty in the mathematical modeling of the 802.11 MAC lies in the large number of states that may exist (scaling exponentially with the number of stations). In his seminal paper, Bianchi [1] addressed this difficulty by assuming that (i) every station is saturated (i.e. always has a packet waiting to be transmitted), (ii) the packet collision probability is constant regardless of the state or station considered and (iii) transmission error is a result of packets colliding and is not caused by medium errors. Provided that every station is indeed saturated, the resulting model is remarkably accurate. However, the saturation assumption is unlikely to be valid in real 802.11 networks. Data traffic such as web and email is typically bursty in nature while streaming traffic such as voice operates at relatively low rates and often in an on-off manner. Hence, for most real traffic the demanded transmission rate is variable with significant idle periods, i.e. stations are usually far from being saturated. Our aim in this paper is to derive a mathematical model of CSMA/CA that relaxes the restriction to saturated operation while retaining as much as possible of the attractive simplicity of Bianchi's model, in particular, the ability to obtain analytic relationships.

In Section II the general model is introduced and solved. In Section III its predictions are verified through ns2 simulation for homogenous stations and heterogenous stations that have one of two distinct arrival rates. In Section IV fairness in the heterogenous case is discussed. In Section V the model limitations, and its possible extension, are discussed. In Section VI other approaches to non-saturated modeling are discussed. Concluding remarks are in section VII.

II. Model of non-saturated heterogenous stations

Following the seminal paper of Bianchi [1], much of the analytic work on 802.11 MAC performance has focused on saturated networks where each station always has a packet to send. For notable examples, see [2], [3]. The saturation assumption enables queueing dynamics to be neglected and avoids the need for detailed modeling of traffic characteristics, making these networks particularly tractable.

Networks do not typically operate in saturated conditions. Internet applications, such as web-browsing, e-mail and voice over IP exhibit bursty or on-off traffic characteristics. Creating an analytic model that includes fine detail of traffic-arrivals and queueing behavior, as well as 802.11 MAC operation, presents a significant challenge. We introduce a model with traffic and buffering assumptions that make it sufficiently simple to give explicit expressions for the quantities of interest (throughput per station and collision probabilities), but still capture key effects of non-saturated operation. Although our traffic assumptions form only a subset of the possible arrival processes, we will see they are useful in modeling a wide range of traffic, including voice conversations. As in [1], our fundamental assumption is that each station has a fixed probability of collision when it attempts to transmit, irrespective of its history.

Bianchi [1] presents a Markov model where each station is modeled by a pair of integers (i, k). The back-off stage, i, starts at 0 at the first attempt to transmit a packet and is increased by 1 every time a transmission attempt results in a collision, up to a maximum value m. It is reset after a successful transmission. The counter, k is initially chosen uniformly between $[0, W_i - 1]$, where typically $W_i = 2^i W_0$ is the range of the counter and W_0 is the 802.11 parameter

Work supported by Science Foundation Ireland grant IN3/03/I346. The authors are with the Hamilton Institute, National University of Ireland, Maynooth, Co. Kildare, Ireland.

[1] Joint first authors.

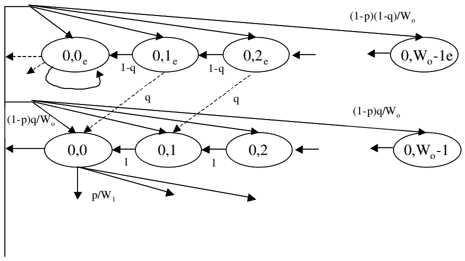


Fig. 1. Non-saturated Markov Chain.

CWmin. While the medium is idle, the counter is decremented. Transmission is attempted when $k = 0$.

We introduce new states $(0, k)_e$ for $k \in [0, W_0 - 1]$, representing a node which has transmitted a packet, but has none waiting. This is called postbackoff. The first two stages of the new chain are depicted in Figure 1. Note that $i = 0$ in all such states, because if $i > 0$ then a collision has occurred, so we must have a packet awaiting transmission.

We assume that for each station there is a constant probability $1 - q$ that the station's buffer has no packets awaiting transmission at the start of each counter decrement. This enables us to derive relationships between the per-station quantities: q, the probability of at least one packet awaiting transmission at the start of a counter decrement; m, the maximum backoff stage; p, the probability of collision given the station is attempting transmission; P, the Markov chain's transition matrix; b, the chain's stationary distribution; and τ, the stationary distribution's probability that the station transmits in a slot. These relationships can be solved for p and τ, and network throughput predicted. It is important to note that the Markov chain's evolution is not real-time, and so the estimation of throughput requires an estimate of the average state duration.

Under our assumptions, we have for $0 < k < W_i$

$$\begin{array}{rrcl} 0 < i \leq m, & P[(i, k-1)|(i, k)] & = & 1, \\ & P[(0, k-1)_e|(0, k)_e] & = & 1 - q, \\ & P[(0, k-1)|(0, k)_e] & = & q. \end{array}$$

If the counter reaches 0 and a packet is queued, then we begin a transmission. We assume there is a station-dependent probability p that other stations transmit at the same time, resulting in a collision. In the case of a collision we must increase the backoff stage (or discard). In the case of a successful transmission we return to backoff stage 0 and the station's buffer is empty with probability $1-q$. In the case with infinitely many retransmission attempts we need introduce no extra per-station parameters and for $0 \leq i \leq m$ and $k \geq 0$ we have

$$\begin{array}{rcl} P[(0, k)_e|(i, 0)] & = & \frac{(1-p)(1-q)}{W_0}, \\ P[(0, k)|(i, 0)] & = & \frac{(1-p)q}{W_0}, \\ P[(\min(i+1, m), k)|(i, 0)] & = & \frac{p}{W_{\min(i+1,m)}}. \end{array}$$

Naturally, these transitions could be adapted to allow discards after a certain number of transmission attempts.

The final transitions are from the $(0, 0)_e$ state, where postbackoff is complete, but the station's buffer is empty. In this case we remain in this state if the station's buffer remains empty. If a packet arrives we have three possibilities: successful transmission, collision or, if the medium is busy, the 802.11 MAC begins another stage-0 backoff, now with a packet. With P_{idle} denoting the probability that the medium is idle during a typical slot, the transitions from the $(0, 0)_e$ state are:

$$\begin{array}{rrcl} & P[(0, 0)_e|(0, 0)_e] & = & 1 - q + \frac{qP_{\text{idle}}(1-p)}{W_0}, \\ k > 0, & P[(0, k)_e|(0, 0)_e] & = & \frac{qP_{\text{idle}}(1-p)}{W_0}, \\ k \geq 0, & P[(1, k)|(0, 0)_e] & = & \frac{qP_{\text{idle}}p}{W_1}, \\ k \geq 0, & P[(0, k)|(0, 0)_e] & = & \frac{q(1-P_{\text{idle}})}{W_0}. \end{array}$$

Observe that p, the probability of a collision given that we are about to transmit, is the probability that at least one other station is transmitting. Using the assumption that station collision probabilities are history independent, this is also the probability that the medium is busy if we know the station under consideration is silent. Hence we substitute $P_{\text{idle}} = 1-p$.

Given the collision probability p for this station in the system and per-station parameters q, W_i and m we may solve for a stationary distribution of this Markov chain. This will enable us to determine the probability, τ, that this station is attempting transmission in a typical slot.

First we make observations that aid in the deduction of the stationary distribution. With $b(i, k)$ and $b(0, k)_e$ denoting the stationary probability of being in states (i, k) and $(0, k)_e$, as b is a probability distribution we have

$$\sum_{i=0}^{m} \sum_{k=0}^{W_i-1} b(i, k) + \sum_{k=0}^{W_0-1} b(0, k)_e = 1. \tag{1}$$

We will write all probabilities in term of $b(0, 0)_e$ and use the normalization in equation (1) to determine $b(0, 0)_e$. We have the following relations. To be in the sub-chain $(1, k)$, a collision must have occurred from state $(0, 0)$ or an arrival to state $(0, 0)_e$ followed by detection of an idle medium and then a collision, so that $b(1, 0) = b(0, 0)p + b(0, 0)_e q(1 - p)p$. Neglecting packet discard, for $i > 1$ we have $b(i, 0) = p^{i-1} b(1, 0)$ and so

$$\sum_{i \geq 1} b(i, 0) = \frac{b(1, 0)}{1 - p} = \frac{b(0, 0)p + b(0, 0)_e q(1 - p)p}{1 - p}. \tag{2}$$

The keystone in the calculation is then the determination of $b(0, W_0-1)_e$. Transitions into $(0, W_0-1)_e$ from $(0, 0)_e$ occur if there is an arrival, the medium is sensed idle and no collision occurs. Transitions into $(0, W_0 - 1)_e$ also occur from $(i, 0)$ if no collision and no arrival occurs

$$b(0, W_0 - 1)_e = b(0, 0)_e \frac{q(1-p)^2}{W_0} + \frac{(1-p)(1-q)}{W_0} \sum_{i \geq 0} b(i, 0). \tag{3}$$

Combining equations (2) and (3) gives

$$b(0, W_0 - 1)_e = b(0, 0)_e \frac{(1 - p)q(1 - pq)}{W_0} + b(0, 0) \frac{1 - q}{W_0}.$$

We then have for $W_0 - 1 > k > 0$, $b(0, k)_e = (1 - q)b(0, k + 1)_e + b(0, W_0 - 1)_e$, with $b(0, k)_e$ on the left hand side replaced by $qb(0, 0)_e$ if $k = 0$. Straightforward recursion leads to expressions for $b(0,k)_e$ in terms of $b(0,0)_e$ and $b(0,0)$, and so we find

$$\frac{b(0,0)_e}{b(0,0)} = \frac{1-q}{q}\left(\frac{1-(1-q)^{W_0}}{qW_0-(1-p)(1-pq)(1-(1-q)^{W_0})}\right). \qquad (4)$$

Using these equations we can determine the second sum in equation (1)

$$\sum_{k=0}^{W_0-1} b(0,k)_e = b(0,0)_e \frac{qW_0}{1-(1-q)^{W_0}}.$$

The $(0,k)$ chain can then be tackled, starting with the relation

$$b(0,W_0-1) = \sum_{i\geq 0} b(i,0)\frac{(1-p)q}{W_0} + b(0,0)_e\frac{qp}{W_0}.$$

Iteration leads to

$$\begin{aligned}\sum_{k=0}^{W_0-1} b(0,k) &= b(0,0)_e \left[\frac{q}{1-q}\frac{W_0+1}{2}\right.\\ &\left(\frac{q^2W_0}{1-(1-q)_0^W} + p(1-q) - q(1-p)^2\right)\\ &\left.+\frac{qW_0(qW_0+q-2)}{2(1-(1-q)_0^W)} + 1 - q\right].\end{aligned}$$

Using equation (4) we can determine $b(1,0)$ in terms of $b(0,0)_e$:

$$b(1,0) = b(0,0)_e\frac{pq^2}{1-q}\left(\frac{W_0}{1-(1-q)_0^W} - (1-p)^2\right).$$

Finally, after algebra, the normalization (1) gives

$$\begin{aligned}1/b_{(0,0)_e} &= (1-q) + \frac{q^2W_0(W_0+1)}{2(1-(1-q)^{W_0})}\\ &+\frac{q(W_0+1)}{2(1-q)}\left(\frac{q^2W_0}{1-(1-q)^{W_0}} + p(1-q) - q(1-p)^2\right)\\ &+\frac{pq^2}{2(1-q)(1-p)}\left(\frac{W_0}{1-(1-q)^{W_0}} - (1-p)^2\right)\\ &\left(2W_0\frac{1-p-p(2p)^{m-1}}{1-2p} + 1\right).\end{aligned} \qquad (5)$$

The main quantity of interest is τ, the probability that the station is attempting transmission. A station attempts transmission if it is in the state $(i,0)$ (for any i) or if it is in the state $(0,0)_e$, a packet arrives and the medium is sensed idle. Thus $\tau = q(1-p)b(0,0)_e + \sum_{i\geq 0} b(i,0)$, which reduces to

$$\tau = b(0,0)_e\left(\frac{q^2W_0}{(1-p)(1-q)(1-(1-q)^{W_0})} - \frac{q^2(1-p)}{1-q}\right), \qquad (6)$$

where $b(0,0)_e$ is given in equation (5), so that τ is expressed solely in terms of p, q, W_0 and m. While q, W_0 and m are fixed for each station, in order to determine the collision probability, p, we must give a relation between the stations competing for the medium.

Consider the case where n stations are present, labeled $l = 1,\ldots,n$. Equation (6) gives an expression for τ_l, the per-station transmission probability, in terms of a per-station arrival process q_l and a per-station collision probability p_l. Observe that

$$1 - p_l = \prod_{j\neq l}(1-\tau_j), \text{ for } l = 1,\ldots,n, \qquad (7)$$

that is, there is no collision for station l when all other stations are not transmitting. With n stations, (6) and (7) provide $2n$ coupled non-linear equations which can be solved numerically for p_l and τ_l. Observe that $(1-p_i)(1-\tau_i)$ is the same for all $i = 1,\ldots,n$ and represents the probability that the medium is idle (as we observed before $1-p_i$ is the probability that other stations are silent and $1-\tau_i$ is the probability that this station is silent). Note that these equations imply that different stations' collision probabilities are not the same unless their transmission probabilities are equal. We remark that in the case where the stations are homogenous, the equations (7) reduce to $1-p = (1-\tau)^{n-1}$. Placing the system in saturation by setting $q = 1$, the model reduces to that of Bianchi [1], as expected.

The length of each state in the Markov chain is not a fixed period of real time. Each state may be occupied by a successful transmission, a collision or the medium being idle. To convert between states and real time, we must calculate the expected time spent per state, which is given by

$$\begin{aligned}E_s &= (1-P_{tr})\sigma + \textstyle\sum_{i=1}^n P_{s_i}T_{s_i}\\ &+ \textstyle\sum_{r=2}^n \sum_{1\leq k_1<\cdots<k_r\leq n} P_{c:k_1\ldots k_r}T_{c:k_1\ldots k_r},\end{aligned} \qquad (8)$$

where: $P_{s_i} = \tau_i\prod_{j\neq i}(1-\tau_j)$ is the probability station i successfully transmits; T_{s_i} is the expected time taken for a successful transmission from station i, which can easily be calculated from expected payload size, physical data rate and MAC parameters; $P_{c:k_1\ldots k_r} = \prod_{i=1}^r \tau_{k_r}\prod_{j\neq k_1\ldots k_r}(1-\tau_j)$, the probability that only the stations labeled k_1 to k_r experience a collision by attempting transmission; $T_{c:k_1\ldots k_r}$ is the expected time taken for a collision from stations labeled k_1 to k_r, which is readily calculated from payload size distributions, physical data rate and MAC parameters; $P_{tr} = 1-\prod_{i=1}^n(1-\tau_i)$ is the probability at least one station attempts transmission; and σ is the slot-time. See Table I for an example calculation of T_s and T_c with fixed payload sizes.

Once the mean state time is known, we can estimate the proportion of time that the medium is used by each station for successfully transferring data:

$$S_i = (P_{s_i}L_i)/E_s, \qquad (9)$$

where L_i is the expected time spent transmitting payload data for source i. The normalized throughput of the system is then

$$S = \sum_{i=1}^n S_i. \qquad (10)$$

Thus in order to determine the throughput and collision probability for each station, and the overall throughput, one first solves equations (7) using equations (5) and (6). Then one uses equations (8), (9) and (10).

To study fairness of the 802.11 MAC layer, we will solve the model for two groups of stations, where all stations within each group have the same station parameters including arrival rate and payload size. Suppose there are n_1 stations in the first class and n_2 stations in the second class, then we may solve for the collision probabilities p_1 and p_2 for a station in each group using (7):

$$\begin{aligned}1-p_1 &= (1-\tau_1)^{n_1-1}(1-\tau_2)^{n_2},\\ 1-p_2 &= (1-\tau_1)^{n_1}(1-\tau_2)^{n_2-1}.\end{aligned}$$

Letting T_s be the time for a successful transmission and T_c be the time for a collision,

$$E_s = (P_{s1} + P_{s2})T_s + (1 - P_{s1} - P_{s2})T_c + (1 - P_{tr})\sigma,$$

where P_{si} is the probability that a station in class i, $i = 1, 2$, successfully transmits. Normalized throughput for each class is $S_1 = P_{s1}L_1/E_s$ and $S_2 = P_{s2}L_2/E_s$, where L_i is the average payload duration for a station in class i.

III. MODEL VERIFICATION

We first consider a homogenous group of stations and then consider the heterogenous setting where each station has one of two arrival rates. Station parameters[1] are shown in Table I.

We compare predictions of the model from Section II with simulations using the ns2 based 802.11 simulator produced by TU-Berlin [4]. We compare model predictions with simulation for various numbers of stations and arrival rates.

In order to move between model and simulation arrival rates, we use the following logic. Queues are set as small as ns2 will permit and traffic arrivals are Poisson. Since we have small buffers, the parameter q_i is the probability that at least one packet arrives in the expected time spent per state, E_s defined in equation (8). In simulation, the probability that at least one packet arrives during E_s is one minus the probability that the first inter-packet time is greater than E_s. Hence, when inter-packet arrival times are exponentially distributed the exponential rate λ_i should be set so that $q_i = 1 - \exp(-\lambda_i E_s)$, i.e. $\lambda_i = -\log(1 - q_i)/E_s$. With λ_i so chosen, the arrival rate in the model and in simulation agree.

For the homogeneous case, Figure 2 shows how collision probability depends on the total normalized offered load. Figure 3 shows how the normalized throughput of the link depends on the total normalized offered load. In all cases there is good agreement between the model and simulations. The model has captured a number of important features of the behavior, including: the linear relationship between the offered load and throughput when well below saturation; the behavior of throughput as predicted by Bianchi's model and simulation at high offered loads (corresponding to saturation); for larger numbers of stations the maximum throughput is achieved before saturation in both the model and simulation; The point at which this maximum occurs is relatively insensitive to the number of stations; and a complex transition from under-loaded to saturated.

For the heterogenous setting of where stations are divided into two classes with each class having a different arrival rate, Figure 4 shows the model's normalized throughput prediction for a station in each class, with $n_1 = 12$ and $n_2 = 24$. The throughput is plotted against normalized arrival rate for a station in each class. We take a representative slice through this surface along the line where the arrival rate to the second group is $1/4$ of that of the first group. Figure 5 shows

[1] Note that the 802.11 standards do not specify a length for ACKTimeout. Thus the length of a collision may depend on whether a station was involved in the collision (including a vendor selected ACKTimeout) or was an onlooker (then using EIFS). We choose $T_c = T_s$, following the spirit for the 802.11 standard. For a model of what occurs when they are set differently in a saturated situation, see Robinson and Randhawa [3].

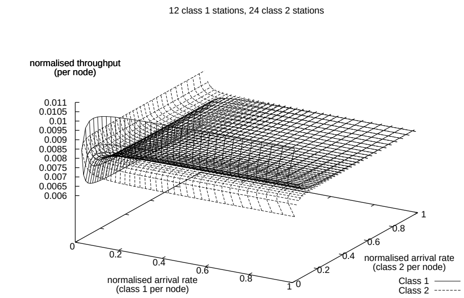


Fig. 4. Per-station throughput for two classes of stations offering different loads, $n_1 = 12$, $n_2 = 24$.

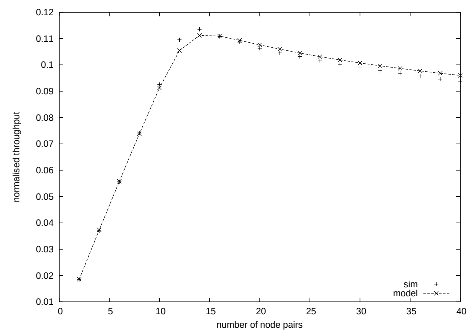


Fig. 6. Throughput for station-pairs sending 64kbps on-off traffic streams.

predicted and simulated throughputs and collision probabilities against overall normalized offered load. There is good match between predicted and observed throughputs, although the simulated collision probabilities are slightly lower than the model predicts. The collision probabilities of a station in each class are always close, but not the same. As commented after equation (7), this is expected because of an asymmetry in the system: a station in class 1 sees 11 other class 1 stations and 24 class 2 stations; a station in class 2 sees 12 class 1 stations and 23 class 2 stations.

We have taken a large number of slices for ranges of values of n_1 and n_2. For smaller numbers of users, we have found that while the predicted throughputs are accurate, the predicted collision probabilities are typically underestimates. For larger number of stations, the estimates' accuracy increases.

As a case-study we consider the predictions of the model in a situation that represents VoIP traffic in an ad-hoc network. Parameters for the voice calls are taken from [5]: 64kbs on-off traffic streams where the on and off periods are distributed with mean 1.5 seconds. Periods of less than 240ms are increased to 240ms in length, to reproduce the minimum talk-spurt period. Traffic is between pairs of stations; the on period of one station corresponds to the off period of another. When modeled, we treat each pair of stations as a single transmitter. Figure 6 shows the predicted and simulated throughput, as the number of station-pairs is increased. It can be seen that the model makes remarkably accurate throughput predictions.

W_0	31	L	364us	=	500.0 bytes @ 11Mbps
m	5	T_s	944us	=	Header + L + SIFS + δ + ACK + δ + DIFS
σ	20us	T_c	944us	=	Header + L + SIFS + δ + ACKTimeout
SIFS	10us	DIFS	50us	=	2σ + SIFS
δ	2us	ACK	304us	=	192 bits @ 1Mbps + 14 bytes @ 1Mbps

TABLE I

PARAMETERS VALUES FOR MODEL AND SIMULATION.

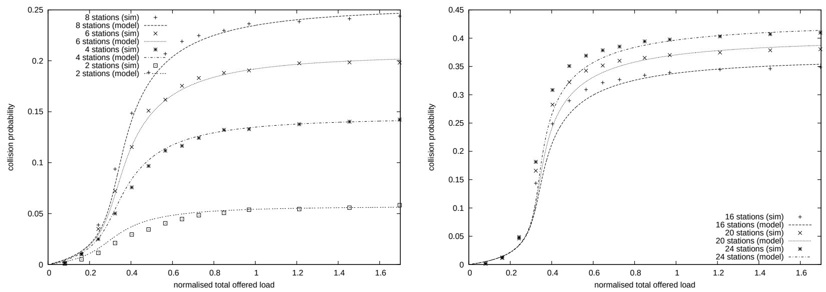


Fig. 2. Collision probability as the traffic arrival rate is varied.

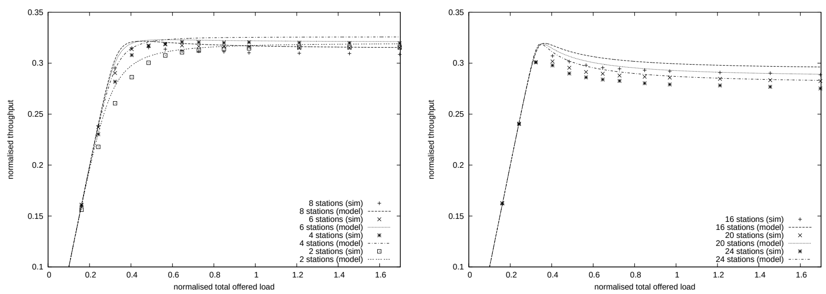


Fig. 3. Throughput as the traffic arrival rate is varied. For throughput rates below those shown there is agreement between the model and simulation.

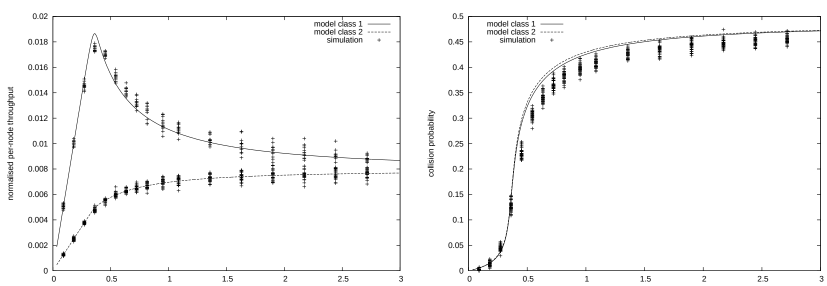


Fig. 5. Normalized per-station throughput and collision probability, where $n_1 = 12$, $n_2 = 24$ and the offered load of a class 2 station is 1/4 of a class 1 station.

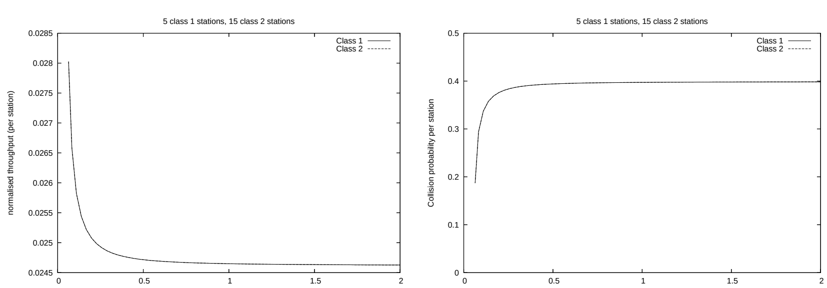


Fig. 8. Per-station throughput and collision probabilities for two classes of stations equal offered load, $n_1 = 5$, $n_2 = 15$. Class 1 and 2 throughput and collision probability are the same.

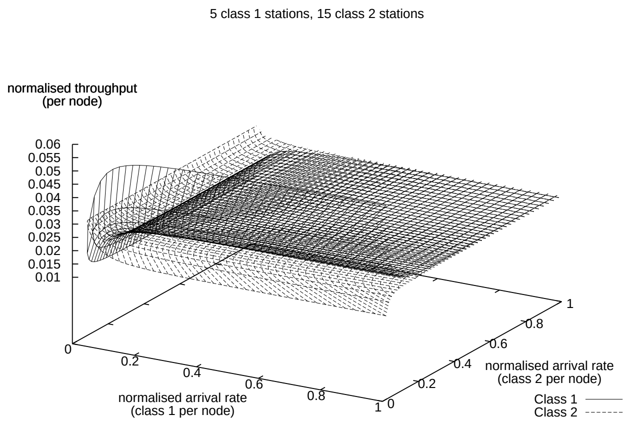


Fig. 7. Per-station throughput for two classes of stations offering different loads, $n_1 = 5$, $n_2 = 15$.

IV. FAIRNESS

Having validated the 2-class model in section III, we consider the model's predictions regarding protocol fairness. With $n_1 = 5$, $n_2 = 15$, Figure 7 shows the normalized throughput of a station in each class against the normalized offered load of a station in each class. Station parameters are those in Table I, but with 1500byte payloads. Taking a slice along the line where the offered load from stations in both classes are equal, shown in Figure 8, demonstrates fairness in this case. The collision probabilities and throughputs of all stations are equal.

Taking slices through Figure 7 when the offered loads of stations in each class differ, however, reveals long term unfairness that is different to known short-term issues. We fix the normalized arrival rate in class 1 per-station to be each of the four values 0.01, 0.02, 0.05 and 0.1 and vary the arrival rate per-station in class 2. Note that when class 1 stations offer 0.1 normalized load, although they are not saturated the offered load exceeds the network's capacity, even when no class 2 stations are present.

Overall normalized throughput and per-station collision probabilities are shown in Figure 9. Collision probabilities of stations in each class are approximately equal, with a maximum difference of 5% for the lowest class 1 offered load (0.01) and heavily loaded class 2 stations. At higher loads the overall channel throughput is insensitive to the class 1 arrival rate, but the bandwidth share does depend on the class 1 arrival rate; this is shown in Figure 10 where normalized throughput for a source in each class is shown against normalized offered load per source for a station in class 2.

In Figures 10 (a), (b) and (c), the network is underloaded for small class 2 offered load, so that the class 1 stations are not adversely affected by class 2. When the class 2 stations offer the same load as class 1 stations, the system is homogeneous and each station gets the same share of bandwidth. However, when the class 2 load ramps up beyond this level, class 1 stations lose their bandwidth share. The biggest drop from bandwidth fairness occurs when class 2 station are saturated, i.e. always have a packet ($q_2 = 1$). The percentage drop in throughput from fair share for these four class 1 offered loads are 16%, 32%, 22% and 8% for Figures 10 (a), (b), (c) and (d) respectively. The network is far from being fair, with greedy stations being able to steal bandwidth.

This unfairness has Quality of Service (QoS) implications. To demonstrate this we consider a scenario representing a single voice-call between two stations competing with stations carrying TCP connections. The voice-call pair is modeled as in Section III. The stations with TCP connections have 1500 byte payloads and are saturated. Figure 11 shows that collision probabilities are approximately equal for the VoIP and TCP stations, but the TCP sources steal bandwidth from the VoIP calls, with 5 TCP flows sufficient to reduce the VoIP throughput by 50%. Note that this is despite the fair-share of the channel for the VoIP station being roughly an order of magnitude above the throughput of the VoIP station (this share is not accessible due to the non-saturated nature of the VoIP traffic).

V. MODEL LIMITATIONS

We have shown that the Markov chain model has a relatively tractable analytic solution. We have considered a simple queue model in this paper, however it is clearly possible to introduce more complex traffic and queue models without adding further states into the Markov chain. For example, the q values could be calculated using more elaborate queueing modeling. Also, the probability that a station's buffer is empty immediately after successful transmission could be made dependent on the backoff stage at which that transmission took place. These probabilities could be obtained from traffic and queue modeling or even estimated from a running system. Alternatively, larger buffers could be explicitly modeled by significantly increasing the number of states in the Markov chain.

VI. RELATED WORK

There are alternative approaches to non-saturated modeling. In [6] a modification of [1] is considered where a probability of not transmitting is introduced that represents a station having no data to send. The model is not predictive as this probability is not known as a function of load and must be estimated from simulation. In [7] idle states are added after packet transmission to represent bursty arrivals in a way that does not account for postbackoff, a key bandwidth saving feature of the 802.11 MAC. In [8] a Markov model where states are of fixed real-time length is introduced, but by virtue of its design it cannot predict the pre-saturation peak in throughput. In [9] a model focusing on multi-rate situations is presented, but not solved analytically and is subject to limited validation. In [10] a non-Markov model is developed, but is based on an unjustified assumption that the saturated setting provides good approximation to certain unsaturated quantities. It appears to produce inaccurate predictions. None of these previous models have gone beyond the homogeneous setting and so have not been able to consider fairness issues for competing traffic types. The p-persistent approach to modeling the 802.11 MAC has also been studied extensively, for recent work see [11] and the references therein.

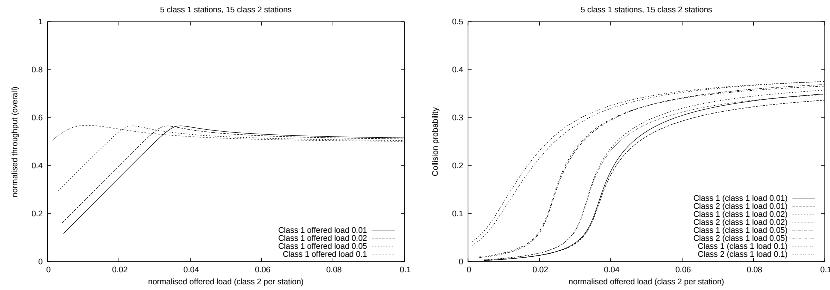
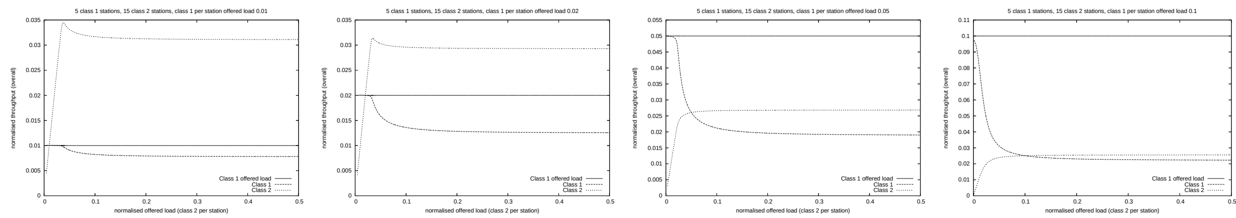


Fig. 9. Overall throughput and per station collision probabilities for two classes of stations with class 1 offering fixed per station load, $n_1 = 5$, $n_2 = 15$.



(a) Class 1 per station offered load 0.01

(b) Class 1 per station offered load 0.02

(c) Class 1 per station offered load 0.05

(d) Class 1 per station offered load 0.1

Fig. 10. Per-station throughput for two classes of stations with class 1 offering fixed per-station load, $n_1 = 5$, $n_2 = 15$.

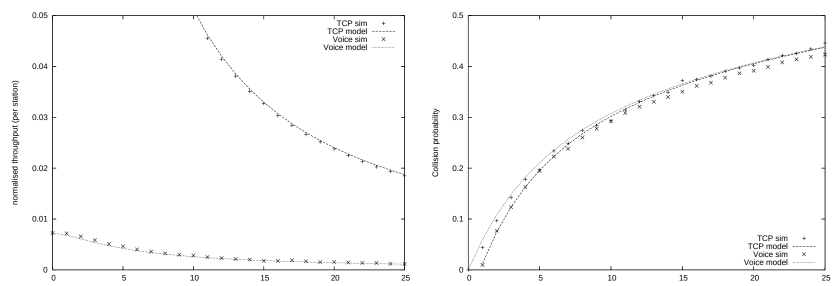


Fig. 11. VoIP and TCP.

VII. Conclusions

We have presented a model and analysis of the 802.11 MAC under non-saturated and heterogenous conditions. The model's predictions were validated against simulation and seen to accurately capture many interesting features of non-saturated operation, including predicting that peak throughput occurs prior to saturation. We address the question of fairness between competing flows showing, for example, that saturated data flows may significantly reduce the bandwidth available to low-rate VoIP flows.

References

[1] G. Bianchi, "Performance analysis of IEEE 802.11 distributed coordination function," *IEEE Journal on Selected Areas in Communications*, vol. 18, pp. 535–547, March 2000.

[2] R. Battiti and B. Li, "Supporting service differentiation with enhancements of the IEEE 802.11 MAC protocol: models and analysis," Tech. Rep. DIT-03-024, University of Trento, May 2003.

[3] J. W. Robinson and T. S. Randhawa, "Saturation throughput analysis of IEEE 802.11e enhanced distributed coordination function," *IEEE Journal on selected areas in communications*, vol. 22, pp. 917–928, June 2004.

[4] S. Wiethölter and C. Hoene, "Design and verification of an IEEE 802.11e EDCF simulation model in ns-2.26," Tech. Rep. TKN-03-019, Technische Universität Berlin, November 2003.

[5] A. Markopoulou, F. Tobagi, and M. Karam, "Assessing the quality of voice communications over internet backbones," *IEEE Transactions on Networking*, vol. 11, pp. 747–760, October 2003.

[6] G.-S. Ahn, A. T. Campbell, A. Veres, and L.-H. Sun, "Supporting service differentiation for real-time and best-effort traffic in stateless wireless ad hoc networks (SWAN)," *IEEE Transactions on Mobile Computing*, vol. 1, no. 3, pp. 192–207, 2002.

[7] M. Ergen and P. Varaiya, "Throughput analysis and admission control in IEEE 802.11a," *ACM-Kluwer MONET Special Issue on WLAN Optimization at the MAC and Network Levels*, 2004.

[8] A. Zaki and M. El-Hadidi, "Throughput analysis of IEEE 802.11 DCF under finite load traffic," in *First International Symposium on Control, Communications and Signal Processing*, pp. 535–538, 2004.

[9] G. Cantieni, Q. Ni, C. Barakat, and T. Turletti, "Performance analysis under finite load and improvements for multirate 802.11," *to appear in Elsivier Computer Communications*, 2005.

[10] O. Tickoo and B. Sikdar, "A queueing model for finite load IEEE 802.11 random access," in *IEEE International Conference on Communications*, vol. 1, pp. 175 – 179, June 2004.

[11] L. Bononi, M. Conti, and E. Gregori, "Runtime optimization of IEEE 802.11 wireless LANs performance," *IEEE Transactions on Parallel and Distributed Systems*, vol. 15, no. 1, pp. 66–80, 2004.