Performance Evaluation and Enhancement of the CSMA/CA MAC Protocol for 802.11 Wireless LANs*

Giuseppe Bianchi, Luigi Fratta, Matteo Oliveri

Politecnico di Milano, Dip. Elettronica, E-mail: bianchi@elet.polimi.it

ABSTRACT. The IEEE 802.11 protocol for Wireless Local Area Networks adopts a CSMA/CA protocol with exponential backoff as medium access control technique. As the throughput performance of such a scheme becomes critical when the number of mobile stations increases, in this paper we propose an *Adaptive Contention Window* mechanism, which dynamically selects the optimal backoff window according to the estimate of the number of contending stations. We show that this technique leads to stable behavior, and it outperforms the standard protocol when the network load and the number of mobile stations are high. We also investigate the CSMA/CA with the optional RTS/CTS technique, and we show that our adaptive technique reaches better performance only when the packet size is short. Finally, the performance of a system environment with hidden terminals show that the RTS/CTS mechanism, which can also be used in conjunction with the adaptive contention window mechanism, provides significant improvements.

1 Introduction

In recent years, much interest has been involved in the design of wireless networks for local area communications [1, 2, 3]. At present, IEEE is being proposing the 802.11 standard [4], whose scope is the definition of the MAC and physical layer specifications for wireless LANs.

The basic mechanism to access the medium is the Distributed Coordination Function (DCF), which is based on the CSMA/CA (Carrier Sense Multiple Access with Collision Avoidance) protocol with binary exponential backoff. In addition to the DCF performance evaluation already available in the literature [1, 2, 6, 7], the main contribution of this paper is a more detailed investigation of the 802.11 performance drawbacks and their dependency on the system parameters.

In order to carefully capture all the 802.11 protocol details, the performance has been evaluated via simulation. We show that the adoption of the exponential backoff technique leads to throughput performance strongly dependent on the initial backoff window size and, most important, on the number of terminals considered in the network. To avoid this drawback, we propose to adopt an adaptive backoff window control technique, based on a dynamic estimate of the number of contending stations. This technique significantly improves the performance, which also become independent from the number of contending stations.

In the 802.11 standard, optional enhancements of the DCF have been introduced to increase the system performance when long messages are transmitted, and to provide a more robust protocol in the presence of fading channels and *hidden terminals* [5]. In particular, a mechanism based on a four way handshaking technique, known as Request-To-Send/Clear-To-Send (RTS/CTS) protocol, has been considered to shorten the duration of the collisions, and to reduce their number in the presence of hidden terminals. In this paper, we extend the performance evaluation also to this optional technique, showing a substantial enhancement of the performance, even in ideal channel conditions, when the size of the packets is large.

2 The 802.11 Distributed Coordination Function

The Distributed Coordination Function (DCF) defined in [4], is based on the CSMA/CA protocol. A station with a packet to transmit, monitors the channel activity until an idle period equal to a Distributed InterFrame Space (DIFS) has been observed. In case the medium is sensed busy, a random backoff interval is selected. 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 figure 1). The station transmits when the backoff time reaches 0. In addition, to avoid channel capture, a station must wait a random backoff time between two consecutive packet transmissions, even if the medium is sensed idle in the DIFS time.

The DCF adopts a slotted binary exponential backoff technique. In particular, the time immediately following an idle DIFS is slotted, and a station is allowed to transmit only at the beginning of each *Slot Time*, which is equal to 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). The values of the parameters used in the simulation program, are summarized in table 1, and are those specified for the FH (Frequency Hopping) PHY layer [4].

The backoff time is uniformly chosen in the interval $(0, CW - 1)$ defined as the Backoff Window (Contention Window). At the first transmission attempt, $CW = CWmin$, and it is doubled at each retransmission up to CWmax. The values suggested in the draft standard [4] are $CWmin = 32$ and $CWmax = 256$.

Since the CSMA/CA does not rely on the capability of the stations to detect a collision by hearing their own transmission, a positive acknowledgement is transmitted by the destination station to signal the successful packet transmission. To allow an immediate response, the acknowledgement is transmitted following the received packet, after a Short InterFrame Space (SIFS). If the transmitting station does not receive the acknowledge 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. As the value of the ACK_Timeout is not specified in the standard, it has been chosen sufficiently long ($300\,\mu s$) to contain a SIFS, the ACK transmission and a round trip delay.

The protocol provides a fragmentation mechanism, which allows the MAC to split an MSDU (the packet delivered to the MAC by the higher layers) into more MPDUs (packets delivered by the MAC to the PHY layer), if the MSDU size exceeds the maximum MPDU payload size. The different fragments are then transmitted in sequence, with only a SIFS between them, so that only the first fragment must contend for the channel access. In our simulations, we have assumed no MSDU fragmentation, i.e. each MSDU corresponds exactly to an MPDU of fixed size. Each MPDU is composed of a payload, a MAC header, and a PHY header, whose sizes, shown in table 1, are those defined in [4], except for the payload length that we have chosen equal to half of the maximum value defined in the standard.

The standard defines an additional mechanism of four-way handshaking to be optionally used in the case an MPDU exceeds a specified length. This mechanism, which has been introduced to reduce the length of the frames involved in the contention process, 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. The 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 the CTS frame is correctly received. Moreover, the frames RTS and CTS carry 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 transmis-

---

Fig. 1. *Basic access mechanism*

Fig. 2. *Throughput versus offered load for the CSMA/CA protocol*
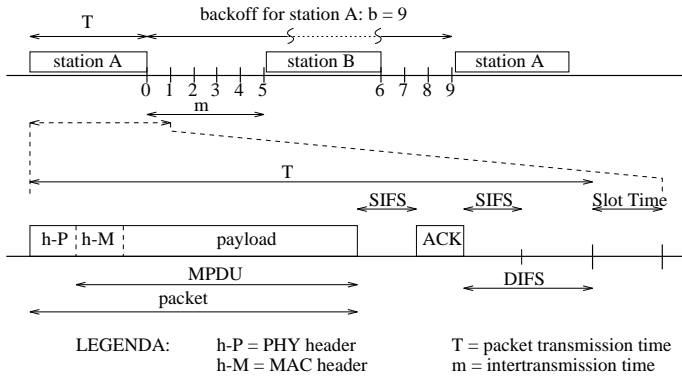
sion of at least one of the RTS or CTS frames, can avoid collision even in the case that it is not able to sense the channel busy.

3 Performance of the basic DCF access scheme

In our simulations, a fixed number of stations, generating packets according to a Poisson distribution, is considered. With the exception of Section 7, we assume an ideal channel with no transmission errors and no *hidden terminal*, i.e. all stations can always hear all the others.

To evaluate the performance of the DCF access scheme, we have measured the system throughput and the access delay. Both offered load and throughput are normalized with respect to the channel capacity, and measured taking into account the packet payload only, thus ignoring the MAC and PHY headers. The access delay is measured from the time a packet becomes head of the line in the transmission queue until the successful transmission of its first bit.

Figure 2 shows the throughput versus the offered load for three different network scenarios with 5, 10 and 20 stations. As the offered load increases, the throughput reaches a saturation value which depends on the number of contending stations: the higher the number of stations, the lower the throughput. Moreover, as expected [8], an unstable behavior, which is more significant as the number of stations increases, is observed.

As shown in figure 3, the saturation throughput highly depends on the number of contending stations and on the values of the contention window limits. For a given pair of contention window limits, the throughput drastically decreases as the number of stations increases. Moreover, we have observed that, regarding the effects of the contention window limits, the most critical parameter is $CWmin$. In fact, especially for a large number of stations, an initially small contention window size does not provide a sufficiently small collision probability.

This result is verified in figure 4, which reports the mean number of attempts needed to transmit a packet in saturation load conditions. The small improvement obtained by increasing $CWmax$ (from 256 to 1024) shows that even with the exponential growth of the backoff window size, the protocol is unable to keep small the number of attempts per packet. A much larger improvement is obtained by increasing $CWmin$.

Its impact on the performance has been investigated by measuring the saturation throughput for different values of $CWmin$, by assuming $CWmax = 8 \cdot CWmin$. The results for the three cases of 5, 20, and 50 contending stations are shown in figure 5. While the throughput highly depends on $CWmin$, the optimal value of

Fig. 3. *Saturation Throughput versus the number of stations*

Fig. 4. *Average number of attempts per packet versus the number of stations, in saturation load conditions*

$CWmin$ depends on the network scenario. For example, an high value of $CWmin$ (1024) which is optimum for 50 contending stations, would drastically penalize the throughput in the case of small number (5) of contending stations. Large values of $CWmin$ may, in fact, strongly limit the throughput of a single station, which, when alone in the channel, is bounded by $P/[T + (CWmin - 1)/2]$, where P and T are the packet payload and the total packet (payload + headers + SIFS + acknowledgement, see figure 1) lengths, respectively.

Similar effects have been observed when considering the access delay performance. For the scenario of 5 contending stations, figure 6 shows the average access delay versus the offered load for several values of $CWmin$ and $CWmax$. Of course, the average access delay increases with the initial contention window size.

packet payload	8184 bits	ACK length	112 bits + PHY
MAC header	272 bits	PHY header	128 bits
RTS payload	160 bits	CTS payload	112 bits
Channel Bit Rate	1 Mbit/s	Propagation Delay	1 μs
RxTx_Turnaround_Time	20 μs	Busy_Detect_Time	29 μs
SIFS	28 μs	DIFS	130 μs
ACK_Timeout	300 μs	CTS_Timeout	300 μs
Slot Time	50 μs		

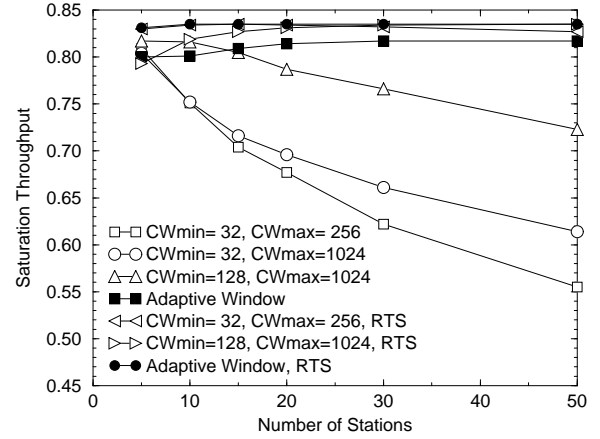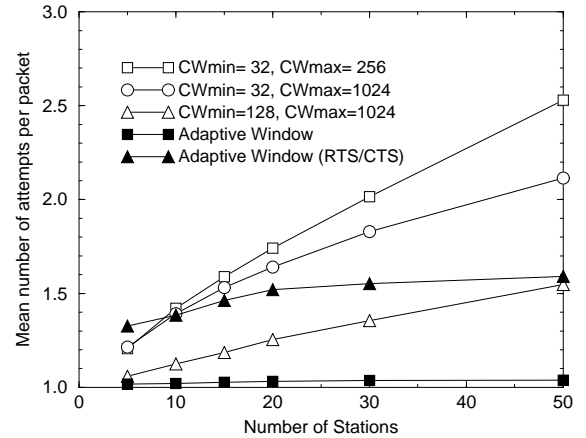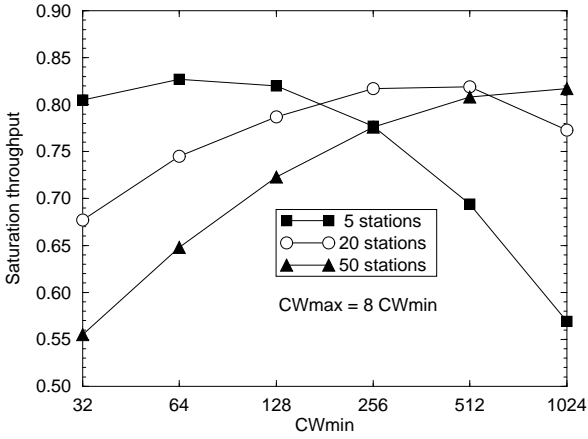**Table 1.** *Parameter values used in the simulations*

Fig. 5. *Throughput versus the initial size of the contention window*
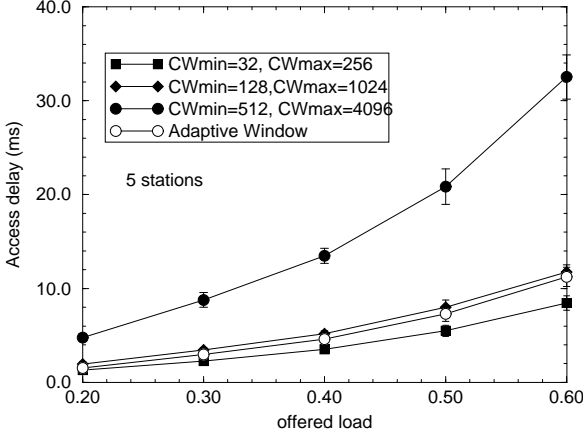

Fig. 6. *Average access delay for a network of 5 station in low load conditions*

4 Adaptive Contention Window

The performance of CSMA/CA, presented in the previous section, has shown that to achieve an optimal operation, the system parameters must be properly selected according to traffic conditions. In particular, the fact that the optimal value of $CWmin$ depends on the number of contending stations, suggests that the CSMA/CA can be improved by dynamically selecting the contention window size according to an estimate of the number of the contending stations based on measurements of the channel activity, performed by each station.

The algorithm is, in principle, trivial. A station which has a packet to transmit, extracts a random backoff b uniformly in the range $(0, W-1)$, where W is its current value of the contention window. Based on the measurements of the channel activity, the station estimates the number $\bar{n}(t)$ of stations contending at time t. According to this estimate, the station continuously modifies the value of the contention window W as:

$$W = \bar{n}\sqrt{2T} \qquad (1)$$

where T is the total packet transmission time (including headers, SIFS, ACK and DIFS, as shown in figure 1), measured in slot times. We will show in Appendix A that this value for W is, in first approximation, the value which maximizes the throughput of the system, given that n stations are contending on the channel.

To this purpose, we assume that each station is able to detect whether at each slot time the channel is idle or busy, and whether its own transmission is successful or collided. If $c(B)$ is the number of slot times observed busy due to other stations transmission, over a period of B slot times, the estimate of the probability that a station will observe a packet transmission in a slot time is given by $P_{tr}^* = c(B)/B$. In the case all stations use the same contention window W,

according to equation (9) of Appendix A applied to $n-1$ stations, the following equality holds:

$$E[c(B)] = B \cdot [1 - (1 - p_0)^{n-1}] \approx B \cdot (n-1)p_0 = \frac{2B \cdot (n-1)}{W+1} \qquad (2)$$

From this relation, we can express n as:

$$n \approx 1 + \frac{E[c(B)](W+1)}{2B} \qquad (3)$$

For convenience, the interval B can be chosen equal to the backoff period b, plus the slot time 0 in which the station transmits. It results that, at each packet transmission attempt, based on the count $c(b+1)$ and on relation 3, the station provides an estimate $\hat{n}$ of the number of contending stations.

The evaluation of the estimate $\bar{n}$, from the single values of $\hat{n}$, is critical since each $\hat{n}$ is computed on the basis of the current measure of $c(b+1)$, instead of its mean value. In order to provide a smoother behavior of the estimate, we weight the current measure $\hat{n}$ with the previous values, by using a linear filter of the type:

$$\bar{n}(t+1) = \alpha\bar{n}(t) + (1-\alpha)\sum_{i=1}^{q}\hat{n}_i \qquad (4)$$

where $\bar{n}(t)$ is the value used to compute W, t is incremented at each packet transmitted by the station, and $\hat{n}_i$ are the last q values estimated by (3), i.e. those computed on the last q packet transmissions. The parameters $\alpha = .8$ and $q = 10$ have been chosen as a result of an heuristic analysis based on the minimization of the estimation error in the assumption of $n(t)$ constant for fixed size intervals.

Since formula (3) involves the previous estimate of the contention window W, it is simple to show [9] that this estimate is instable, i.e. that, if a station has a current estimate $\bar{n}$ significantly different from the actual value of n, such an estimate keeps stuck at this arbitrarily wrong value. To avoid this problem, we have added a non linear function in the computation of the adaptive W, which becomes:

$$W = s(n)\sqrt{2Tn} \qquad (5)$$

where $s(n) = (1 + h/\sqrt{n})$ has been chosen as a smooth decreasing function with the constraint $s(n) \rightarrow 1$ for $n \rightarrow \infty$ (The proof of stability of the estimate is given in [9]).

Simulation results, not reported here (see [9]), show that the rate of convergence increases with h, but at the same time the accuracy of the value of W decreases, especially for a small number of stations. As a good compromise, in the following results we have adopted $h = 2$.

Figure 7 shows the throughput as a function of W, computed by equation (11) of Appendix A. The window sizes W at which the system operates according to equations (1) and (5) are also marked on the throughput curves. The stabilizing function introduced in equation (5) increases the values of W, which turn out to be larger than the optimum ones. However, the consequent throughput degradation is almost negligible, especially when the number of stations is large.

5 Performance of the Adaptive Window Algorithm

The maximum throughput performance of the adaptive window technique is reported in figure 3. We can note that the maximum throughput is practically independent from the number of stations in the network. The slight reduction of the throughput for a small number of stations is due to the increase of the contention window size introduced by the stabilization factor. However, this fact does not excessively penalize the access delay, as shown in figure 6.

The effectiveness of the adaptive window mechanism is due to its capability to keep a very low collision probability, regardless of the number of stations in the system. This is shown in figure 4, where the mean number of transmission attempts per packet, achieved by
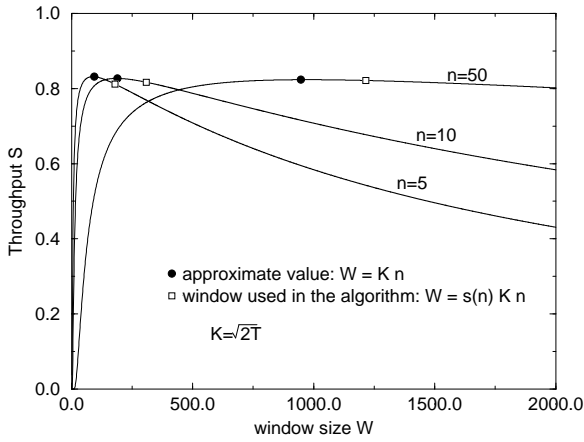
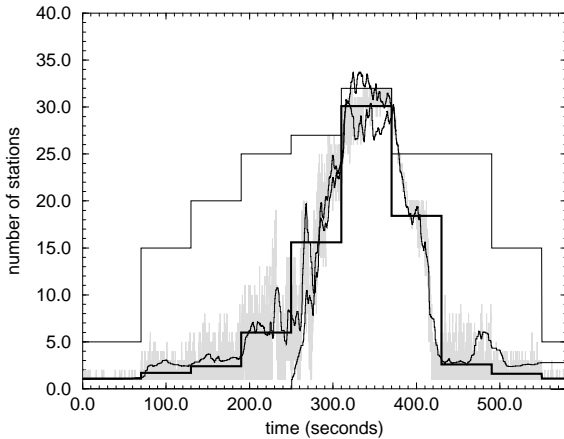**Fig. 7.** *Analytical throughput versus the contention window size*

Fig. 8. *A test of the Adaptive Window Algorithm with the number of stations variable during the time*

Fig. 9. *Saturation Throughput versus packet size*

Fig. 10. *Saturation Throughput versus the hidden terminals probability*

the different systems, is compared. In the case of adaptive window, the average number of retransmissions is less that 0.05.

In order to verify that the adaptive window algorithm effectively behaves in presence of traffic fluctuations, we have simulated a system with a variable number of active stations. The results in figure 8 refer to the case where each active station offers a traffic equal to 0.033 of the channel capacity. The variation pattern of the active stations is represented by the upper stair case function, while the lower stair case function shows the actual average number of stations participating to each contention. The shaded line represents the instantaneous value of $n(t)$. The two solid curves report the estimates $\bar{n}(t)$ done by two stations in the network, and show that these estimates basically follow the average value of $n(t)$ as the offered load changes.

6 Performance of the RTS/CTS mechanism

The maximum throughput obtained by operating the RTS/CTS mechanism in both cases of exponential backoff and adaptive contention window is shown in figure 3. When the RTS/CTS mechanism is used in conjunction with the adaptive window technique, it is necessary to recompute the optimal value of the contention window W, which is given by $W_{RTS/CTS} \approx n\sqrt{2K}$, where $K = H_p + RTS + \tau + DIFS$, and H_p and RTS are the lengths of the physical header and the RTS frame, respectively (the computation is similar, although slightly more complex, to the one considered in Appendix A: see details in [9])

The throughput of the standard protocol with the RTS/CTS mechanism is very close to that of the adaptive window mechanism, as the dimensioning of the contention window in the RTS/CTS mechanism
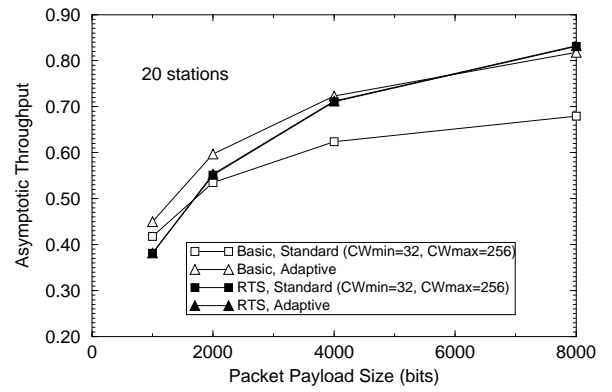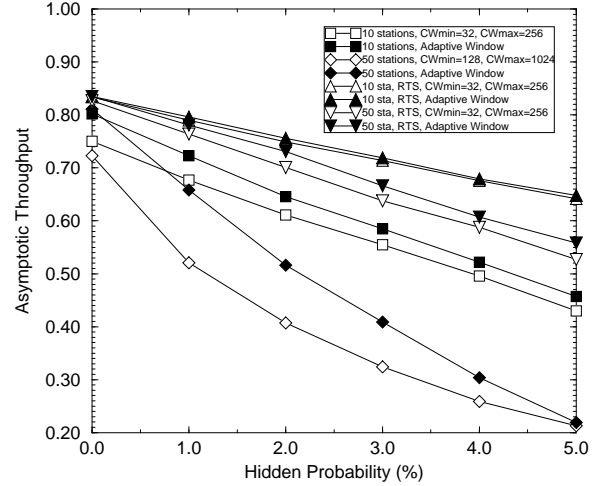
is much less critical than in the standard protocol. In all simulated cases, the optional mechanism proves to be effective, and it shows to be almost mandatory if the adaptive contention window technique is not adopted.

Note that the RTS/CTS mechanism reaches better performance by forcing more collisions, as observed by the increased number of attempts per packet shown in figure 4. This behavior is explained by observing that, in this case, collisions are much shorter than in the basic protocol, and the corresponding throughput degradation is reduced. On the other hand, more attempts reduce the probability of empty slots, with a consequent throughput increase.

Finally, figure 9 shows that the effectiveness of the RTS/CTS mechanism depends on the packet payload size, and, because of the RTS/CTS frame transmission overhead, it increases as the packet length increases. The trade off between pros and cons of RTS/CTS show that it is profitably applied in conjunction with the adaptive contention window only for long packets (> 6000 bits), while in the basic standard, it provides benefits for packets longer than 2000 bits.

7 Performance of CSMA/CA with hidden terminals

The use of CSMA/CA for wireless networks has been criticized in the literature [5, 1, 2] for its low robustness in the presence of interference and fading channels. In the latter case, a phenomenon known as the problem of *hidden terminals* may arise. A terminal A is said to be hidden from a terminal B if, during a packet transmission by terminal B, A senses the channel idle. Hence, during this period, A may start a packet transmission, causing an unexpected collision at the stations that hear both A and B.

In order to provide preliminary numerical results, we have modeled the hidden terminal phenomenon by the *hidden probabilities* P_{ij}, i.e. the probability that the station i does not hear the station j. In our simulations, we have assumed the hidden probability constant and statistically independent for each packet transmission, and for all pairs of stations, i.e. $P_{ij} = P_h$.

Figure 10 compares the throughput performance of both the basic DCF and the Adaptive window technique versus the hidden probability P_h, with and without the RTS/CTS mechanism. It can be noticed that, for both protocols, if the RTS/CTS mechanism is not used, the throughput degradation is dramatic even for very small P_h, and it is more critical as the number of stations increases. In any case, the adaptive window mechanism outperforms the basic protocol.

When the RTS/CTS mechanism is considered, performance are good even with a very high hidden probability. Moreover, we note from figure 10 that the adaptive window algorithm performs slightly better that the basic mechanism with RTS/CTS when the number of stations is high (50).

8 Conclusions

The numerical results reported in this paper clearly show that the CSMA/CA basic protocol suffers of several performance drawbacks. In particular, the throughput performance is strongly dependent on the number of active stations, and on the total load offered to the system.

Performance can be substantially enhanced if the exponential backoff window adopted in the CSMA/CA (DCF) protocol is substituted by an adaptive contention window, depending on the number of contending stations and optimized to maximize the throughput of the system. As shown in this paper, the number of contending stations can be easily estimated on the basis of the number of contentions detected on the channel by each station. The proposed algorithm is able to follow the traffic fluctuations, and to rapidly converge to an accurate estimate of the current number of stations.

Another technique to enhance the performance is to adopt the RTS/CTS mechanism optionally considered in the 802.11 standard. This mechanism can be used in conjunction with both basic and adaptive window mechanism, although a substantial performance improvement is obtained only in the first case. However, the RTS/CTS mechanism results of vital importance when the phenomenon of hidden terminals is taken into account. In fact, in this case, both basic and adaptive mechanisms suffer a severe performance degradation, even with a very small hidden probability.

Appendix A

Consider a fixed number n of *greedy* contending stations, i.e. stations that always have packets to transmit. For a fixed contention window W, and assuming that the backoff value $b(t)$, initially randomly chosen in the range $(0, W - 1)$, is decremented at each slot, the stochastic process $b(t)$ can be modeled by the following discrete Markov chain:

$$\begin{cases} P\{b(t+1) = k\} = P\{b(t) = k+1\} + \frac{P\{b(t)=0\}}{W}, \ k < W - 1 \\ P\{b(t+1) = W - 1\} = \frac{1}{W} P\{b(t) = 0\} \qquad k = W - 1 \end{cases}$$
(6)

where the term $P\{b(t) = 0\}/W$ accounts for the fact that, after a packet transmission, the new backoff value is uniformly chosen in the range $(0, W - 1)$, and the term $P(b(t) = k + 1)$ corresponds to the decrement of the backoff value at each slot time. The steady state probability $P\{b(t) = k\}$ is given by:

$$\lim_{t \to \infty} P\{b(t) = k\} = p_k = \frac{2(W - k)}{W(W + 1)}$$
(7)

As a station transmits in a slot with probability

$$p_0 = \frac{2}{W + 1}$$
(8)

the probability that in a slot there is at least one transmission given n active stations, is given by:

$$P_{tr} = 1 - (1 - p_0)^n$$
(9)

Therefore, the average number of consecutive idle slots is $E[m] = 1/P_{tr} - 1$. Since the activity of the channel may be defined as the ratio between a busy period and a busy + idle period, it is given by $T/(T + E[m])$, where T is the total packet transmission time measured in slot times (including headers, SIFS, ACK and DIFS, as shown in figure 1).

The system throughput S is derived from the average activity by observing that only active slots with successful transmissions contribute to the throughput. The probability that a transmission is successful is given by:

$$P_s = \frac{np_0(1 - p_0)^{n-1}}{1 - (1 - p_0)^n}$$
(10)

Therefore,

$$S = P_s \frac{\gamma T}{T + E[m]}$$
(11)

where γ $(\gamma < 1)$ accounts for the constant fraction of T reserved to the packet payload field. It is clear from the equation (11) that, as S depends on the contention window size via p_0, the optimum value of W is obtained by the solution of the equation:

$$\frac{dS}{dW} = \frac{dS}{dp_0} \cdot \frac{dp_0}{dW} = 0$$

Which implies, according to eq. (8), that $\frac{dS}{dp_0} = 0$, and therefore,

$$(1 - p_0)^n - T\{np_0 - [1 - (1 - p_0)^n]\} = 0$$
(12)

Under the condition $W >> n > 1$,

$$(1 - p_0)^n \approx 1 - np_0 + \frac{n(n - 1)}{2}p_0^2$$

holds, and provides the following approximate solution:

$$p_0 = \frac{\sqrt{n[n + 2(n - 1)(T - 1)]} - n}{n(n - 1)(T - 1)} \approx \frac{\sqrt{2}}{n\sqrt{T}}$$
(13)

From equation (8), the approximate value of the contention window which maximizes the saturation throughput is thus given by:

$$W \approx n\sqrt{2T}$$
(14)

References

1. A. De Simone, S. Nanda, "Wireless Data: Systems, Standards, Services", Wireless Networks, Vol. 1, No. 3, February 1996, pp. 241-254.
2. K. C. Chen, "Medium Access Control of Wireless LANs for Mobile Computing", IEEE Networks, September/October 1994, pp. 50-63.
3. K. Pahlavan, A. H. Levesque, "Wireless Data Communications", Proc. of the IEEE, Vol. 82, No. 9, September 1994, pp. 1398-1430.
4. P802.11, IEEE Draft Standard for Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specifications, D2.0, July 1995.
5. L. Kleinrock, 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. Comm., Vol. COM-23, No. 12, December 1975, pp. 1417-1433.
6. H. S. Chhaya, S. Gupta, "Throughput and Fairness Properties of Asynchronous Data Transfer Methods in the IEEE 802.11 MAC Protocol", Proc. of IEEE PIMRC '95, Toronto, September 1995, pp. 613-617.
7. K. C. Huang, K. C. Chen, "Interference Analysis of Nonpersistent CSMA with Hidden Terminals in Multicell Wireless Data Networks", Proc. of IEEE PIMRC '95, Toronto, September 1995, pp. 907-911.
8. D. Bertsekas, R. Gallager, *Data Networks*, Prentice Hall, Inc., 1987.
9. G. Bianchi, L. Fratta, M. Oliveri, "On the performance of the 802.11 protocol for Wireless Local Area Networks", submitted for publication.