IEEE 802.11—Saturation Throughput Analysis

Giuseppe Bianchi

***Abstract*—To satisfy the emerging need of wireless data communications, IEEE is currently standardizing the 802.11 protocol for wireless local area networks. This standard adopts a CSMA/CA medium access control protocol with exponential backoff. In this letter, we present a simple analytical model to compute the saturation throughput performance in the presence of a finite number of terminals and in the assumption of ideal channel conditions. The model applies to both basic and RTS/CTS access mechanisms. Comparison with simulation results shows that the model is extremely accurate in predicting the system throughput.**

***Index Terms*—IEEE 802.11 protocol, multiple access control, performance evaluation.**

I. Introduction

To ACCESS the medium, IEEE 802.11 employs a CSMA/CA (carrier sense multiple access with collision avoidance) MAC protocol with binary exponential backoff, called distributed coordination function (DCF) [1]. DCF defines a basic access method, and an optional four-way handshaking technique, known as *request-to-send/clear-to-send* (RTS/CTS) method.

This letter provides a simple but nevertheless very accurate analysis to compute the throughput of both basic and RTS/CTS access schemes, in the assumption of ideal channel (see [2] and [3] for approximate models that account for hidden terminals and capture). We focus on the saturation throughput, which is a fundamental performance figure defined as the limit reached by the system throughput as the offered load increases, and it represents the maximum load that the system can carry *in stable conditions* (as in most random access protocols, the maximum throughput may be greater than the saturation one, but it is of no practical importance as not sustainable for "long" time—see a general discussion about stability in [4]).

II. Distributed Coordination Function

The DCF basic access method [1] is shortly summarized as follows. A station with a packet to transmit, monitors the channel activity until an idle period equal to a distributed interframe space (DIFS) is detected. The time immediately following an idle DIFS is slotted, and a station is allowed to transmit only at the beginning of each *slot time*, defined as the time needed at any station to detect the transmission of a packet from any other station. It accounts for the propagation delay, for the time needed to switch from the receiving to the transmitting state (RX_TX_Turnaround_Time), and for the time to signal to the MAC layer the state of the channel (busy detect time).

Manuscript received February 9, 1998. The associate editor coordinating the review of this letter and approving it for publication was Dr. K. Sriram.

The author is with the Dipartimento di Elettronica e Informazione, Politecnico di Milano, 20133 Milano, Italy (e-mail: bianchi@elet.polimi.it).

Publisher Item Identifier S 1089-7798(98)09502-7.

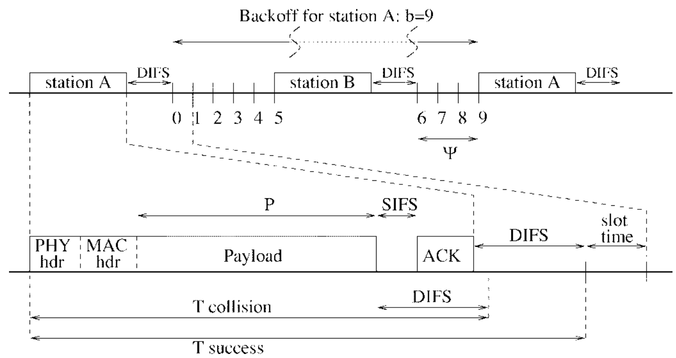

Fig. 1. Basic access mechanism.

After sensing an idle DIFS, the station generates a random backoff interval before transmitting. The backoff time counter is decremented as long as the channel is sensed idle, stopped when a transmission is detected on the channel, and reactivated when the channel is sensed idle again for more than a DIFS (see Fig. 1). The station transmits when the backoff time reaches zero. At each transmission, the backoff time is uniformly chosen in the range $(0, w-1)$. At the first transmission attempt, $w = W$, namely the minimum backoff window. After each unsuccessful transmission, w is doubled, up to a maximum value $2^m W$.

Since the CSMA/CA does not rely on the capability of the stations to detect a collision by hearing their own transmission, a positive acknowledgment (ACK) is transmitted by the destination station to signal the successful packet transmission. To allow an immediate response, the ACK is transmitted following the received packet, after a short interframe space (SIFS). If the transmitting station does not receive the ACK within a specified ACK_Timeout, or it detects the transmission of a different packet on the channel, it reschedules the packet transmission according to the previous backoff rules.

The standard defines an additional mechanism of four-way handshaking to be optionally used in the case a packet exceeds a specified length, to improve the system throughput by shortening the duration of the collisions. This mechanism requires the transmission of special short *request to send* (RTS) and *clear to send* (CTS) frames prior to the transmission of the actual data frame. As shown in Fig. 2, an RTS frame is transmitted by a station which needs to transmit a packet. When the receiving station detects an RTS frame, it responds, after a SIFS, with a CTS frame. The transmitting station is thus allowed to transmit its packet only if it correctly receives the CTS frame. Moreover, the frames RTS and CTS carry

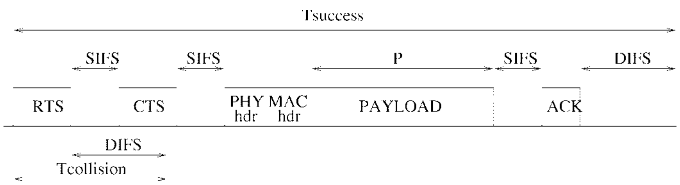

Fig. 2. RTS/CTS access mechanism.

the information of the length of the packet to be transmitted. This information can be read by each station, which is then able to update a *network allocation vector* (NAV) containing the information of the period of time in which the channel will remain busy. This latter technique has been introduced to combat the system degradation due to hidden terminals [5]. In fact, a station able to detect the transmission of at least one of the RTS or CTS frames, can avoid collision even when unable to sense the channel busy.

III. Throughput Analysis

Consider a fixed number n of contending stations. In saturation conditions, each station has immediately a packet available for transmission, after the completion of each successful transmission. Let $b(t)$ be the stochastic process representing the size of the backoff window for a given station at slot time t (note that, as shown in Fig. 1, the time is stopped when the channel is sensed busy). Clearly, this process is non-Markovian. However, define for convenience $W_i = 2^i W$, where $i \in (0, m)$ is called "backoff stage," and let $s(t)$ be the stochastic process representing the backoff stage $(0, \cdots, m)$ of the station at time t.

The key approximation in our model is that the probability p that a transmitted packet collides is independent on the state $s(t)$ of the station (this is more accurate as W and n are larger). In this condition, the bidimensional process $\{s(t), b(t)\}$ is a discrete-time Markov chain, for convenience depicted in Fig. 3, with the only nonnull one-step transition probabilities being[1]

$$\begin{cases} P\{i, k|i, k+1\} = 1, & k \in (0, W_i - 2); \\ & i \in (0, m) \\ P\{0, k|i, 0\} = (1-p)/W_0, & k \in (0, W_0 - 1); \\ & i \in (0, m) \\ P\{i, k|i-1, 0\} = p/W_i, & k \in (0, W_i - 1); \\ & i \in (1, m) \\ P\{m, k|m, 0\} = p/W_m, & k \in (0, W_m - 1). \end{cases}$$

These transition probabilities account, respectively, for: 1) the decrement of the backoff time counter; 2) the fact that a new packet following a successful transmission starts with a backoff stage 0; and 3), 4) the fact that after an unsuccessful transmission at backoff stage i, the backoff interval is uniformly chosen in the range $(0, W_{\min(i+1, m)})$.

Let $b_{i,k} = \lim_{t\to\infty} P\{s(t) = i, b(t) = k\}$, $i \in (0, m)$, $k \in (0, W_i - 1)$ be the stationary distribution of the chain.

[1] We adopt the short notation: $P\{i_1, k_1|i_0, k_0\} = P\{s(t+1) = i_1, b(t+1) = k_1|s(t) = i_0, b(t) = k_0\}$.

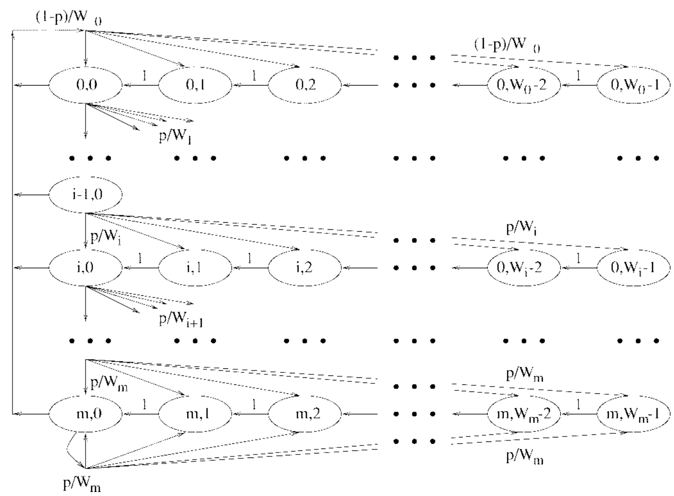

Fig. 3. Markov chain model for the backoff window size.

Owing to the chain regularities, the following relations hold:

$$\begin{aligned} b_{i,0} &= p^i b_{0,0}, \qquad i \in (0, m-1) \\ b_{m,0} &= \frac{p^m}{1-p} b_{0,0} \\ b_{i,k} &= \frac{W_i - k}{W_i} b_{i,0}, \qquad k \in (0, W_i - 1). \end{aligned}$$

The value of $b_{0,0}$ is determined by imposing the normalization condition

$$\begin{aligned} 1 &= \sum_{i=0}^{m} \sum_{k=0}^{W_i - 1} b_{i,k} = \sum_{i=0}^{m} b_{i,0} \sum_{k=0}^{W_i - 1} \frac{W_i - k}{W_i} \\ &= \sum_{i=0}^{m} b_{i,0} \frac{W_i + 1}{2} \\ &= \frac{b_{0,0}}{2} \left[W \left(\sum_{i=0}^{m-1} (2p)^i + \frac{(2p)^m}{1-p} \right) + \sum_{i=0}^{m-1} p^i + \frac{p^m}{1-p} \right] \end{aligned}$$

from which

$$b_{0,0} = \frac{2(1-2p)(1-p)}{(1-2p)(W+1) + pW(1-(2p)^m)}. \tag{1}$$

Let τ be the probability that a station transmits in a generic slot time. As any transmission occurs when the backoff window is equal to zero, regardless of the backoff stage, it is

$$\tau = \sum_{i=0}^{m} b_{i,0} = \frac{b_{0,0}}{1-p} = \frac{2(1-2p)}{(1-2p)(W+1) + pW(1-(2p)^m)}. \tag{2}$$

To finally compute the probability p that a transmitted packet collides, note that p is the probability that, in a time slot, at least one of the $n-1$ remaining stations transmits

$$p = 1 - (1-\tau)^{n-1} \tag{3}$$

where τ is given in (2) as function of the only unknown p. Numerically solving (3), the value of p, and therefore, of τ, is found. Once τ is known, the probability $P_{\rm tr}$ that in a slot time there is at least one transmission, given n active stations, and the probability P_s that a transmission is successful, are readily obtained as

$$P_{\mathrm{tr}} = 1 - (1-\tau)^n \tag{4}$$

$$P_s = \frac{n\tau(1-\tau)^{n-1}}{P_{\mathrm{tr}}} = \frac{n\tau(1-\tau)^{n-1}}{1-(1-\tau)^n}. \tag{5}$$

Being Ψ the r.v. representing the number of consecutive idle slots between two consecutive transmissions on the channel (see Fig. 1), it is

$$E[\Psi] = \frac{1}{P_{\mathrm{tr}}} - 1. \tag{6}$$

We are finally in the condition to determine the normalized system throughput S, defined as the fraction of time the channel is used to successfully transmit payload bits. As the instants of time right after the end of a transmission are renewal points, it is sufficient to analyze a single renewal interval between two consecutive transmissions, and express S as the ratio

$$\begin{aligned} S &= \frac{E[\text{timeused for successful transm. in interval}]}{E[\text{length of a renewal interval}]} \\ &= \frac{P_s E[P]}{E[\Psi] + P_s T_s + (1-P_s) T_c} \end{aligned} \tag{7}$$

where $E[P]$ is the average packet length, T_s is the average time the channel is sensed busy because of a successful transmission, and T_c is the average time the channel is sensed busy by the stations during a collision. The times $E[P]$, T_s and T_c must be measured in slot times, as this is the time unit of $E[\Psi]$.

To conclude the analysis, it remains only to specify the values T_s and T_c. Let $H = \mathrm{PHY}_{\mathrm{hdr}} + \mathrm{MAC}_{\mathrm{hdr}}$ be the packet header, and δ be the propagation delay. For the basic access method it is (see Fig. 1):

$$\begin{cases} T_s^{\mathrm{bas}} = & H + E[P] + \mathrm{SIFS} + \delta + \mathrm{ACK} + \mathrm{DIFS} + \delta \\ T_c^{\mathrm{bas}} = & H + E[P^*] + \mathrm{DIFS} + \delta \end{cases}$$

where $E[P^*]$ is the average length of the longest packet payload involved in a collision[2]: in the case all packets have the same fixed size, $E[P^*] = E[P] = P$. T_c is the time in which the channel is sensed busy by the *noncolliding stations*: we have neglected the fact that the two or more colliding stations, before sensing the channel again, need to wait an additional SIFS plus an ACK timeout (the same approximation holds in the following RTS/CTS case, with a CTS timeout instead of the ACK timeout).

For the RTS/CTS access method, assuming for simplicity of presentation that all the stations use the RTS/CTS mechanism for all the transmitted packets, regardless of the packet's length, it is (see Fig. 2)

$$\begin{cases} T_s^{\mathrm{rts}} = \mathrm{RTS} + \mathrm{SIFS} + \delta + \mathrm{CTS} + \mathrm{SIFS} + \delta + H \\ \qquad\quad + E[P] + \mathrm{SIFS} + \delta + \mathrm{ACK} + \mathrm{DIFS} + \delta \\ T_c^{\mathrm{rts}} = \mathrm{RTS} + \mathrm{DIFS} + \delta \end{cases}$$

[2] When the probability of three of more packets simultaneously colliding is neglected, this is given by $E[\max(P_1, P_2)]$, where P_i are independent and identically distributed r.v.s. To proceed further it is then necessary to assume a suitable pdf for the packet's length.

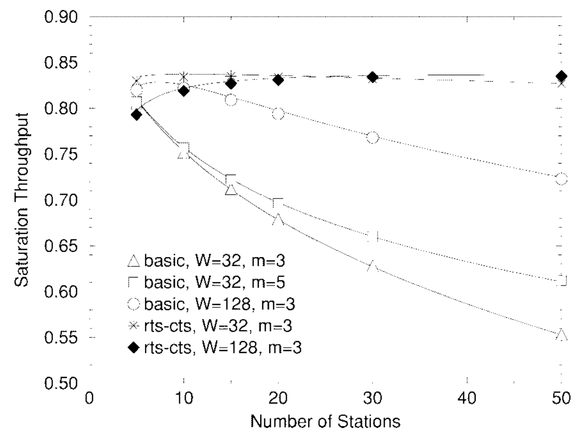

Fig. 4. Saturation throughput: Analysis versus simulation.

IV. Model Validation

To validate the model, we have compared its results with that obtained with the 802.11 DCF simulator used in [6], which accounts for all the protocol details. The packet payload length has been chosen constant and equal to 8184 bits. The other protocol and channel parameters adopted are those specified for the FH (frequency hopping) PHY layer [1], and are reported in the following table (additional parameters are specified in [6]):

packet payload	8184 bits
MAC header	272 bits
PHY header	128 bits
ACK length	112 bits + PHY header
RTS length	160 bits + PHY header
CTS length	112 bits + PHY header
Channel Bit Rate	1 Mbit/s
Propagation Delay	1 μs
SIFS	28 μs
Slot Time	50 μs
DIFS	128 μs

Fig. 4 shows that the analytical model is highly accurate: analytical results (lines) practically coincide with the simulation results (simbols), in both basic access and RTS/CTS cases. All simulation results are obtained with a 95% confidence interval lower than 0.002.

References

[1] P802.11, "IEEE standard for wireless LAN medium access control (MAC) and physical layer (PHY) specifications," Nov. 1997.
[2] H. S. Chhaya and S. Gupta, "Performance modeling of asynchronous data transfer methods of IEEE 802.11 MAC protocol," *Wireless Networks*, vol. 3, pp. 217–234, 1997.
[3] K. C. Huang and K. C. Chen, "Interference analysis of nonpersistent CSMA with hidden terminals in multicell wireless data networks," in *Proc. IEEE PIMRC'95*, Toronto, Ont., Canada, Sept. 1995, pp. 907–911.
[4] D. Bertsekas and R. Gallager, *Data Networks.* Englewood Cliffs. NJ: Prentice Hall, 1987.
[5] L. Kleinrock and F. Tobagi, "Packet switching in radio channels: Part II—The hidden terminal problem in carrier sense multiple access and the busy tone solution," *IEEE Trans. Commun.*, vol. COM-23, pp. 1417–1433, Dec. 1975.
[6] G. Bianchi, L. Fratta, and M. Oliveri, "Performance evaluation and enhancement of the CSMA/CA MAC protocol for 802.11 wireless LAN's," in *Proc. PIMRC 1996*, Taipei, Taiwan, R.O.C., Oct. 1996, pp. 392–396.