Mobile Networks and Applications 10, 705–716, 2005

Throughput Analysis and Admission Control for IEEE 802.11a

MUSTAFA ERGEN and PRAVIN VARAIYA
Department of Electrical Engineering and Computer Science, University of California Berkeley

Abstract. We propose a new Markov model for the distributed coordination function (DCF) of IEEE 802.11. The model incorporates carrier sense, non-saturated traffic and SNR, for both basic and RTS/CTS access mechanisms. Analysis of the model shows that the throughput first increases, and then decreases with the number of active stations, suggesting the need for an admission control mechanism.

We introduce such a mechanism, which tries to maximize the throughput while maintaining a fair allocation. The maximum achievable throughput is tracked by the mechanism as the number of active stations increases. An extensive performance analysis shows that the mechanism provides significant improvements.

Keywords: mobile networks, wireless LAN, IEEE 802.11, distributed coordination function, admission control

1. Introduction

The IEEE 802.11 protocol specifies both Medium Access Control (MAC) and Physical (PHY) layers. The 1997 standard [5] was updated in 1999 with two new physical layers, IEEE 802.11b [7] and IEEE 802.11a [6]. The standard includes a contention based and a polling based medium access protocol, called Distributed Coordination Function (DCF) and Point Coordination Function (PCF), respectively. This paper is concerned with DCF.

DCF employs a carrier sense multiple access with collision avoidance (CSMA/CA) protocol, with two phases: carrier sense and exponential backoff. Carrier sense, comprising physical carrier sense (PCS) and virtual carrier sense (VCS), ensures that the medium is idle for some time before a transmission attempt. PCS is a notification from the physical layer about any ongoing transmission; VCS sets the network allocation vector (NAV) after inspecting the duration field of each packet during a transmission from another station (STA). After receiving a positive notification from PCS and VCS, the station backs off.

DCF is modelled in various ways. The most important Markov chain model is in [1]. That model only considers saturation throughput, meaning that the stations always have packets to transmit; non-saturated conditions are not considered. Furthermore, the models of [1] and [9] do not take into account freezing of the backoff counter: According to the standard, STA stops decrementing the backoff counter if there is an ongoing transmission, and resumes after waiting for at least DIFS time.

The model in [10] approximates freezing of the backoff counter and also introduces an additional state for the carrier sense period. However, the carrier sense model in [10] is inaccurate since, if a new module is added, the number of states that are used to represent that model should be proportional to a slot time.

In this paper we propose a Markov chain model that is closer to the standard, and compare it with previous models [3]. Unlike previous models, which only consider the saturated case, we augment the chain with an additional module to consider non-saturated conditions. Our model also reflects the behavior when stations have different signal to noise ratio (SNR) levels, hence transmit at different data rates.

As traffic increases, the DCF mechanism shows an interesting behavior: The throughput first increases and then starts to decrease, indicating congestion. This suggests the need for an admission control (AC) mechanism to maintain high throughput as the offered traffic increases. Such a mechanism can be deployed within the basic service sets[1] (BSS) or between the basic service sets [5]. Inter-access point coordination is being studied in IEEE 802.11f working group. The AC mechanism introduced here requires signaling between APs and STAs. The control signals can use the unlicensed or the dedicated band, proposed to be reserved for signaling between APs.

The paper is organized as follows. Section 2 briefly reviews IEEE 802.11 DCF, with both basic and RTS/CTS access mechanisms. Section 3 presents our model in detail, and Section 4 examines the throughput according to our model. Section 5 considers different data rates. Section 6 describes the admission control mechanism and analysis its performance. Section 7 concludes the paper.

2. IEEE 802.11 distributed coordination function

The Distributed Coordination Function (DCF) is a contention based medium access scheme. It uses a carrier sense multiple access with collision avoidance (CSMA/CA) MAC protocol with binary exponential backoff. Before transmission, a station senses the wireless medium to determine if the channel is idle. There are two sensing mechanisms: physical carrier sense (PCS) and virtual carrier sense (VCS). PCS is a signal from the physical layer to the MAC layer, indicating the

[1] Basic service set is a set of stations that communicate with one another. When a BSS includes an access point (AP), the BSS is called infrastructure BSS and the traffic is sent through the AP.

detection of a signal in the channel. VCS sets the network allocation vector (NAV) according to the duration field of each received packet. In carrier sensing, if the medium is in use, the station backs off; if it is idle for a distributed inter frame space (DIFS) interval, it transmits. The station transmits immediately if the medium is detected idle for more than DIFS time, and performs backoff prior to transmitting another frame.

DCF employs a discrete-time backoff and each slot time is one discrete-time unit. The slot time size σ is fixed in the standard according to the chosen physical layer. The station selects a backoff time uniformly in the range $(0, W-1)$. W is selected according to the station's backoff level and it is between *CWmin* and *CWmax*. If the station has experienced i collisions, $W_i = 2^i(CWmin+1)$. W is incremented up to a maximum value $CWmax+1 = 2^m(CWmin+1)$. The values for IEEE 802.11a are specified in Table 2. The backoff counter decrements as long as the channel is sensed idle for an empty slot time σ, otherwise it freezes. It reactivates when channel is detected idle for more than DIFS [1,5].

An ACK mechanism guarantees transmission. After each data transmission, a station waits for an ACK. The receiving station sends an ACK immediately after a period of time called short inter-frame space (SIFS), which is shorter than DIFS. If no ACK is received within a specified ACK_Timeout, the transmission is unsuccessful and the station applies backoff.

DCF uses two different access mechanisms. With the basic access mechanism, STA reserves the channel immediately by sending a data packet and waits for ACK; if ACK is not received within ACK_Timeout, STA performs backoff. Stations other than receiver or transmitter adjust their NAV according to the duration field of the packets.

The second access mechanism, called RTS/CTS, uses Request-To-Send (RTS) and Clear-To-Send (CTS) frames to reserve the channel. STA first sends a RTS frame and waits for a CTS frame from the intended receiver. Reception of the CTS frame corresponds to a channel reservation for the RTS transmitter. Figure 1 illustrates this operation. Stations not involved in transmission or reception update their NAV according to the duration values specified in RTS, CTS, Data and ACK.

RTS/CTS improves system efficiency. It solves the hidden terminal problem. As RTS and CTS frames are very short, time wasted in a collision during the contention period is small compared to the basic access method, in which the time for a data packet is wasted.

3. Throughput analysis

The discrete Markov chain model proposed here is similar those in [1,9,10]. We consider both the saturated case, in which stations always have packets to transmit, and the non-saturated case, in which stations transmit with probability λ. The derivation of the throughput is carried out in two steps. We first obtain τ—*the probability that the station transmits a packet in a given slot time*. We then use τ to calculate throughput. We first obtain τ for the saturated case and then modify the result for the normal or non-saturated case.

3.1. Saturated case formulation

Observing the medium in figure 1 leads us to describe its continuous-time operation in terms of discrete (virtual) slots or epochs. There is either an empty slot (of duration σ) during which the backoff counter is decremented, or a transmission (of duration equal to a data or RTS/CTS packet) during which backoff is frozen. So the events that cause a state transition in the Markov chain are an empty slot or a transmission. Hence

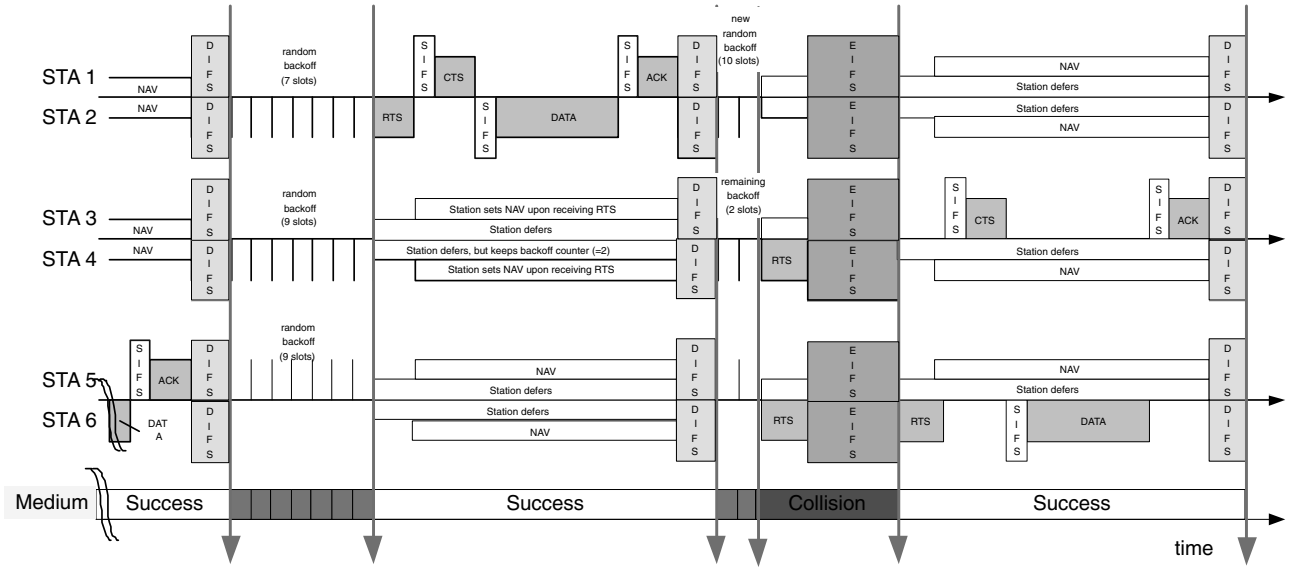


Figure 1. DCF operation in IEEE 802.11 with RTS/CTS.

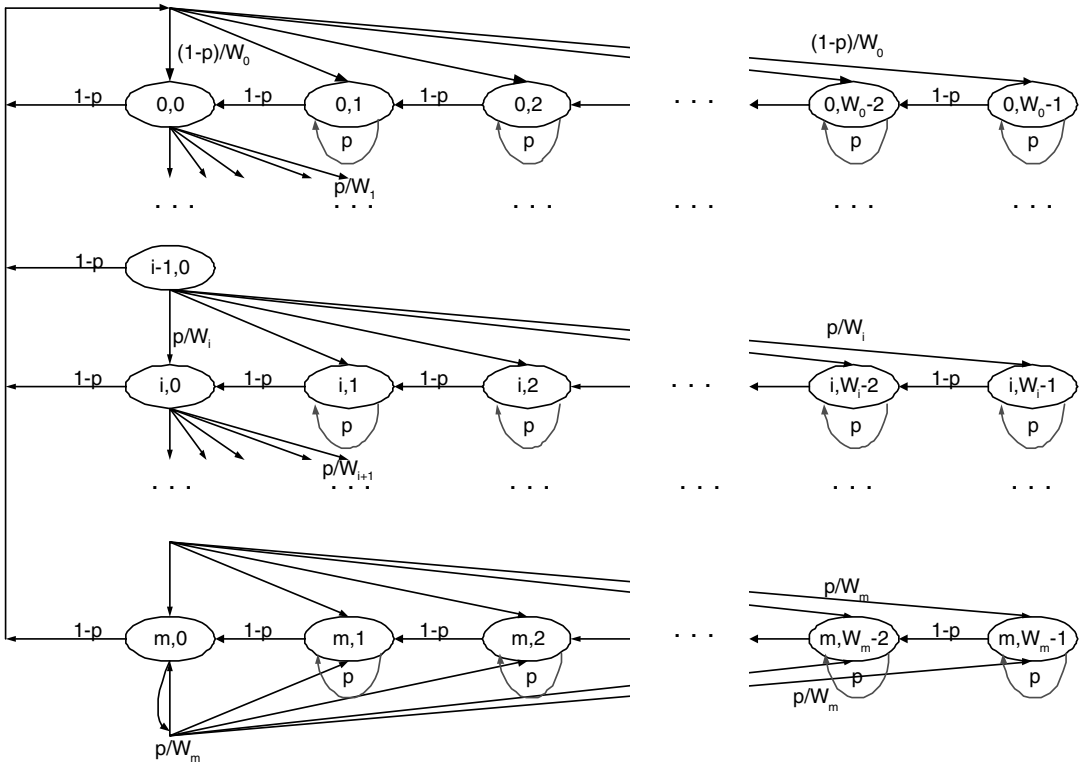


Figure 2. Markov Chain model for IEEE 802.11 DCF.

the Markov chain shown in figure 2 has two dimensions, $s(t)$ and $b(t)$, representing state number and backoff stage, respectively.

We adopt the key approximation in [1]. We represent by p the *conditional collision probability* and we assume that it is independent and constant, regardless of the number of retransmissions experienced. p also stands for *probability of detecting the channel busy*. During the deference period, if a station detects the channel busy, it resets its timer, since the condition is to detect the channel idle at least for one DIFS period.

A slot in our model is either an empty slot or a transmission. This is different from [1], in which a slot is either an empty slot or a transmission + empty slot. (Thus every slot in [1] includes an empty slot.) For this reason, our Markov model (figure 2) has the self loop in the backoff stages to model freezing of the backoff counter when the medium is busy.

$$\begin{cases} P\{i,k|i,k+1\} = (1-p) & k \in [0, W_i - 2] \quad i \in [0,m] \\ P\{i,k|i,k\} = p & k \in [1, W_i - 1] \quad i \in [0,m] \\ P\{i,k|i-1,0\} = p/W_i & k \in [0, W_i - 1] \quad i \in [1,m] \\ P\{m,k|m,0\} = p/W_m & k \in [0, W_m - 1] \end{cases} \tag{1}$$

The two-dimensional chain $(s(t), b(t))$ is governed by the one-step transition probabilities (1). The first and second equations respectively indicate that at the beginning of each slot, the backoff counter is decremented if the channel is sensed idle (which happens with probability with $(1-p)$) and frozen if the channel is sensed busy (which happens with probability p).

The third and fourth equations respectively indicate that following an unsuccessful transmission, the station in backoff stage $(i-1)$ selects a backoff interval uniformly in the range $(0, W_i - 1)$ and when the backoff stage reaches m, W_m stays constant.

We can solve the balance equations to obtain the stationary distribution denoted by $b_{i,k}, i \in [0,m], k \in [0, W_i - 1]$.

$$\begin{aligned} b_{i-1,0} \cdot p = b_{i,0} \rightarrow b_{i,0} = p^i b_{0,0}, \quad 0 < i < m \\ b_{m-1,0} \cdot p = (1-p) b_{m,0} \rightarrow b_{m,0} = \frac{p^m}{1-p} b_{0,0} \end{aligned} \tag{2}$$

From (2), the stationary distribution is

$$b_{i,k} = \frac{W_i - k}{W_i(1-p)} \cdot \begin{cases} (1-p) \cdot \sum_{j=0}^{m} b_{j,0} & i = 0 \\ p \cdot b_{i-1,0} & 0 < i < m \\ p \cdot (b_{m-1,0} + b_{m,0}) & i = m \end{cases} \tag{3}$$

or

$$b_{i,k} = \frac{W_i - k}{W_i(1-p)} b_{i,0} \quad i \in [0,m], \ k \in [0, W_i - 1]. \tag{4}$$

$$1 = \sum_{i=0}^{m} \sum_{k=0}^{W_i - 1} b_{i,k} = backoff \tag{5}$$

All $b_{i,k}$ can be expressed in terms of $b_{0,0}$, which can then be obtained because all probabilities add to one (6). This finally

yields $b_{0,0}$ in terms of p, W, m in equation (7).

$$backoff = \sum_{i=0}^{m} \sum_{k=0}^{W_i - 1} b_{i,k}$$

$$= \frac{b_{0,0}}{(1-p)} \left(\sum_{i=0}^{m} p^i \left(\frac{W_i + 1}{2} \right) + \frac{p^{m+1}}{(1-p)} \left(\frac{W_m + 1}{2} \right) \right), \quad (6)$$

$$b_{0,0} = \frac{1}{\frac{1}{(1-p)} \sum_{i=0}^{m} p^i \left(\frac{W_i+1}{2} \right) + \frac{p^{m+1}}{(1-p)^2} \left(\frac{W_m+1}{2} \right)}. \quad (7)$$

A packet is transmitted in states $b_{i,0}, i \in [0, m]$, so τ, the probability of transmission in a plot, is given by (8),

$$\tau = \sum_{i=0}^{m} b_{i,0} = \frac{b_{0,0}}{(1-p)}$$

$$= \frac{1}{\sum_{i=0}^{m} p^i \left(\frac{W_i+1}{2} \right) + \frac{p^{m+1}}{(1-p)} \left(\frac{W_m+1}{2} \right)}. \quad (8)$$

Taking the contention window $CW_{\max} = 2^m CW_{\min}$, so $W_i = 2^i W, i \in [0, m]$, and $W = CW_{\min}$, gives a simpler expression for τ,

$$\tau = \frac{1}{\frac{(1-2p)(W+1) + pW(1-(2p)^m)}{2(1-2p)(1-p)}}. \quad (9)$$

For purposes of comparison, the transmission probability $\tau_{[1]}$ of [1] is

$$\tau_{[1]} = \frac{\tau}{1-p}. \quad (10)$$

3.2. Normal, unsaturated case formulation

To model normal, non-saturated conditions, we introduce additional states, giving the chain of figure 3.

Take $W_{-2} = 2$, so we introduce only two states. The one-step transition probabilities are slightly changed:

$$\begin{cases} P\{-2, W_{-2} - 1 | -2, W_{-2} - 1\} = (1 - \lambda) \\ P\{0, j | -2, 0\} = \lambda(1-p)/W_0 \quad j \in [0, W_0 - 1] \\ P\{-2, 1 | -2, 0\} = (1 - \lambda) \\ P\{-2, 0 | -2, 1\} = \lambda \\ P\{-2, 0 | i, 0\} = (1 - p) \quad i \in [0, m] \end{cases} \quad (11)$$

Under non-saturated conditions, a station may now wait in the idle state for a packet from upper layers. This corresponds to a delay in the idle state, represented by the box in figure 3. The delay in the idle state is geometric with parameter λ. The transition probabilities in (11) are straightforward modifications of those previously obtained for the saturated case.

The stationary probabilities add up to 1,

$$1 = \sum_{i=0}^{m} \sum_{k=0}^{W_i - 1} b_{i,k} + \sum_{k=0}^{W_{-2} - 1} b_{-2,k} = backoff + idle. \quad (12)$$

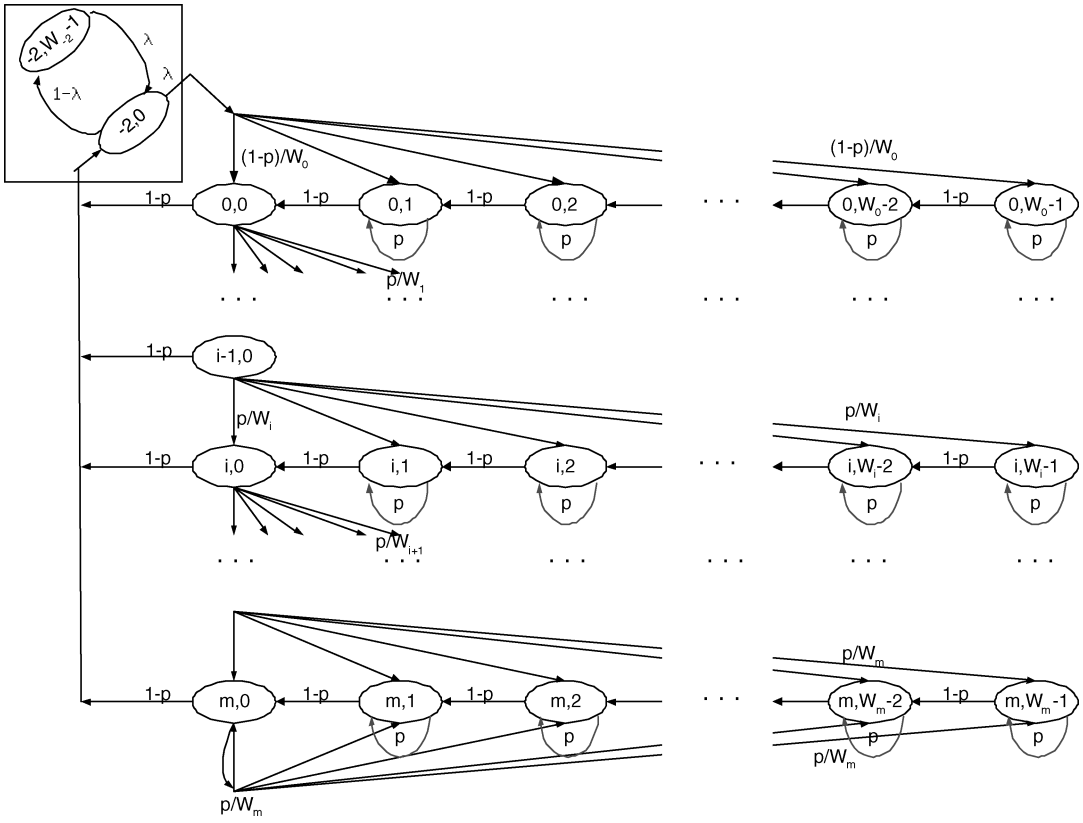


Figure 3. Markov Chain model for the IEEE 802.11 DCF model in normal operating condition.

The probabilities $b_{-2,0}$ and $b_{-2,1}$ can be expressed in terms of $b_{0,0}$ using (2), (11), and by representing the probability *idle* in terms of $b_{0,0}$ by

$$idle = \sum_{k=0}^{W_{-2}-1} b_{-2,k} - 1 = \frac{b_{0,0}}{\lambda^2} - 1.$$

The new τ is given by (13), which reduces to (9) for the saturated case ($\lambda = 1$),

$$\tau = \frac{1}{\frac{(1-2p)(W+1)+pW(1-(2p)^m)}{2(1-2p)(1-p)} + (1-p)(\frac{1}{\lambda^2}-1)}. \quad (13)$$

From (13) we see that $\tau = \tau(p, m, W, \lambda)$ depends on the unknown p. Also, as in [1],

$$p = 1-(1-\tau)^{n-1} \quad \text{or} \quad \tau(p) = 1-(1-p)^{\frac{1}{(n-1)}}. \quad (14)$$

Equations (13) and (14) together determine τ and p. Figure 4 plots the collision probability p and transmission probability τ as the number of stations varies, for five cases: the model in [1], and the proposed model for four different $\lambda = 1, 0.2, 0.1, 0.05$. The model in [1] gives higher values of p and τ than our model for $\lambda = 1$. In general, as expected, p increases and τ decreases with n.

Also, as expected, τ increases with load λ, which is readily appreciated by taking $n \to 1$, $p \to 0$, for which

$$\lim_{p\to 0} \tau = \frac{1}{\frac{(W+1)}{2} + \left(\frac{1}{\lambda^2}-1\right)}. \quad (15)$$

For the saturated case, $\lambda = 1$, and $m = 0$ (no exponential backoff), we can compare τ with $\tau_{[1]}$ in (10),

$$\tau(p, 0, W, 1) = \frac{2(1-p)}{W+1} < \tau_{[1]} = \frac{2}{W+1}. \quad (16)$$

Unlike in [1], τ depends on the collision probability p (and hence on n). Intuitively of course, τ should depend on n: if there are more stations, the medium will be busy more often, and a station will transmit less frequently.

3.3. Throughput analysis

As in [1], "throughput is the fraction of time the channel is used to successfully transmit payload bits." Define P_{tr} as the probability that there is at least one transmission in a slot, and P_s as the probability that a transmission is successful, so

$$P_{tr} = 1-(1-\tau)^n,$$
$$P_s = n\tau(1-\tau)^{n-1}. \quad (17)$$

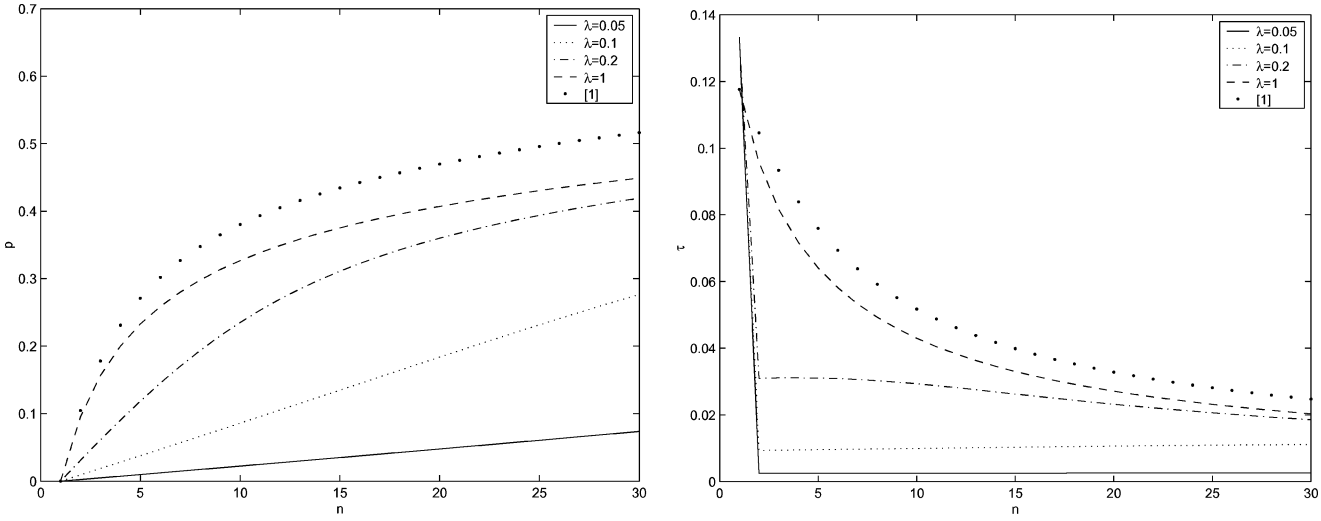
The throughput S is the ratio

$$S = \frac{P_s E[P]}{(1-P_{tr})\sigma + P_s T_s + P_{tr}(P_{tr}-P_s)T_c}, \quad (18)$$

in which $E[P]$ is the average packet payload size. The denominator is the average duration of a slot, which may be an idle slot (of duration σ), a successful transmission (of duration T_s), or a collision (of duration T_c). These durations for the basic and RTS/CTS mechanisms are given below:

$$\begin{aligned} T_s^{\text{basic}} &= T_{DATA} + tSIFSTime + \delta + T_{ACK} + \delta + tDIFSTime \\ T_c^{\text{basic}} &= T^*_{DATA} + \delta + tEIFSTime \\ T_s^{rts} &= T_{RTS} + tSIFSTime + \delta + T_{CTS} + tSIFSTime + \delta \\ &\quad + T_{DATA} + tSIFSTime + \delta + T_{ACK} + \delta + tDIFSTime \\ T_c^{rts} &= T_{RTS} + \delta + tEIFSTime. \end{aligned} \quad (19)$$

In (19), T_{DATA} is the duration of a packet of size $E[P]$ and T_{RTS}, T_{CTS}, T_{ACK} are the durations of the corresponding frames. T^*_{DATA} is the average time to send $E[P^*]$ bytes, which is the average length of the longest packet payload involved in a collision. When all packets have the same size, $E[P] = P = E[P^*]$. δ is the propagation delay. Unlike in basic access



(a) Probability of collision (b) Probability of transmission

Figure 4. p and τ versus n.

Table 1
Eight PHY Modes of the IEEE 802.11a PHY.

Mode	Modulation	Code rate	Data rate	BpS	SNR
1	BPSK	1/2	6 Mbps	3	25
2	BPSK	3/4	9 Mbps	4.5	27
3	QPSK	1/2	12 Mbps	6	30
4	QPSK	3/4	18 Mbps	9	32
5	16-QAM	1/2	24 Mbps	12	35
6	16-QAM	3/4	36 Mbps	18	40
7	64-QAM	2/3	48 Mbps	24	42
8	64-QAM	3/4	54 Mbps	27	45

T_c^{rts} only contains T_{RTS} since a collision can only occur in the RTS frame transmission.

3.4. IEEE 802.11a OFDM physical layer

The IEEE 802.11a PHY uses OFDM modulation, and provides eight modes with different modulation schemes and coding rates. Table 1 shows the supported rates depending on SNR. (SNR values are vendor-proprietary.)

As shown in figure 5, each MAC data frame or MAC Protocol Data Unit (MPDU), consists of the MAC Header, Frame Body, and Frame Check Sequence (FCS). The MAC header and FCS together are 28 octets, the RTS frame is 18 octets, and the CTS and ACK are 12 octets long.

When a MPDU is passed to the PLCP layer it is called PSDU. In order to create a PLCP Protocol Data Unit (PPDU), PLCP headers are added. Figure 6 shows the PPDU format. During transmission, a PLCP preamble and a PLCP header are added to a PSDU to create a PLCP Protocol Data Unit (PPDU). The PLCP preamble field, with the duration of *tPLCPPreamble*, is composed of 10 repetitions of a short training sequence (0.8 μs) and two repetitions of a long training sequence (4 μs). The PLCP header except the SERVICE field, with the duration of *tPLCP_SIG* constitutes a single OFDM symbol. Each OFDM symbol interval is denoted by *tSymbol* and its duration is 4 μs. The 16-bit SERVICE field of the PLCP header and the MPDU (along with six tail bits and pad bits), represented by DATA, are transmitted at the data rate specified in the RATE field.

The BSS basic rate set is {6 Mbps, 12 Mbps, 24 Mbps} and each station should support these rates and control information should be sent at these rates. We assume that all rates specified in Table 1 are supported and control signalling can be in any rate.

We can thus obtain the duration of each packet. The time to transmit a frame with $E[P]$ octets of data payload with the IEEE 802.11a PHY 1 is given below:

$$\begin{aligned} T_{DATA}(m) &= tPLCPPreamble + tPLCPHeader \\ &\quad + MACHeader + E[P] + FCS + Tailbits + PadBits \\ &= 20\ \mu s + \left\lceil \frac{28 + (16+6)/8 + E[P]}{BpS(m)} \right\rceil \cdot 4\ \mu s \\ T_{RTS}(m) &= tPLCPPreamble + tPLCPHeader \\ &\quad + MACHeader + FCS \\ &= 20\ \mu s + \left\lceil \frac{20 + (16+6)/8}{BPS(m)} \right\rceil \cdot 4\ \mu s \end{aligned}$$

DATA

Frame Control	Duration/ ID	Address1	Address2	Address3	Sequence Control	Address4	Frame Body	FCS
Octets:2	2	6	6	6	2	6	0-2312	4

RTS

Frame Control	Duration	RA	TA	FCS
Octets:2	2	6	6	4

CTS & ACK

Frame Control	Duration	RA	FCS
Octets:2	2	6	4

Figure 5. Frame formats.

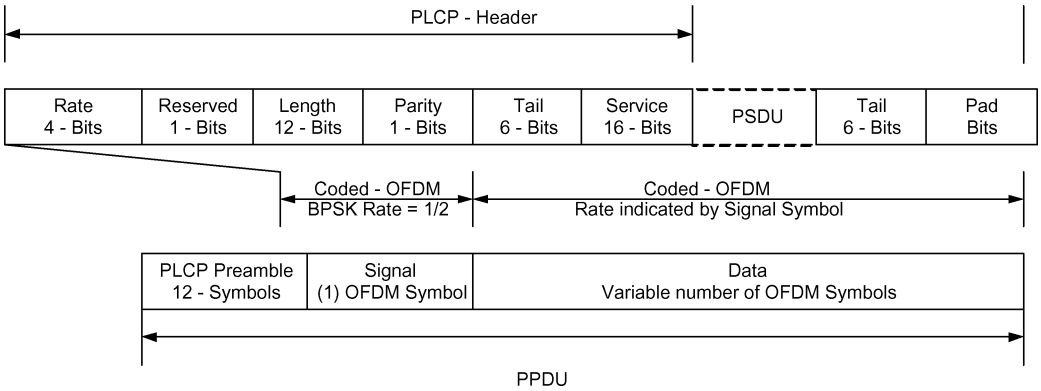


Figure 6. IEEE 802.11a OFDM Packet.

$$T_{CTS}(m) = T_{ACK}(m) = tPLCPPreamble + tPLCPHeader + MACHeader + FCS = 20\ \mu s + \left\lceil \frac{14 + (16+6)/8}{BPS(m)} \right\rceil \cdot 4\ \mu s.$$

The Bytes-per-Symbol for PHY mode m is $BpS(m)$ in Table 1. The BpS value in our example depends on SNR.

4. Throughput characteristics

The performance results below are based on the physical layer parameters specified in Tables 1 and 2. The payload is constant, $E[P] = 1024$ bytes.

Figure 7(a) shows the throughput for the basic access scheme for the model of [1] and our model for four different values of λ. The throughput first increases with the number of stations until congestion sets in, after which throughput decreases. As the traffic intensity decreases, the maximum throughput is reached with a larger number of active stations.

Figure 7(b) reports the throughput for RTS/CTS access mechanism. The effect of congestion is now less severe. However, as with basic access, the throughput decreases and goes to 0, as $n \to \infty$.

Figure 8(a) shows the effect of different SNR levels (all stations have the same SNR level). Figure 8(b) shows the effect of traffic intensity on throughput for a fixed number of stations. We again observe onset of congestion.

Table 2
IEEE 802.11a OFDM PHY characteristics.

Characteristics	Value	Definition
tSlotTime	9 μs	Slot time
tSIFSTime	16 μs	SIFS time
tDIFSTime	34 μs	DIFS = SIFS + 2 × Slot
aCWmin	15	min CW
aCWmax	1023	max CW
tPLCPPreamble	16 μs	PLCP preamble duration
tPLCP_SIG	4 μs	PLCP SIGNAL field duration
tSymbol	4 μs	OFDM symbol interval

5. Formulation for different data rates

We evaluate the throughput when different stations have different SNR ratios, hence different data rates. The protocol gives each station the same chance to transmit, and different data rates only affect the slot duration. Suppose there are n stations, D different data rates, $R^1 < \cdots < R^D$, and n^i stations have rate R^i with corresponding slot durations T_s^i and T_c^i. The average slot durations are given by (20), (21), and the throughput of a station is given by (22). Note that the throughput S is the *same* for all stations [4].

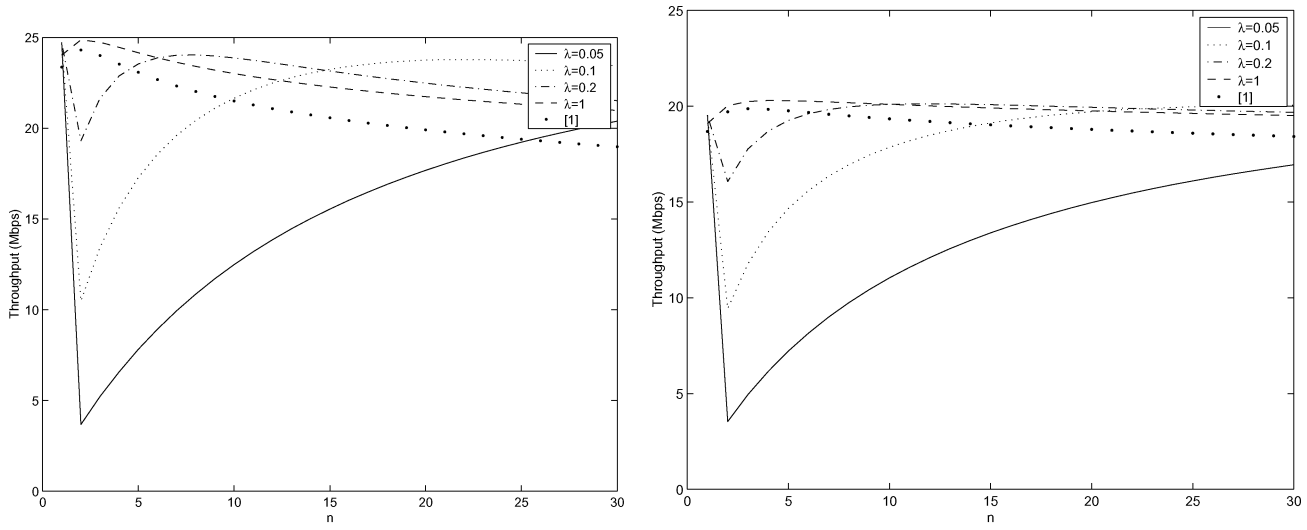
$$\bar{T}_s = \frac{P_s}{n} \sum_{i=1}^{D} n^i T_s^i \tag{20}$$

$$\bar{T}_c = \sum_{i=1}^{n-1} \sum_{j=1}^{D} \sum_{k=1}^{n^j} \binom{n - k - \sum_{l=1}^{j-1} n^l}{i} \times T_c^j \tau^{i+1} (1-\tau)^{n-1-i} \tag{21}$$

$$S = \frac{1}{n} \frac{P_s E[P]}{(1 - P_{tr})\sigma + \bar{T}_s + \bar{T}_c} \tag{22}$$

6. Admission control

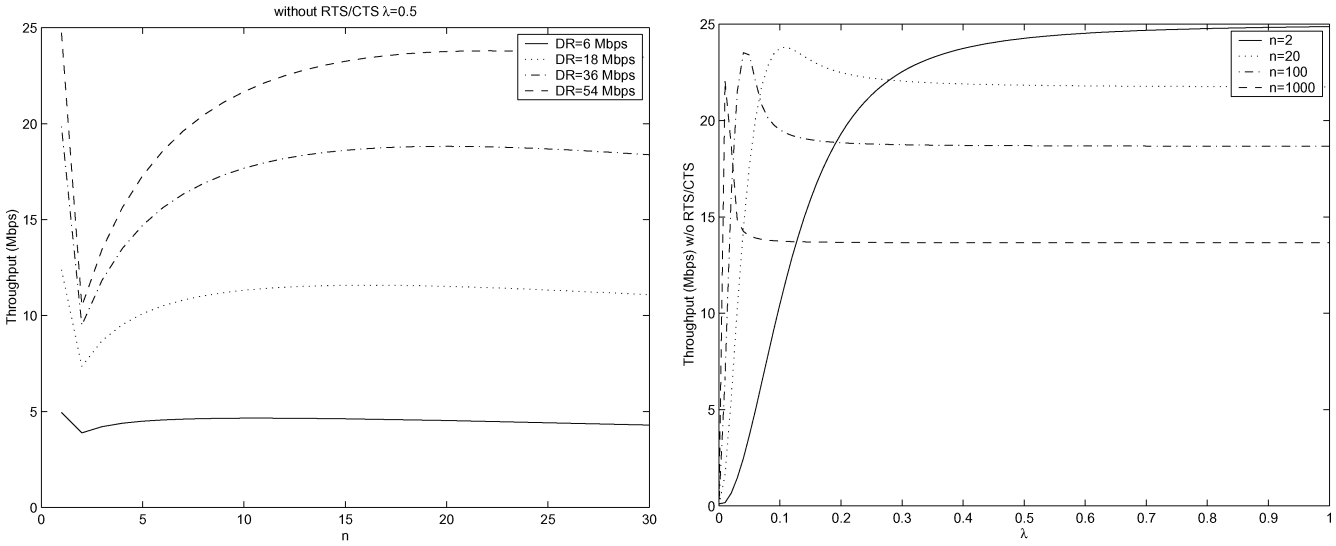
As we have seen, the overall throughput at first increases and later decreases with the number of stations. It also depends on the individual data rates. Thus there is a need for an admission



(a) Data Rate=54Mbps w/o RTS/CTS

(b) Data Rate=54Mbps with RTS/CTS

Figure 7. Throughput versus number of active nodes.



(a) Different Data Rate w/o RTS/CTS

(b) Data Rate=54Mbps w/o RTS/CTS when n fixed

Figure 8. Throughput versus number of active nodes.

control (AC) mechanism that restricts access in order to maintain high system throughput. We introduce one AC mechanism.

We consider AC within the BSS using DCF, so we formulate our problem for a network with one access point (AP) connected to several stations. The goal is to maintain maximum throughput as the number of active stations increases. In inter-BSS admission control, there is more than one AP and in addition to restricting access, the mechanism may assign stations to different APs.

6.1. Intra BSS admission control

The goal is to maximize system throughput, while maintaining fairness in the sense of equalizing the chance each station has to transmit. As suggested in figure 9, the algorithm makes two decisions. It first determines the number of active stations in each period to maximize throughput; it then selects the stations to achieve fairness.

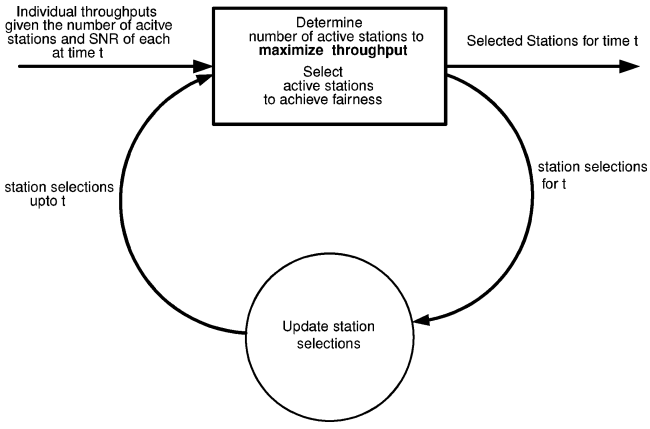


Figure 9. Control algorithm.

We use the following notation to describe the algorithm. The time interval is divided into periods indexed $t \in [1, \mathit{Endtime}]$. For each station $i \in [1, N]$, $x_i(t) = 1$ or 0, accordingly as station i is or is not selected in period t, and

$$x_i^{\text{Total}}(t) = \sum_{s \le t} x_i(s).$$

For $I \subset \{1, \ldots, N\}$, let $S(I)$ be the system throughput if subset I of stations is selected. Given the data rates of every station, $S(I)$ is given by (22).

At each t, the mechanism activates the subset $I(t)$ of stations that solves the following optimization problem:

$$\begin{aligned} \max_{I(t)} \quad & S(I(t)) - KC(t) \\ \text{s. t.} \quad & C(t) = \max_i x_i^{\text{Total}}(t) - \min_i x_i^{\text{Total}}(t) \end{aligned} \tag{23}$$

Here $K > 0$ is a constant. $C(t)$ is the inequality among stations at time t, so the objective function strikes a balance between throughput and fairness, depending on K.

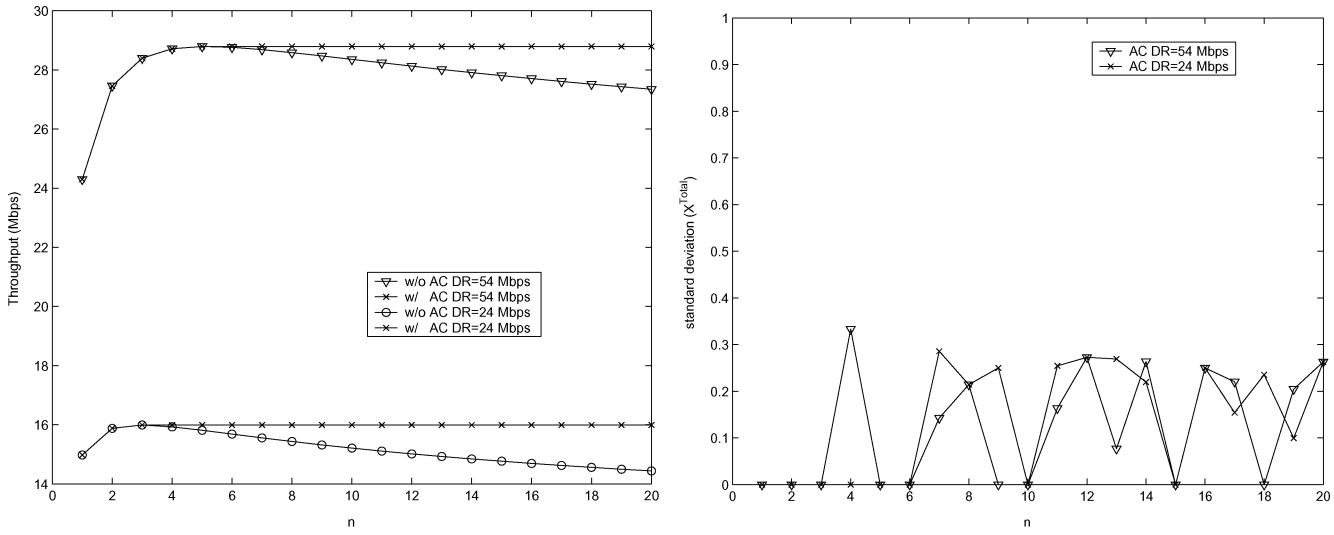
6.2. Performance results

The performance of the algorithm is evaluated in three scenarios, for basic access and with $\lambda = 0.2$: (1). All stations have the same data rate; (2) Stations have different (but fixed) data rates; and (3) Station data rates change with time, suggesting that they are moving.

6.2.1. Same data rate

Figure 10(a) indicates that the algorithm correctly determines the number of active stations that maximizes throughput. As N increases, the number of active stations remains constant. This constant depends on the data rate.

Figure 10(b) shows that the stations are selected to ensure fairness. It plots the standard deviation of the samples



(a) Throughput (b) Fairness Metric

Figure 10. Admission control when SNR is constant.

$\{x_1^{\text{Total}}, \ldots, x_N^{\text{Total}}\}$. Since $Endtime = 20$, the standard deviation could be as large as 10, but the algorithm keeps it well below 1.

6.2.2. *Different data rates*

When stations have different SNR values, the throughput depends not only on the number of active stations, but on their data rates. From (22) we know that if we wish to maximize the system throughput, only stations with large SNR would be selected. Thus in this case, as K increases, maintaining fairness occurs at the cost of reduced system throughput.

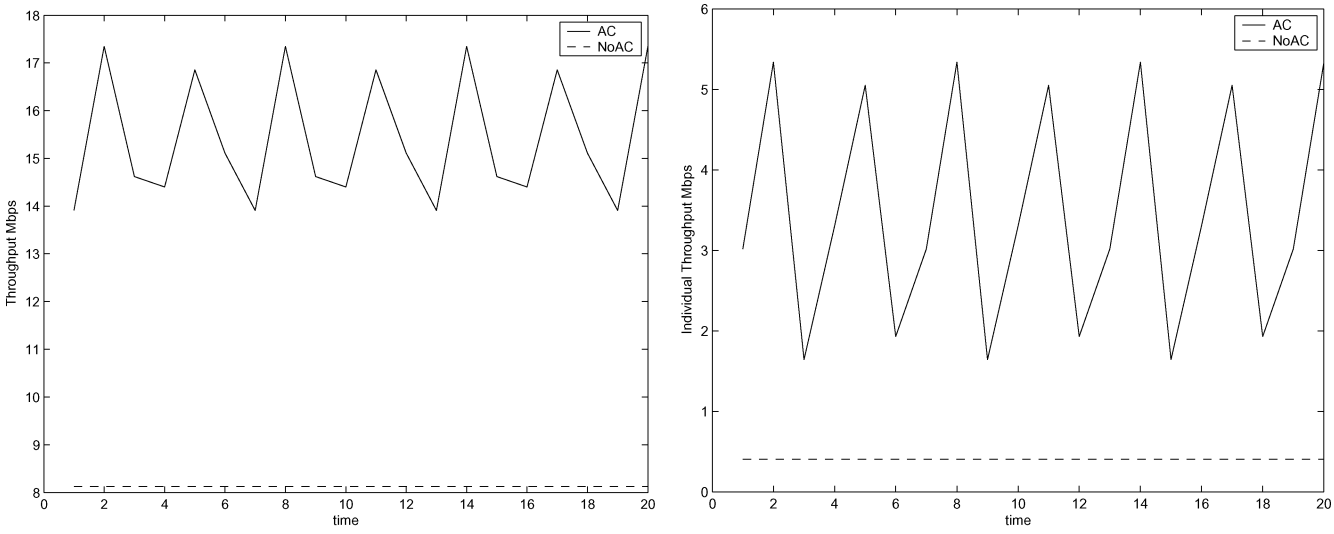
Figures 11(a) and (b) show the algorithm triples individual throughput and doubles total throughput compared with the situation with no admission control.

As stations have different data rates, the number of selected stations varies significantly as seen in figure 12(b). The mechanism follows a pattern and movement is first to higher throughput and then to fairness. Figure 12(a) shows that fairness is achieved among station since they all selected with equal probability.

In figure 13(a), the fairness metric C is around 1 meaning that the stations are selected almost equally. Figure 13(b) represents the throughput distribution for different data rates. As can be seen, as the data rate increases, the throughput also increases.

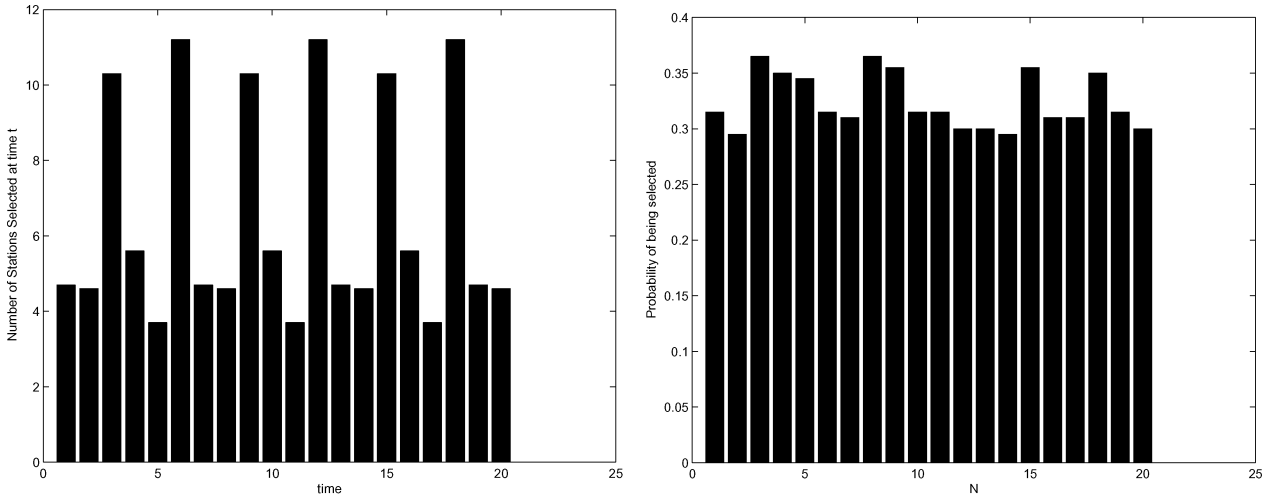
6.2.3. *Mobile stations*

When the stations move, the SNR of each station changes with time. The movement model selects a random data rate set each



(a) Throughput (b) Individual Throughput

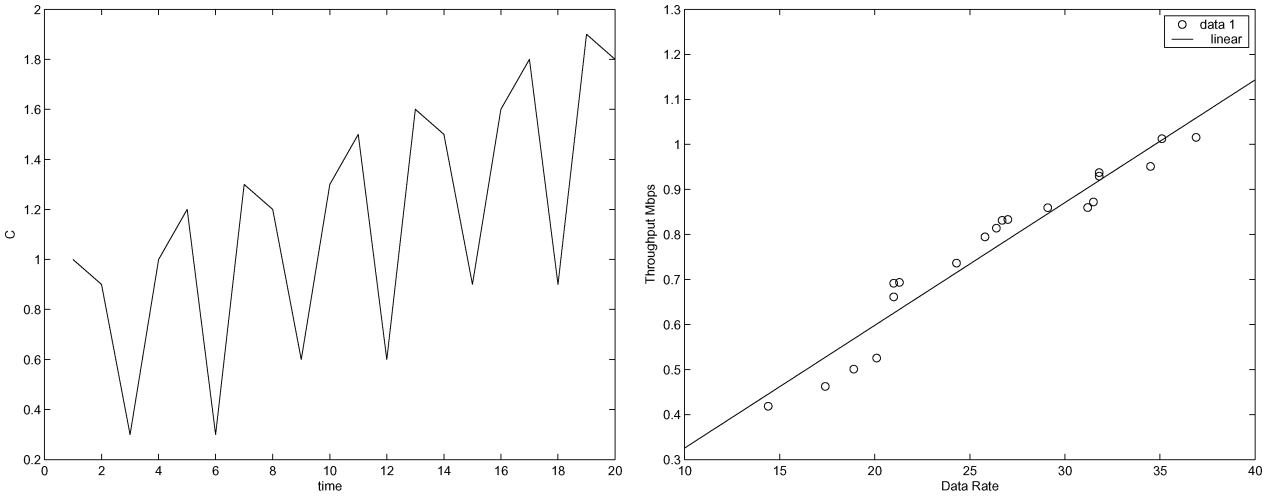
Figure 11. Admission control when there is no mobility.



(a) Number of Active Stations over time

(b) Probability of being selected at time t

Figure 12. Admission control when there is no mobility.



(a) Fairness Constraint

(b) Data Rate vs Throughput

Figure 13. Admission control when there is no mobility.

time. Hence, previous data rates do not impose any constraint in future selections.

The difference between figures 11(a) and (b) and figures 14(a) and (b) is due to the fact that when there is no mobility the initial SNR values shape the selection and impose strong constraints that can not be disregarded by the AC mechanism. For this reason, when there is no mobility the shape of the curve is almost periodic.

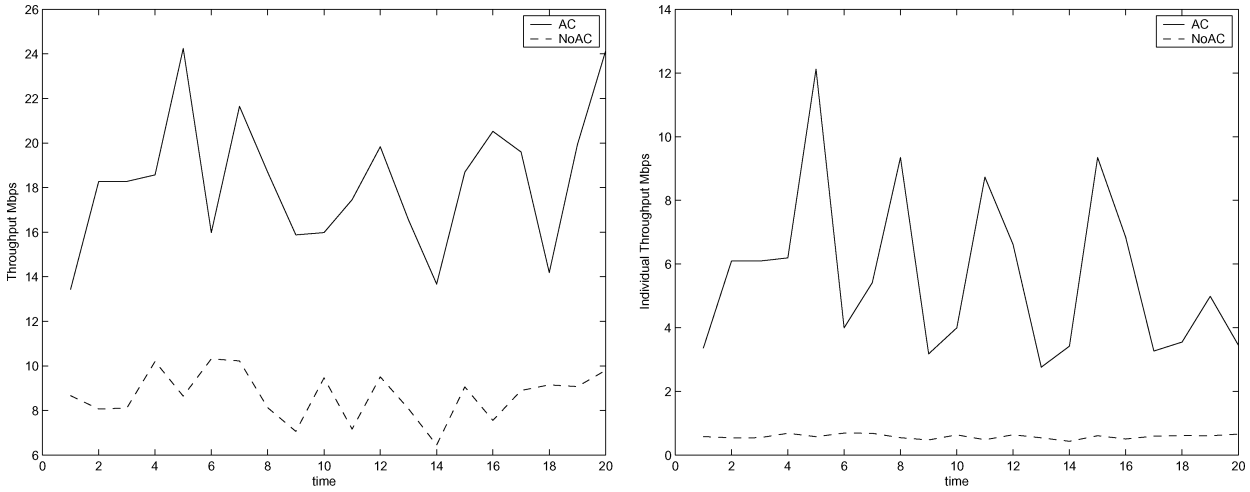
Figures 15(a) and (b) show consistent behavior. The average number of stations selected at time t is not smooth as before because it depends on the data rate vector of the time. It is important to note that probability of being selected for each station is almost equal.

Figure 16(a) shows that fairness constraint is around one and individual throughput allocation increases as the increase in the data rate. The points in figure 16(b) are interpolated by a cubic function and one can compare the plot with the no mobility case and infer that stations with higher data rate are favored as the time passes, since the data rate changes all the time in the mobility scenario, throughput allocation is not linear with data rate.

6.3. Implementation issues

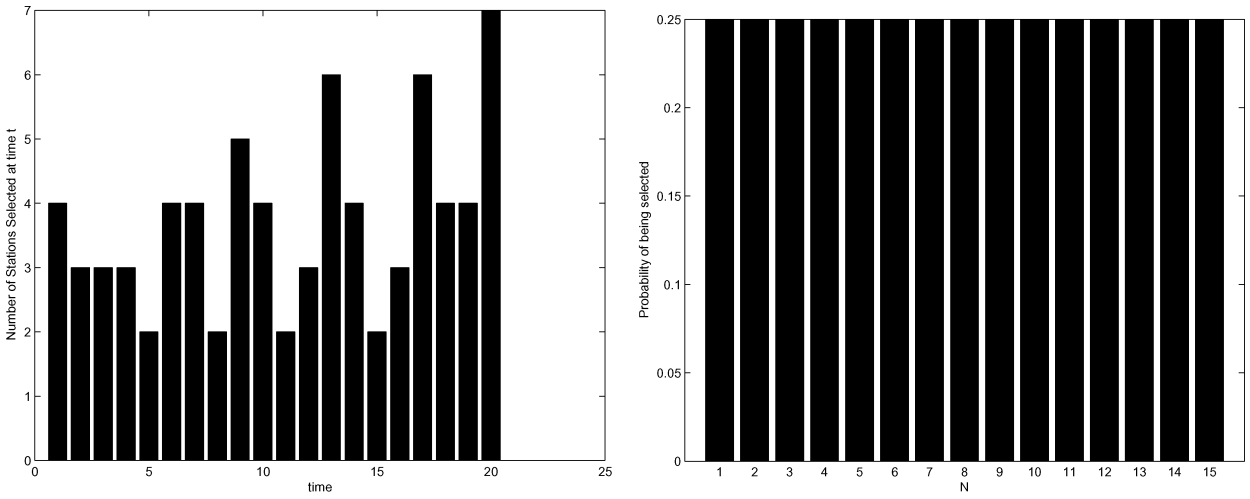
The optimization problem (23) requires selection of one of 2^N subsets I, so a large data base is needed to store all the values of $S(I)$. A more scalable method might use clustering or on-demand scheduling.

The proposed AC mechanism is a centralized approach, but DCF is designed for decentralized networks. In infrastructure BSS, the access point is where the AC mechanism can be implemented. The access point monitors the channel and notifies the stations through the same channel or a dedicated channel. A dedicated channel is being proposed for inter-AP communication.



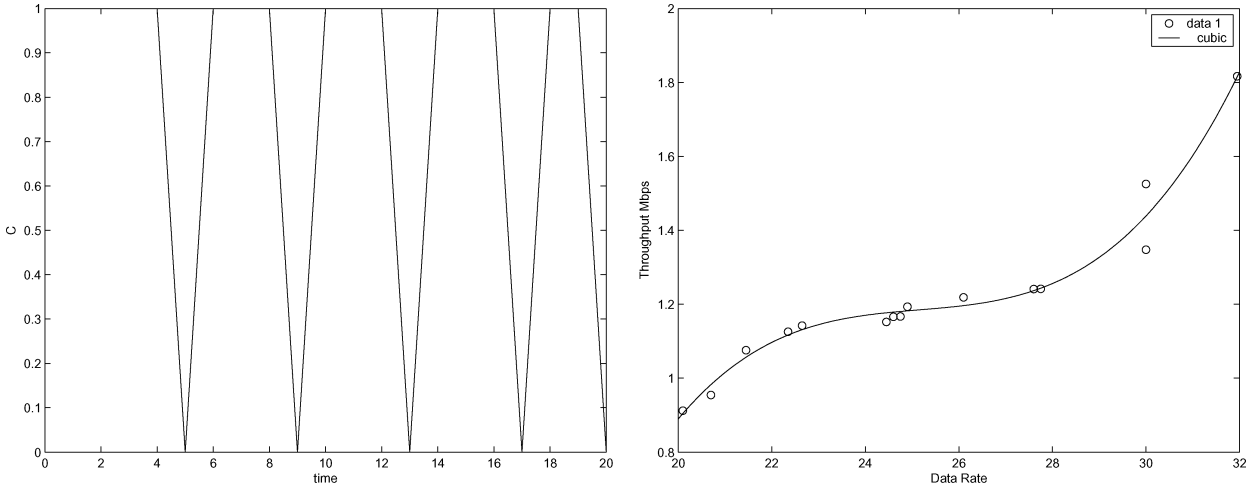
(a) Throughput (b) Individual Throughput

Figure 14. Admission control when there is mobility.



(a) Number of Active Stations over time (b) Probability of being selected at time t

Figure 15. Admission control when there is mobility.



(a) Fairness Constraint (b) Throughput vs Data Rate

Figure 16. Admission control when there is mobility.

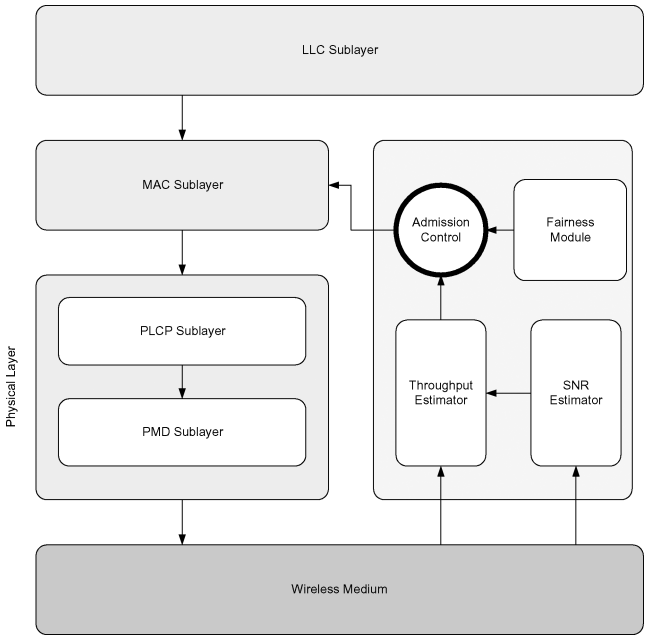


Figure 17. System architecture.

If the AP does not belong to the same ESS, the communication can be carried in a dedicated channel without any interference. This method could be extended to ad hoc networks, in which a station is selected to monitor and control the network in terms of only selecting the stations.

The system architecture in figure 17 introduces three modules. The SNR estimator estimates the SNR of each station. The throughput estimator estimates the total throughput and individual throughput. The selection mechanism uses the throughput estimator and the fairness module, which keeps each station's history.

7. Conclusion

We introduced a novel Markov chain model for IEEE 802.11, which includes three features: carrier sense, backoff freezing, and unsaturated traffic. The model is analyzed to obtain throughput. The throughput increases as the number of stations increases until the network gets congested, after which throughput starts to decrease.

We introduced an admission control mechanism that avoids congestion and maintains the system at its highest achievable throughput level, which maintaining a fair resource distribution.

References

[1] G. Bianchi, Performance analysis of the IEEE 802.11 distributed coordination function IEEE Std 802.16-2001, Vol., 2002 Pages: 0_1-322 URL: IEEE Journal on Selected Areas in Communications 18(3), (March 2000).

[2] M. Ergen, IEEE 802.11 Tutorial http://www.eecs.berkeley.edu/∼ergen/docs/ieee.pdf

[3] M. Ergen and P. Varaiya, Understanding of analytical markov model for throughput analysis in distribution coordination function of IEEE 802.11, preprint.

[4] M. Ergen and P. Varaiya, Throughput formulation and WLAN optimization in mixed data rates for IEEE 802.11 DCF Mode. GLOBECOM-CAMAD 2004.

[5] IEEE 802.11, Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specifications, Standard, IEEE, Aug. 1999.

[6] IEEE 802.11a, Part 11: Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specifications: High-Speed Physical Layer Extension in the 5 GHz Band, supplement to IEEE 802.11 Standard, Sept. 1999.

[7] Part 11: Wireless LAN, Medium Access Control (MAC) and Physical Layer (PHY) Specifications: High-Speed Physical Layer Extension in the 2.4 GHz Band, supplement to IEEE 802.11 Standard, Sept. 1999.

[8] D. Qiao, S. Choi and K.G. Shin, Goodput analysis and link adaptation for IEEE 802.11a wireless LANs IEEE Transactions on Mobile Computing 1(4) (October-December 2002).

[9] H. Wu, Y. Peng, K. Long, S. Cheng and J. Ma, Performance of reliable transport protocol over IEEE 802.11 wireless LAN: Analysis and Enhancement.

[10] Y. Xiao, A simple and effective priority scheme for IEEE 802.11, IEEE Communication Letters 7(2) (February 2003).

Mustafa Ergen received the B.S. degree in electrical engineering from Middle East Technical University (METU) and was the METU Valedictorian in 2000. He received the M.S. and Ph.D. degrees in electrical engineering in 2002 and 2004, the MOT certificate of HAAS Business School in 2003, and the M.A. degree in International and Area Studies in 2004 from the University of California, Berkeley.

Dr. Ergen has been conducting research in wireless communication networks with an emphasis on sensor networks, wireless LAN and OFDM systems and is the author of many works in the field, including the book (with A.R.S. Bahai and B.R. Saltzberg) Multi-Carrier Digital Communications: Theory and Applications of OFDM (New York: Springer, 2004).

He is National Semiconductor Post Doctoral Fellow and was awarded eight times Bulent Kerim Altay Award by department of electrical engineering in METU and received Best Student Paper Award in IEEE ISCC 2003 and has an invited paper in IEEE GLOBECOM CAMAD 200.
E-mail: ergen@eecs.berkeley.edu

Pravin Varaiya is Nortel Networks Distinguished Professor in the Department of Electrical Engineering and Computer Sciences at the University of California, Berkeley. From 1975 to 1992, he was also Professor of Economics at Berkeley. His research is concerned with communication networks, transportation, and hybrid systems. He has taught at MIT and the Federal University of Rio de Janeiro, Varaiya has held a Guggenheim Fellowship and a Miller Research Professorship. He received an Honorary Doctorate from LíInstitut National Polytechnique de Toulouse, and the Field Medal of the IEEE Control Systems Society. He is a Fellow of IEEE and a member of the National Academy of Engineering. He is on the editorial board of several journals, including "Discrete Event Dynamical Systems" and "Transportation Research-C." He has co-authored three books and more than 250 technical papers. The second edition of "High-Performance Communication Networks" (with Jean Walrand) was published by Morgan-Kaufmann in 2000. "Structure and interpretation of signals and systems" (with Edward Lee) was published in 2002 by Addison-Wesley.
E-mail: varaiya@eecs.berkeley.edu