

Investigation of the IEEE 802.11 Medium Access Control (MAC) Sublayer Functions

B. P. Crow, I. Widjaja, J. G. Kim, and P. Sakai

Department of Electrical and Computer Engineering
University of Arizona
Tucson, AZ 85721

Abstract

Analysis of the draft IEEE 802.11 wireless local area network (WLAN) standard is needed to characterize the expected performance of the standard's ad hoc and infrastructure networks. The performance of the medium access control (MAC) sublayer, which consists of Distributed Coordination Function (DCF) and Point Coordination Function (PCF), is determined by simulating asynchronous data traffic in a 1 Mbps ad hoc network, and asynchronous data and packetized voice traffic in a 1 Mbps infrastructure network. The simulation models incorporate the effect of burst errors, packet size, RTS threshold and fragmentation threshold on network throughput and delay. The results show that the IEEE 802.11 WLAN can achieve a reasonably high efficiency when the medium is almost error-free, but may degrade appreciably under harsh fading. The results also show that time-sensitive traffic such as packet voice can be supported together with other intensive traffic such as packet data. However, an echo canceler is required for packet voice systems.

1 Introduction

Wireless local area networks (WLAN) are being developed to provide high bandwidth to users in a limited geographical area. Typical data rates for WLANs can be expected to range from 1 Mbps to 20 Mbps [5]. WLANs will then be attached to backbone networks like the Internet, or interface to wireless wide area networks (WAN) for range extension. Multimedia traffic requiring constant, variable, and available bit rate services will be employed on WLANs in the future.

The IEEE is in the process of developing an international WLAN standard identified as IEEE 802.11 [7]. The scope of the standard is the physical layer (PHY) and medium access control (MAC) sublayer implementation. The IEEE 802.11 draft standard describes mandatory support for a 1 Mbps WLAN with optional support for a 2 Mbps data transmission rate. Mandatory support for asynchronous data transfer is specified as well as optional support for distributed time bounded services (DTBS). These services, such as packetized voice and video, must be transferred with minimal delay for an acceptable quality of service. The purpose of this paper is to characterize the performance of the IEEE 802.11 MAC sublayer by determining the throughput and delay via discrete

event simulation. An analytical method is not done because it would be very difficult to accurately evaluate an infrastructure network incorporating both the Distributed Coordination Function (DCF) and Point Coordination Function (PCF).

At present, little work has been done to model the performance of the IEEE 802.11 standard. All performance analysis work to-date has concentrated on the DCF MAC protocol [1][2][4][6]. The contributions made in this paper include a detailed performance evaluation of the DCF, and the DCF with PCF co-existing together. We also investigate the effect of channel errors on the performances of PCF and DCF, which is absent in all previous studies. The joint simulation of the DCF and PCF will include asynchronous data traffic using variable packet lengths on the DCF, and packetized voice traffic (which is time-bounded) on the PCF. Channel degradation, in terms of burst errors due to multipath fading, will be factored into the simulations and the effects on throughput and delay will be determined. We also develop an efficient polling scheme used during the PCF to drop inactive stations from the polling list for a polling cycle, thereby providing more bandwidth to currently active stations.

Readers desiring more detailed information should consult the standard and the papers submitted to the working group, which are located on the Internet at uniform resource locator (url): <file://atg.apple.com/pub/802.11/>. Due to space constraint, this paper can only give a few performance results. Many more results can be found in [3].

2 Description of Standard

2.1 Architecture

The Basic Service Set (BSS) is the fundamental building block of the IEEE 802.11 architecture. A BSS is defined as a group of stations that are under the direct control of a single coordination function, i.e., a DCF or PCF which is defined in section 2.3 below. The geographical area covered by the BSS is known as the Basic Service Area (BSA), which is analogous to a cell in a cellular communications network. Conceptually, all stations in a BSS can communicate directly with all other stations in a BSS (i.e., no "hidden terminals").

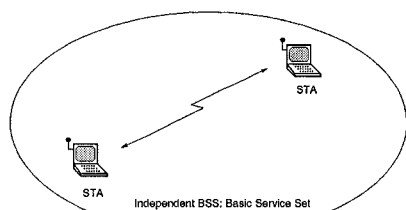


Figure 1: Sketch of an ad hoc network.

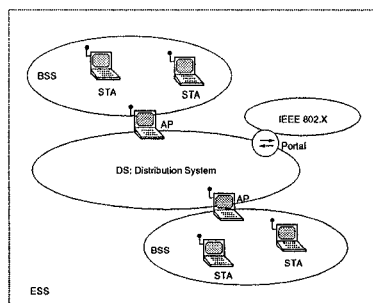


Figure 2: Sketch of an infrastructure network.

A single BSS can be used to form an ad hoc network. An ad hoc network is a deliberate grouping of stations into a single BSS for the purposes of internetworked communications without the aid of an infrastructure network. Figure 1 is an illustration of an ad hoc network. Note that a minimum of two stations can make an ad hoc network. The formal name of an ad hoc network in the IEEE 802.11 standard is an Independent Basic Service Set (IBSS). Any station can establish a direct communications session with any other station in the BSS, without the requirement of funneling traffic through a centralized access point (AP).

In contrast to the ad hoc network, infrastructure networks are established to provide wireless users with specific services and range extension. Infrastructure networks in the context of IEEE 802.11 are established using APs. The AP is analogous to the base station in a cellular communications network. The AP supports range extension by providing the integration points necessary for network connectivity between multiple BSSs, thus forming an Extended Service Set (ESS). The ESS consists of multiple BSSs that are integrated together using a common Distribution System (DS). The DS can be thought of as a backbone network that is responsible for MAC level transport of MAC service data units (MSDU). An ESS can also provide gateway access for wireless users into a wired network such as the Internet. This is accomplished via a device known as a *portal*. The portal is a logical entity that specifies the integration point on the DS where the IEEE 802.11 network integrates with a non-IEEE 802.11 network. Figure 2 illustrates a simple ESS developed with two BSSs, a DS, and a portal access to a wired LAN.

2.2 Physical Layer

The IEEE 802.11 draft specification calls for three different physical layer implementations: Frequency Hopping Spread Spectrum (FHSS), Direct Sequence Spread Spectrum (DSSS), and Infrared (IR). The performance issues identified in this paper are primarily concerned with the MAC sublayer parameters. Pertinent DSSS physical layer parameters are used in the simulation and are illustrated in Tables 1 and 2. Readers interested in further physical layer details should consult the draft specification.

2.3 MAC Sublayer

The MAC sublayer is responsible for the channel allocation procedures, protocol data unit (PDU) addressing, frame formatting, error checking, and fragmentation and reassembly. The transmission medium can operate in the contention mode exclusively, requiring all stations to contend for the channel for each packet transmitted. The medium can also alternate between the contention mode, known as the contention period (CP), and a contention free period (CFP). During the CFP, the medium usage is controlled (or mediated) by the AP, thereby eliminating the need for stations to contend for channel access. IEEE 802.11 supports three different types of frames: management, control, and data. The management frames are used for station association and disassociation with the AP, timing and synchronization, and authentication and deauthentication. Control frames are used for handshaking during the CP, for positive acknowledgments during the CP, and to end the CFP. Data frames are used for the transmission of data during the CP and CFP, and can be combined with polling and acknowledgments during the CFP.

2.3.1 Distributed Coordination Function

The DCF is the fundamental access method used to support asynchronous data transfer on a best effort basis. As identified in the specification, all stations must support the DCF. The MAC architecture is depicted in Figure 3, where the DCF sits directly on top of the physical layer and supports contention services. Contention services imply that each station with an MSDU queued for transmission must contend for the channel, and once the MSDU is transmitted, must contend for the channel for all subsequent frames.

The DCF is based on the Carrier Sense Multiple Access with Collision Avoidance (CSMA/CA) protocol. The reason is that, even though the wireless LAN is a broadcast medium, the traditional CSMA/CD (Collision Detection) will not function properly because the station is unable to listen to the channel for collisions while transmitting. In IEEE 802.11, carrier sensing is performed at both the air interface, referred to as (*physical carrier sensing*), and at the MAC sublayer, referred to as (*virtual carrier sensing*). Physical carrier sensing detects the presence of other IEEE 802.11 WLAN users by analyzing all detected packets and also detects activity in the channel via relative signal strength from other sources.

Virtual carrier sensing is used by a source station to inform all other stations in the BSS of how long

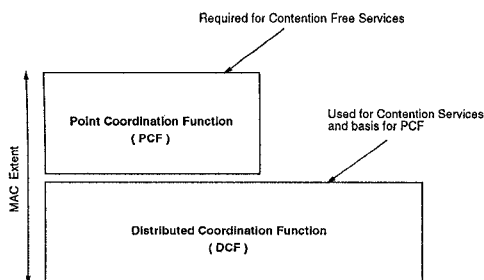


Figure 3: MAC architecture.

the channel will be utilized for the successful transmission of a MAC protocol data unit (MPDU). An MPDU is a complete data unit that is passed from the MAC sublayer to the physical layer. The MPDU contains header information, payload, and a 32-bit CRC. The source stations set the duration field in the MAC header of data frames, or Request to Send (RTS) and Clear to Send (CTS) control frames. The duration field indicates the amount of time (in microseconds) after the end of the present frame that the channel will be utilized to complete the successful transmission of the frame. Stations detecting a duration field in a transmitted MSDU adjust their Network Allocation Vector (NAV), which indicates the amount of time that must elapse until the current transmission session is complete and the channel can be sampled again for idle status. The channel is marked busy if either the physical or virtual carrier sensing mechanism indicates the channel is busy.

Priority access to the wireless medium is controlled through the use of Inter-Frame Space (IFS) time intervals between the transmission of frames. The IFS intervals are mandatory periods of idle time on the transmission medium. Three IFS intervals are specified in the standard: Short-IFS (SIFS), Point Coordination Function-IFS (PIFS), and Distributed Coordination Function-IFS (DIFS). The SIFS interval is the smallest IFS, followed by PIFS and DIFS, respectively. Stations only required to wait a SIFS have priority access over those stations required to wait a PIFS or DIFS before transmitting.

Figure 4 is a timing diagram illustrating the successful transmission of a data frame. When the data frame is transmitted, the duration field of the frame is used to let all stations in the BSS know how long the medium will be busy. All stations hearing the data frame, adjust their NAV based on the duration field value, which includes the SIFS interval and the acknowledgment frame following the data frame.

Since a source station in a BSS cannot hear simultaneously, when a collision occurs, the source continues transmitting the complete MPDU. If the MPDU is large, channel bandwidth is wasted due to a corrupted MPDU. RTS and CTS control frames can be used by a station to reserve channel bandwidth prior to the transmission of an MPDU so that bandwidth wasted due to collisions is minimized. RTS and CTS control frames are relatively small (RTS is 20 octets and CTS is 14 octets) when compared to the maxi-

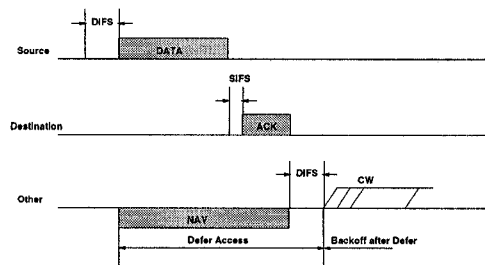


Figure 4: Transmission of MPDU without RTS/CTS.

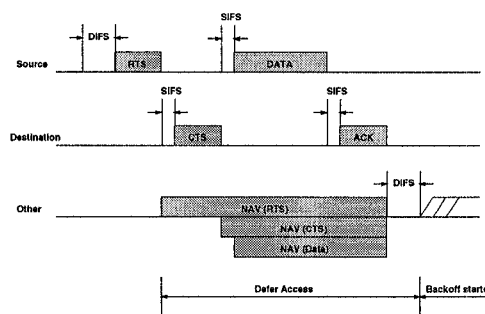


Figure 5: Transmission of MPDU using RTS/CTS.

imum size data frame (2346 octets). Imbedded in the RTS and CTS control frames is a duration value used by listening stations to update their NAVs for the duration of the data exchange. Figure 5 illustrates the transmission of an MPDU using the RTS/CTS mechanism. Stations can choose to never use RTS/CTS, use RTS/CTS whenever the MSDU exceeds the value of `RTS.Threshold` (manageable parameter), or always use RTS/CTS. If a collision occurs with an RTS or CTS MPDU, far less bandwidth is wasted when compared to a large data MPDU. However, for a lightly loaded medium, additional delay is imposed by the overhead of the RTS/CTS frames.

Large MSDUs handed down from the logical link sublayer (LLC) to the MAC may require fragmentation to increase transmission reliability. To determine whether to perform fragmentation, MSDUs are compared to the manageable parameter, `Fragmentation.Threshold`. If the MSDU size exceeds the value of `Fragmentation.Threshold`, then the MSDU is fragmented into multiple MPDUs, which include a header and trailing CRC.

The collision avoidance portion of CSMA/CA is performed through a random backoff procedure. If a station with a frame to transmit initially senses the channel to be busy, then the station waits until the channel becomes idle for a DIFS period and then computes a random backoff time. For IEEE 802.11, time is slotted in time periods corresponding to a `Slot.Time`. Unlike slotted Aloha, where the slot time is equal to the transmission time of one packet, the `Slot.Time` used in IEEE 802.11 is much smaller than an MPDU and is used to define the IFS intervals and determine the backoff time for stations in the CP. The random backoff time is an integer value that corresponds to a

number of time slots. Initially, the station computes a backoff time uniformly in the range 0–7 time slots. After the medium becomes idle after a DIFS period, stations decrement their backoff timer until the medium becomes busy again, or the timer reaches zero. If the timer has not reached zero and the medium becomes busy, the station freezes its timer. When the timer is finally decremented to zero, the station transmits its frame. If two or more stations decrement to zero at the same time, then a collision will occur and each station will have to generate a new backoff time in the range 0–15 time slots. For each retransmission attempt, the backoff time grows exponentially as:

$$\lfloor 2^{2+i} * \text{ranf}() \rfloor * \text{Slot_Time}, \quad (1)$$

where i is the number of consecutive times a station attempts to send a MPDU, $\text{ranf}()$ is a uniform random variate in $(0,1)$, and $\lfloor x \rfloor$ represents the largest integer less than or equal to x . The idle period after a DIFS period is referred to as the contention window (CW). The advantage to this channel access method is that it promotes fairness among stations, but its weakness is that it is not able to support DTBS. Fairness is maintained because each station must contend for the channel after every transmission of an MSDU. All stations have equal probability of gaining access to the channel after each DIFS interval. Time-bounded services typically support applications such as packetized voice or video that must be maintained with a specified delay. With DCF, there is not a mechanism to guarantee delay to stations supporting time-bounded services.

2.3.2 Point Coordination Function

The PCF is an optional capability, which is connection-oriented, and provides contention-free services enabling polled stations to transmit without contending for the channel. The function of the PC is performed by the AP within each BSS. Stations within the BSS that are capable of operating in the CF-period (CFP) are known as CF_Aware stations. The method by which polling tables are maintained and the polling sequence is determined by the PC is left to the implementor.

The CFP repetition interval is used to determine the frequency with which the PCF occurs. Within a repetition interval, a portion of the time is allotted to contention free traffic, and the remainder is provided for contention-based traffic. The CFP repetition interval is initiated by a Beacon frame, where the Beacon frame is transmitted by the AP. One of its primary functions is synchronization and timing. The duration of the CFP repetition interval is a manageable parameter that is always an integral number of Beacon frames. The maximum size of the CFP is determined by the manageable parameter, CFP_Max_Duration . At a minimum, time must be allotted for at least one MPDU to be transmitted during the CP. It is up to the AP to determine how long to operate the CFP during any given repetition interval. If traffic is very light, the AP may shorten the CFP and provide the remainder of the repetition interval for the DCF. The CFP may

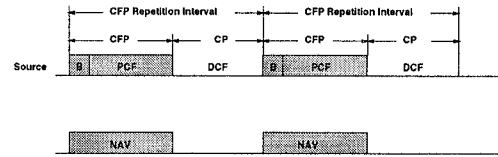


Figure 6: Coexistence of the PCF and DCF.

also be shortened if DCF traffic from the previous repetition interval carries over into the current interval. The maximum amount of delay that can be incurred is the time it takes to transmit an RTS/CTS handshake, maximum MPDU, and an acknowledgment. Figure 6 is a sketch of the CFP repetition interval, illustrating the coexistence of the PCF and DCF.

At the nominal beginning of each CFP repetition interval, all stations in the BSS update their NAV to the maximum length of the CFP (i.e., CFP_Max_Duration). During the CFP, the only time stations are permitted to transmit is in response to a poll from the PC or for transmission of an acknowledgment a SIFS interval after receipt of an MPDU. At the nominal start of the CFP, the PC senses the medium. If the medium remains idle for a PIFS interval, the PC transmits a Beacon frame to initiate the CFP.

3 Simulation Model

Two different simulation models are presented in this paper. The first model represents an ad hoc network, where all stations in the BSS are capable of directly communicating with all other stations in the BSS. All stations in the ad hoc network are assumed to be asynchronous data users. The second model represents an infrastructure network which characterizes a single BSS with an AP. The infrastructure network operates with asynchronous data users in the CP and packetized voice terminals operating in the CFP. Both simulation models are implemented using the physical layer parameters specified in the standard for the DSSS implementation.

Several assumptions have been made to reduce the complexity of the simulation models. A short description of each of the assumptions is provided below: (1) we neglect the effects of propagation delay which is a fairly realistic assumption if transmission distances are on the order of 100 feet between stations; (2) the “hidden terminal” problem is not addressed in the simulation models; (3) the basic rate of 1 Mbps was simulated for the DSSS; (4) no stations operate in the “power-saving” mode (PS-Mode); (5) no interference is considered from nearby BSSs reusing the same DSSS spreading sequence; (6) when the PCF and DCF coexist together in the infrastructure network, all stations operating during the CP are asynchronous data users, and all users operating during the CFP are packetized voice users; (7) a finite transmit buffer is maintained for each station. If the finite buffer fills, all newly generated MSDUs will be considered dropped without returning.

For the ad hoc and infrastructure network simulations, a burst error model is introduced to character-

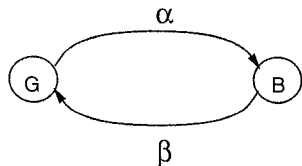


Figure 7: Markov chain representing the burst error model

ize fading in the communications channel. A 2-state continuous-time Markov chain is used to represent the burst error model, as shown in Figure 7. State *G* represents the channel in a ‘good’ state. This indicates that the channel is operating with a very low bit error rate (denoted by `Good_BER`). State *B* indicates the channel is operating in a fading condition with a higher error rate denoted by `Bad_BER`. The transition rate from the ‘good’ state to the ‘bad’ state is denoted by α , while the transition rate from the ‘bad’ state to the ‘good’ state is denoted by β . A frame is considered to be corrupted if it contains one or more bit errors.

3.1 Ad Hoc Network Model

With the ad hoc network model, all users are assumed to be asynchronous data users and they shall operate in a self-contained BSS. The arrival of frames from a station’s higher layer protocol to the MAC sublayer is modeled with exponential inter-arrival times and a truncated geometric distribution for the frame lengths. The truncated geometric distribution is used to ensure that the MSDU does not exceed the maximum length established by the specification (i.e., 2312 octets). However, the simulation model can easily accommodate other arrival processes and frame length distributions.

During the simulation, if collisions or bit errors affect the transmission of a frame, retransmission will occur according to the backoff procedure described previously. The number of retransmissions is limited before the frame is dropped from the station’s transmit queue. In the case of MSDUs smaller than `RTS_Threshold`, the number of retransmissions is limited to `Short_Retry_Limit`. For MSDUs larger than `RTS_Threshold`, the maximum number of retransmissions is set by `Long_Retry_Limit`. The number of retransmissions is extended in this case since short RTS frames are not as wasteful of bandwidth as larger data payloads. Table 1 lists the default values for various attributes that are used for the simulation of the ad hoc network.

3.2 Infrastructure Network Model

We simulate the effect of a single BSS with an AP, where asynchronous data users transmit during the CP and packetized voice users transmit during the CFP. The coexistence of the DCF and PCF is illustrated in Figure 6, where for the purposes of this simulation, the value of `CFP_Max_Duration` is provided in Table 2. The duration of the `CFP_Repetition_Interval` is approximately 0.4096 seconds, therefore approximately 95% of the repetition interval can be allocated by the AP for contention-free services.

Table 1: Default attribute values for the ad hoc network, unless otherwise specified.

Attribute	Typical Value
Channel Rate	1 Mbps
Good_BER	10^{-10}
α	30.0 sec^{-1}
β	10.0 sec^{-1}
Fragmentation_Threshold	800 octets
RTS_Threshold	250 octets
Short_Retry_Limit	5
Long_Retry_Limit	7
DSSS Preamble	144 bits
DSSS Header	48 bits
Station Buffer Size	300 frames
Slot_Time	$20 \mu\text{s}$
SIFS_Time	$10 \mu\text{s}$
DIFS_Time	$50 \mu\text{s}$

During the CFP, if a station is polled by the AP to transmit, the station can transmit directly to another station in the BSS or to a station in another BSS. When the transmission is directed to a station in another BSS, the source station transmits the frame to the AP, who is responsible for forwarding the frame through the DS to the remote AP servicing the destination station. Since the BSS is assumed to be fairly small, it would be unrealistic to establish voice connections over a small distance. We thus assume that all voice traffic in the simulation will be forwarded to the AP for transmission into the DS.

The polling scheme during the CFP uses a ‘round-robin’ algorithm, where each station is polled sequentially in the order in which it is placed in the polling list. When the CFP ends, the AP keeps track of the location in the polling list where it stopped, and resumes polling at that same point when the CFP starts. Since all stations operating during the CFP are packetized voice users, they all have the same QoS requirements, and therefore priority polling mechanisms are not required. We propose a simple polling scheme to allocate unused bandwidth to currently active voice users. When the AP is prepared to poll a station during the CFP, if the AP has an MPDU queued for transmission to the station, the poll and MPDU can be combined and transmitted as a single frame. Otherwise, the AP sends a sole CF-Poll (no Data) to the station. If the AP sends k consecutive CF-Polls to a station, and the station responds each time without any payload to transmit *Null Function*, then the station is dropped from the polling list for that `CFP_Repetition_Interval`. During the next interval, the station will be added back into the polling list and the process will start over. The polling scheme will drop stations that are not active either transmitting or receiving voice packets. When all voice stations have been dropped from the polling list, the AP will send a `CF_END` frame indicating that the asynchronous users can start using the channel until the start of the next CFP interval.

The voice stream is modeled using an ON/OFF process, where stations are either transmitting (ON) or

listening (OFF). The amount of time sitting in the OFF or ON state is exponentially distributed, where the mean value of the silence (OFF) period is 1.35 seconds, and the mean value of the talkspurt (ON) period is 1 second. The voice transmission rate in the ON state is 64 kbps. The length of the voice payload should be chosen so that voice packetization delay is minimized and header overhead is not large, which is a conflicting goal. No retransmissions will be performed for voice frames since this traffic is delay sensitive. Quality of service parameters for voice typically limit maximum delay to 25 ms without echo canceling, and 500 ms using echo canceling. Asynchronous data frames are transmitted in the CP portion of the repetition interval in the same manner as described in paragraph 3.1. Table 2 lists the additional default values that are used for the simulation of the infrastructure network.

Table 2: Default attribute values for the infrastructure network, unless otherwise specified.

Attribute	Typical Value
No. of Voice Stations	10
Voice Transmission Rate	64 kbps
Voice Station Buffer Size	100 frames
CFP_Max_Duration	0.39 sec
CFP_Repetition_Interval	0.41 sec
PIFS_Time	30 μ s

4 Simulation Results

In this section, we present the simulation results for both the ad hoc and infrastructure networks. Our main interest is to study the end-to-end performance at the MAC sublayer and identify crucial operating points. Whenever applicable, the plots are provided with 95% confidence intervals.

4.1 Ad Hoc Network

For the ad-hoc network, we assume all mobile stations generate asynchronous data traffic with the same intensity. Figure 8 shows the aggregate data throughput in Mbps versus the offered load in Mbps for several cases of bit error rates (i.e., the Bad_BER). The offered load is defined to be the average number of bits per second passed down to the MAC sublayer at the source. The throughput is the average number of bits per second passed up from the MAC sublayer at the destination. The average length of the MSDU, l_{data} , is taken to be 1000 octets long and the number of stations, N_{data} , is fixed to 10. Other parameters are given in Table 1.

When the medium is relatively clean (Bad_BER is less than 10^{-6}), a maximum throughput of 75% and above is possible. However, the maximum throughput can drop to below 20% under harsh fading. Thus, it is clear that channel condition can adversely affect the throughput performance of the 802.11 system.

Figure 9 shows the effect of the average MSDU length on the throughput performance. This result is generated for the case when the Bad_BER is equal to 10^{-5} , and the same value for the number of stations

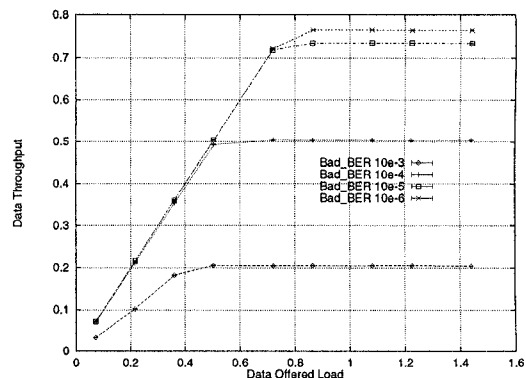


Figure 8: Effect of Bad_BER on data throughput.

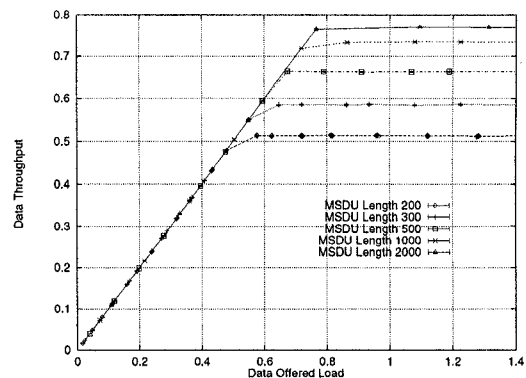


Figure 9: Effect of l_{data} on data throughput.

as above. The IEEE 802.11 MAC and PHY layers add a total of 58 octets for overhead. As confirmed by the figure, the longer the MSDU is, the more efficient the system becomes. However, when the MSDU is longer than 1000 octets, the throughput improvement becomes marginal.

Given an MSDU, it is of interest to decide whether fragmentation is needed. If we decide to fragment a small MSDU, we may incur a large overhead. On the other hand, deferring fragmentation to very large MSDUs may waste more bandwidth due to transmission errors that are more likely to occur in large MSDUs.

The optimal Fragmentation_Threshold with respect to Bad_BER is depicted in Figure 10. Here, 'optimal' refers to the operating point that gives maximum throughput. The plots are obtained by using $N_{data} = 10$ and $l_{data} = 2312$ octets long which is the maximum possible payload length. From the figure, we can conclude that larger Fragmentation_Threshold should be applied when the channel is cleaner. In general, a Fragmentation_Threshold of 500-1000 octets seems to be the best choice. Fragmentation_Threshold beyond 1000 octets gives no significant gain in throughput for low BER.

The original MSDU length is a driving factor that determines whether the exchange of RTS and CTS control packets should be used. The RTS_Threshold manageable parameter should be set at an optimum value that will enable maximum throughput.

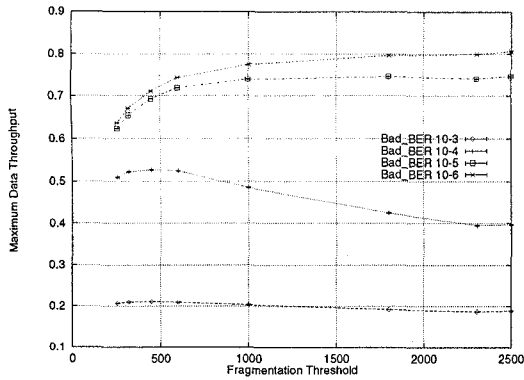


Figure 10: Optimal fragmentation threshold.

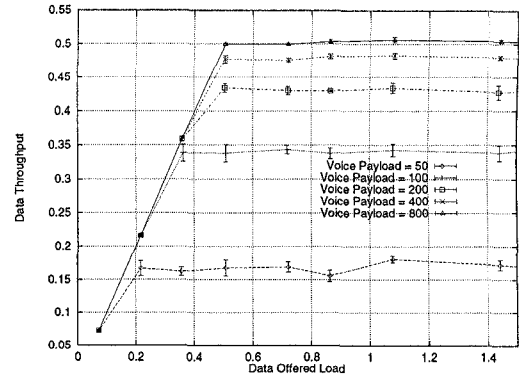


Figure 12: Effect of l_{voice} on data throughput.

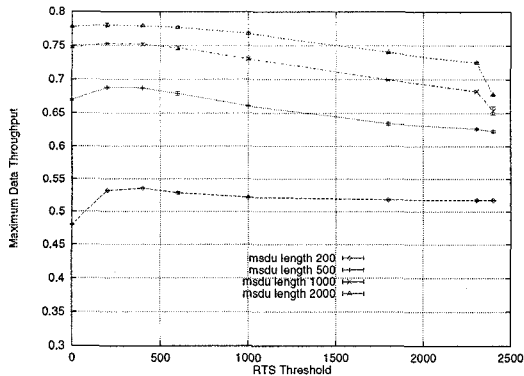


Figure 11: Optimal RTS threshold.

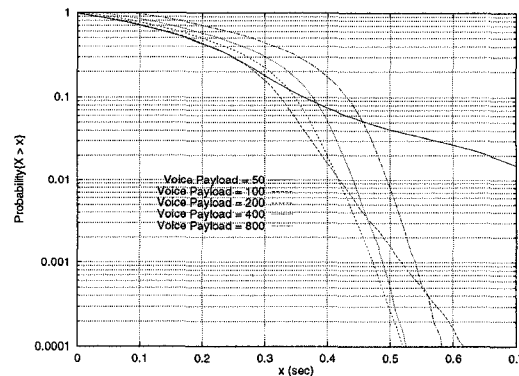


Figure 13: Complementary cumulative distribution for voice delay.

Figure 11 illustrates the impact on maximum data throughput for various values of MSDU length when the RTS_Threshold value is varied. The impact of using RTS_Threshold is more pronounced as the MSDU length increases. The optimum value for RTS_Threshold is in the approximate range of 250–500 octets. As the MSDU length increases, the sensitivity of the value selected for the RTS threshold becomes more pronounced. This is to be expected since large packets have a much larger impact on aggregate throughput when they are involved in collisions and must be retransmitted.

4.2 Infrastructure Network

We now turn our attention to an infrastructure network supporting voice and data traffic. Data traffic is transported through the CFP and voice traffic through the CP. For all the cases below, $l_{data} = 1000$ octets, $N_{data} = 10$, and $Bad_BER = 10^{-5}$. With 10 voice stations, the aggregate voice throughput is approximately 272 Kbps

We first investigate the effect of voice payload length, l_{voice} , on both data and voice performances. We set the number of voice stations, N_{voice} , to 5 pairs. The first 5 voice stations are located in the BSS. The others are located elsewhere and are generated through the AP. For voice traffic, we only consider the delay between an AP and another mobile station in the same BSS. All measurements are done

at the MSAP (MAC Service Access Point).

Figure 12 displays the influence of the voice payload length on data traffic performance. As we can see from figure, data traffic will suffer more as the voice payload length is decreased. This is because the voice stations will incur more overhead, making less available bandwidth for data traffic. Thus, from the point of view of data traffic, the voice payload should be made very long.

Let us now consider the voice delay performance as shown in Figure 13. The random variable X denotes the end-to-end delay between an AP and a mobile station. Here the delay is measured from the time the first bit is generated at the transmitter until the time the last bit is received at the receiver. The figure illustrates the complementary cumulative distribution, $P\{X > x\}$, for voice delay in seconds. As discussed previously, voice delay can tolerate as much as 0.5 seconds if an echo canceler is used. Without an echo canceler, a much more stringent voice delay (under 25 ms) must be satisfied. It is obvious from the figure that an echo canceler must be used since a large proportion of the voice traffic exceeds the 25-ms requirement in delay. This is mainly due to the fact that voice packets would have to wait during the CP which may last for approximately 0.2 s. Thus, we assume that an echo canceler is employed and that voice packets that are delayed by more than 0.5 seconds at the receiver

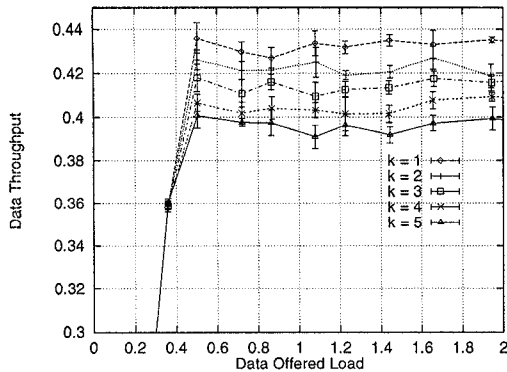


Figure 14: Data throughput versus offered load for several values of k .

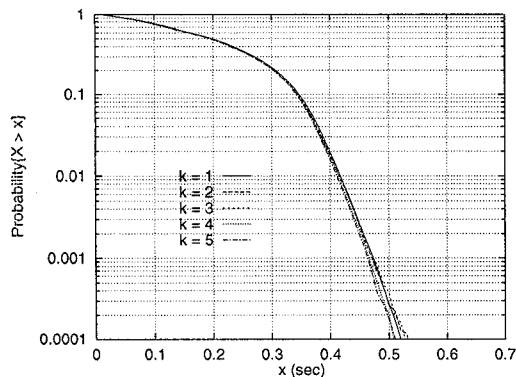


Figure 15: Effect of k on voice delay.

become useless and have to be discarded. Thus, the performance measure of interest for voice traffic is the probability that a voice packet will be discarded due to its late arrival at the receiver. Shorter payloads incur larger overheads, translating into longer delays. On the other extreme, longer payloads imply longer packetization delays. Thus, these two parameters must be traded-off. As we can see from Figure 13, the best operating points appear to be around 100-400 octets long for voice payload, where less than 0.2% of voice packets would have to be discarded.

As mentioned previously, an AP drops a station from the polling list if the station does not transmit and receive any data for k consecutive polls in the current CFP interval. To see the appropriate values of k , we plot the effect of k on data throughput and voice delay, illustrated in Figure 14 and Figure 15, respectively.

As can be seen from Figure 14, reducing k increases the throughput for data traffic. This is intuitively clear since reducing k implies that voice stations will be dropped from the polling list sooner, thereby giving more bandwidth to data stations. A seemingly surprising result is illustrated in Figure 15 which shows that the performance of voice stations is almost unaffected. The reason is due to the fact that voice stations operate on an ON/OFF basis. When a voice station does not have any data to send during the OFF pe-

riod, it is likely that it will not have any data to send in the near future. Thus, when a communicating pair of voice buffers are empty, the best policy is to drop the stations from the polling list immediately ($k = 1$).

5 Conclusion

The general conclusions derived from our study are:

- The efficiency delivered by the DCF is reasonably high if the average MSDU length is longer than 500 octets, the Fragmentation_Threshold is set to 500-1000 octets, the RTS_Threshold is set to 250-500 octets, and the medium is relatively clean (BER better than 10^{-6}).
- Real-time services such as packet voice can be transported by the PCF. However, packet-voice systems must employ an echo canceler since the end-to-end delay cannot be bounded under 25 ms.
- Compromised performance for both data and voice traffic is achieved when the voice payload length is between 100 to 400 octets long.
- When a voice station does not have any data to receive and transmit during a poll, the station should be dropped from the list immediately so that the remaining bandwidth can be allocated to other stations.

We are currently investigating the effect of hidden terminals and station mobility on the performance of the overall system. The results of these studies will be reported elsewhere.

References

- [1] K.C. Chen, "Medium access control of wireless LANs for mobile computing," *IEEE Network*, vol. 8, no. 5, pp. 50-63, Sep. 1994.
- [2] H.S. Chhaya and S. Gupta, "Throughput and fairness properties of asynchronous data transfer methods in the IEEE 802.11 MAC protocol," *Personal, Indoor, Mobile, and Radio Communication Conference*, pp. 613-617, 1995.
- [3] B. P. Crow, *Performance Evaluation of the IEEE 802.11 Wireless Local Area Network Protocol*, M.S. Thesis, University of Arizona, 1996.
- [4] W. Diepstraten, "A wireless MAC protocol comparison," IEEE P802.11-92/51.
- [5] R.O. LaMaire, A. Krishna, P. Bhagwat, and J. Panian, "Wireless LANs and mobile networking: standards and future directions," *IEEE Commun. Mag.*, vol. 34, no. 8, pp. 86-94, Aug. 1996.
- [6] J. Weinmiller, H. Woesner, and A. Wolisz, "Analyzing and improving the IEEE 802.11-MAC protocol for wireless LANs," *MASCOTS '96*, San Jose, USA, pp. 200-206, Feb. 1996.
- [7] Wireless Medium Access Control and Physical Layer Working Group, P802.11 IEEE Draft Standard-Wireless Lan, (IEEE Standards Department, D3, Jan. 1996).