ABSTRACT

The draft IEEE 802.11 Wireless Local Area Network (WLAN) specification is approaching completion. In this article, the IEEE 802.11 protocol is explained, with particular emphasis on the medium access control sublayer. Performance results are provided for packetized data and a combination of packetized data and voice over the WLAN. Our performance investigation reveals that an IEEE 802.11 network may be able to carry traffic with time-bounded requirements using the point coordination function. However, our findings suggest that packetized voice traffic must be handled in conjunction with an echo canceler.

IEEE 802.11 Wireless Local Area Networks

Brian P. Crow, The MITRE Corporation

Indra Widjaja, Fujitsu Network Communications

Jeong Geun Kim, University of Arizona

Prescott T. Sakai, Cypress Semiconductor

Wireless computing is a rapidly emerging technology providing users with network connectivity without being tethered off of a wired network. Wireless local area networks (WLANs), like their wired counterparts, are being developed to provide high bandwidth to users in a limited geographical area. WLANs are being studied as an alternative to the high installation and maintenance costs incurred by traditional additions, deletions, and changes experienced in wired LAN infrastructures. Physical and environmental necessity is another driving factor in favor of WLANs. Typically, new building architectures are planned with network connectivity factored into the building requirements. However, users inhabiting existing buildings may find it infeasible to retrofit existing structures for wired network access. Examples of structures that are very difficult to wire include concrete buildings, trading floors, manufacturing facilities, warehouses, and historical buildings. Lastly, the operational environment may not accommodate a wired network, or the network may be temporary and operational for a very short time, making the installation of a wired network impractical. Examples where this is true include ad hoc networking needs such as conference registration centers, campus classrooms, emergency relief centers, and tactical military environments.

Ideally, users of wireless networks will want the same services and capabilities that they have commonly come to expect with wired networks. However, to meet these objectives, the wireless community faces certain challenges and constraints that are not imposed on their wired counterparts.

Frequency Allocation — Operation of a wireless network requires that all users operate on a common frequency band. Frequency bands for particular uses must typically be approved and licensed in each country, which is a time-consuming process due to the high demand for available radio spectrum.

Interference and Reliability — Interference in wireless communications can be caused by simultaneous transmissions (i.e., collisions) by two or more sources sharing the same frequency band. Collisions are typically the result of multiple stations waiting for the channel to become idle and then beginning transmission at the same time. Collisions are also caused by the "hidden terminal" problem, where a station, believing the channel is idle, begins transmission without successfully detecting the presence of a transmission already in progress. Interference is also caused by multipath fading, which is characterized by random amplitude and phase fluctuations at the receiver. The reliability of the communications channel is typically measured by the average bit error rate (BER). For packetized voice, packet loss rates on the order of 10^{-2} are generally acceptable; for uncoded data, a BER of 10^{-5} is regarded as acceptable. Automatic repeat request (ARQ) and forward error correction (FEC) are used to increase reliability.

Security — In a wired network, the transmission medium can be physically secured, and access to the network is easily controlled. A wireless network is more difficult to secure, since the transmission medium is open to anyone within the geographical range of a transmitter. Data privacy is usually accomplished over a radio medium using encryption. While encryption of wireless traffic can be achieved, it is usually at the expense of increased cost and decreased performance.

Power Consumption — Typically, devices connected to a wired network are powered by the local 110 V commercial power provided in a building. Wireless devices, however, are meant to be portable and/or mobile, and are typically battery powered. Therefore, devices must be designed to be very energy-effi-

The views and opinions expressed in this article are those of the authors and do not reflect MITRE's or Fujitsu Network Communications' current position.

cient, resulting in "sleep" modes and low-power displays, causing users to make cost versus performance and cost versus capability trade-offs.

Human Safety — Research is ongoing to determine whether radio frequency (RF) transmissions from radio and cellular phones are linked to human illness. Networks should be designed to minimize the power transmitted by network devices. For infrared (IR) WLAN systems, optical transmitters must be designed to prevent vision impairment.

Mobility — Unlike wired terminals, which are static when operating on the network, one of the primary advantages of wireless terminals is freedom of mobility. Therefore, system designs must accommodate handoff between transmission boundaries and route traffic to mobile users.

Throughput — The capacity of WLANs should ideally approach that of their wired counterparts. However, due to physical limitations and limited available bandwidth, WLANs are currently targeted to operate at data rates between 1–20 Mb/s. To support multiple transmissions simultaneously, spread spectrum techniques are frequently employed.

Currently, there are two emerging WLAN standards: the European Telecommunications Standards Institute (ETSI) High-Performance European Radio LAN (HIPERLAN) and the IEEE 802.11 WLAN. Both draft standards cover the physical layer and medium access control (MAC) sublayer of the open systems interconnection (OSI) seven-layer reference model. The HIPERLAN committee has identified the 5.15–5.30 GHz and 17.1–17.2 GHz bands for transmission. The 5 GHz band has been ratified for HIPERLAN use by the Conference of European Postal and Telecommunications Administrations (CEPT). Data rates up to 23.529 Mb/s are projected, and multihop routing, time-bounded services, and power-saving features are expected. For further information regarding HIPERLAN, see the article by LaMaire *et al.* [1] or the HIPERLAN specification [2].

The IEEE is developing an international WLAN standard identified as IEEE 802.11 [3]. This project was initiated in 1990, and several draft standards have been published for review. The scope of the standard is "to develop a Medium Access Control (MAC) and Physical Layer (PHY) specification for wireless connectivity for fixed, portable and moving stations within a local area." The purpose of the standard is twofold:

- "To provide wireless connectivity to automatic machinery, equipment, or stations that require rapid deployment, which may be portable, or hand-held or which may be mounted on moving vehicles within a local area"
- "To offer a standard for use by regulatory bodies to standardize access to one or more frequency bands for the purpose of local area communication" [3].

The IEEE 802.11 draft standard describes mandatory support for a 1 Mb/s WLAN with optional support for a 2 Mb/s data transmission rate. Mandatory support for asynchronous data transfer is specified as well as optional support for distributed time-bounded services (DTBS). Asynchronous data transfer refers to traffic that is relatively insensitive to time delay. Examples of asynchronous data are available bit rate traffic like electronic mail and file transfers. Time-bounded traffic, on the other hand, is traffic that is bounded by specified time delays to achieve an acceptable quality of service (QoS) (e.g., packetized voice and video).

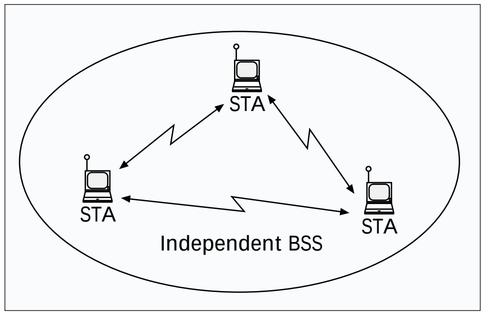


Figure 1. *Sketch of an ad hoc network.*

Of particular interest in the specification is the support for two fundamentally different MAC schemes to transport asynchronous and time-bounded services. The first scheme, distributed coordination function (DCF), is similar to traditional legacy packet networks supporting best-effort delivery of the data. The DCF is designed for asynchronous data transport, where all users with data to transmit have an equally fair chance of accessing the network. The point coordination function (PCF) is the second MAC scheme. The PCF is based on polling that is controlled by an access point (AP). The PCF is primarily designed for the transmission of delay-sensitive traffic. While the DCF has been studied by several researchers [4–7], the combined performance of the DCF and PCF operating in a common repetition interval is much less understood. In this article, the performance of an ad hoc network (DCF-only) and an infrastructure network (DCF and PCF) are investigated by means of simulation. We also investigate the effect of channel errors on the performances of PCF and DCF, which is absent in all previous studies. 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.

In the remainder of the article, we will summarize the IEEE 802.11 WLAN specification (emphasis on the MAC sublayer), briefly describe the simulation model which supports asynchronous data and packetized voice traffic, and provide performance results from the simulation.

Description of the IEEE 802.11 Draft Standard

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 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. However, transmission medium degradations due to multipath fading, or interference from nearby BSSs reusing the same physical-layer characteristics (e.g., frequency and spreading code, or hopping pattern), can cause some stations to appear "hidden" from other stations.

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 *independent BSS* (IBSS), which is the formal name of an ad hoc network in the IEEE 802.11 standard. Any station can establish a direct communications session with any other station in the BSS, without the requirement of channeling all 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 has the appearance of one large BSS to the *logical link control* (LLC) sublayer of each station (STA). 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 (MSDUs). The DS, as specified by IEEE 802.11, is implementation-independent. Therefore, the DS could be a wired IEEE 802.3 Ethernet LAN, IEEE 802.4 token bus LAN, IEEE 802.5 token ring LAN, fiber distributed data interface (FDDI) metropolitan area network (MAN), or another IEEE 802.11 wireless medium. Note that while the DS could physically be the same transmission medium as the BSS, they are logically different, because the DS is solely used as a transport backbone to transfer packets between different BSSs in the ESS.

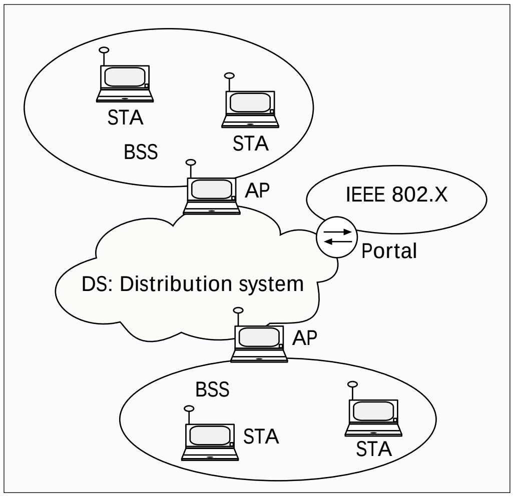


Figure 2. *Sketch of an infrastructure network.*

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. If the network is an IEEE 802.X, the portal incorporates functions which are analogous to a bridge; that is, it provides range extension and the translation between different frame formats. Figure 2 illustrates a simple ESS developed with two BSSs, a DS, and a portal access to a wired LAN.

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 IR. The FHSS utilizes the 2.4 GHz Industrial, Scientific, and Medical (ISM) band (i.e., 2.4000–2.4835 GHz). In the United States, a maximum of 79 channels are specified in the hopping set. The first channel has a center frequency of 2.402 GHz, and all subsequent channels are spaced 1 MHz apart. The 1 MHz separation is mandated by the FCC for the 2.4 GHz ISM band. The channel separation corresponds to 1 Mb/s of instantaneous bandwidth. Three different hopping sequence sets are established with 26 hopping sequences per set. Different hopping sequences enable multiple BSSs to coexist in the same geographical area, which may become important to alleviate congestion and maximize the total throughput in a single BSS. The reason for having three different sets is to avoid prolonged collision periods between different hopping sequences in a set [3]. The minimum hop rate permitted is 2.5 hops/s. The basic access rate of 1 Mb/s uses two-level Gaussian frequency shift keying (GFSK), where a logical 1 is encoded using frequency $F_c + f$ and a logical 0 using frequency $F_c - f$. The enhanced access rate of 2 Mb/s uses four-level GFSK, where 2 bits are encoded at a time using four frequencies.

The DSSS also uses the 2.4 GHz ISM frequency band, where the 1 Mb/s basic rate is encoded using differential binary phase shift keying (DBPSK), and a 2 Mb/s enhanced rate uses differential quadrature phase shift keying (DQPSK). The spreading is done by dividing the available bandwidth into 11 subchannels, each 11 MHz wide, and using an 11-chip Barker sequence to spread each data symbol. The maximum channel capacity is therefore (11 chips/symbol)/(11 MHz) = 1 Mb/s if DBPSK is used [8]. Overlapping and adjacent BSSs can be accommodated by ensuring that the center frequencies of each BSS are separated by at least 30 MHz [3]. This rigid requirement will enable only two overlapping or adjacent BSSs to operate without interference.

The IR specification identifies a wavelength range from 850 to 950 nm. The IR band is designed for indoor use only and operates with nondirected transmissions. The IR specification was designed to enable stations to receive line-of-site and reflected transmissions. Encoding of the basic access rate of 1 Mb/s is performed using 16-pulse position modulation (PPM), where 4 data bits are mapped to 16 coded bits for transmission. The enhanced access rate (2 Mb/s) is performed using 4-PPM modulation, where 2 data bits are mapped to 4 coded bits for transmission.

MEDIUM ACCESS CONTROL 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 access to 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, 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. The stan-

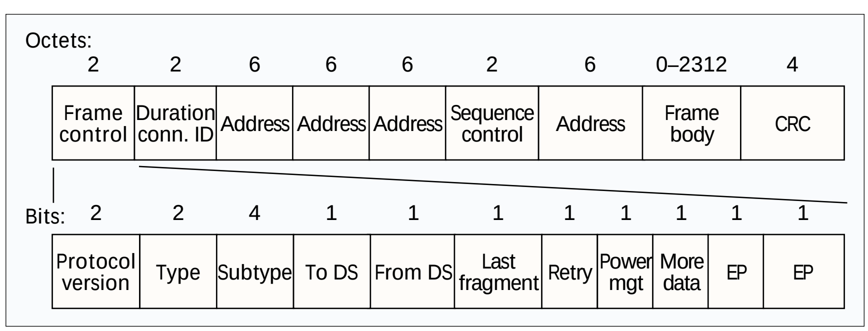


Figure 3. *Standard IEEE 802.11 frame format.*

dard IEEE 802.11 frame format is illustrated in Fig. 3. Note that the frame body (MSDU) is a variable-length field consisting of the data payload and 7 octets for encryption/decryption if the optional Wired Equivalent Privacy (WEP) protocol is implemented. The IEEE standard 48-bit MAC addressing is used to identify a station. The 2 duration octets indicate the time (in microseconds) the channel will be allocated for successful transmission of a MAC protocol data unit (MPDU). The type bits identify the frame as either control, management, or data. The subtype bits further identify the type of frame (e.g., Clear to Send control frame). A 32-bit cyclic redundancy check (CRC) is used for error detection.

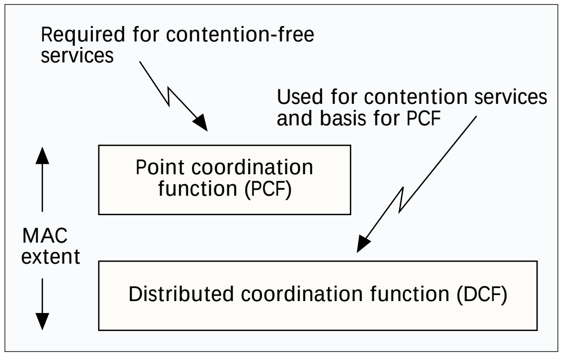


Figure 4. *MAC architecture.*

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 DCF operates solely in the ad hoc network, and either operates solely or coexists with the PCF in an infrastructure network. The MAC architecture is depicted in Fig. 4, where it is shown that 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 access to the channel and, once the MSDU is transmitted, must recontend for access to the channel for all subsequent frames. Contention services promote fair access to the channel for all stations.

The DCF is based on carrier sense multiple access with collision avoidance (CSMA/CA). CSMA/CD (collision detection) is not used because a 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.

A source station performs virtual carrier sensing by sending MPDU duration information in the header of request to send (RTS), clear to send (CTS), and data frames. 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 duration field indicates the amount of time (in microseconds) after the end of the present frame the channel will be utilized to complete the successful transmission of the data or management frame. Stations in the BSS use the information in the duration field to 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 mechanisms indicate the channel is busy.

Priority access to the wireless medium is controlled through the use of interframe 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 DCF-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; therefore, SIFS has the highest-priority access to the communications medium. For the basic access method, when a station senses the channel is idle, the station waits for a DIFS period and samples the channel again. If the channel is still idle, the station transmits an MPDU. The receiving station calculates the checksum and determines whether the packet was received correctly. Upon receipt of a correct packet, the receiving station waits a SIFS interval and transmits a positive acknowledgment frame (ACK) back to the source station, indicating that the transmission was successful. Figure 5 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 ACK following the data frame.

Since a source station in a BSS cannot hear its own transmissions, when a collision occurs, the source continues transmitting the complete MPDU. If the MPDU is large (e.g., 2300 octets), a lot of channel bandwidth is wasted due to a corrupt MPDU. RTS and CTS control frames can be used by a station to reserve channel bandwidth prior to the transmission of an MPDU and to minimize the amount of bandwidth wasted when collisions occur. RTS and CTS control frames are relatively small (RTS is 20 octets and CTS is 14 octets) when compared to the maximum data frame size (2346 octets). The RTS control frame is first transmitted by the source station (after successfully contending for the channel) with a data or management frame queued for transmission to a specified destination station. All stations in the BSS, hearing the RTS packet, read the duration field (Fig. 3) and set their NAVs accordingly. The destination station responds to the RTS packet with a CTS packet after an SIFS idle period has elapsed. Stations hearing the CTS packet look at the duration field and again update their NAV. Upon successful reception of the CTS, the source station is virtually assured that the medium is stable and reserved for successful transmission of the MPDU. Note that stations are capable of

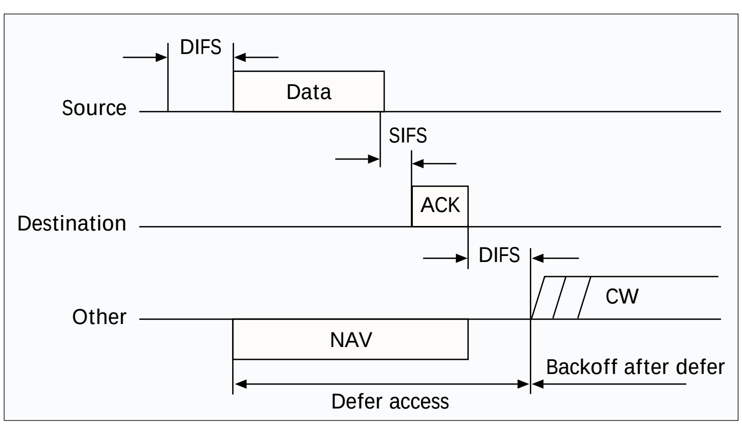


Figure 5. *Transmission of an MPDU without RTS/CTS.*

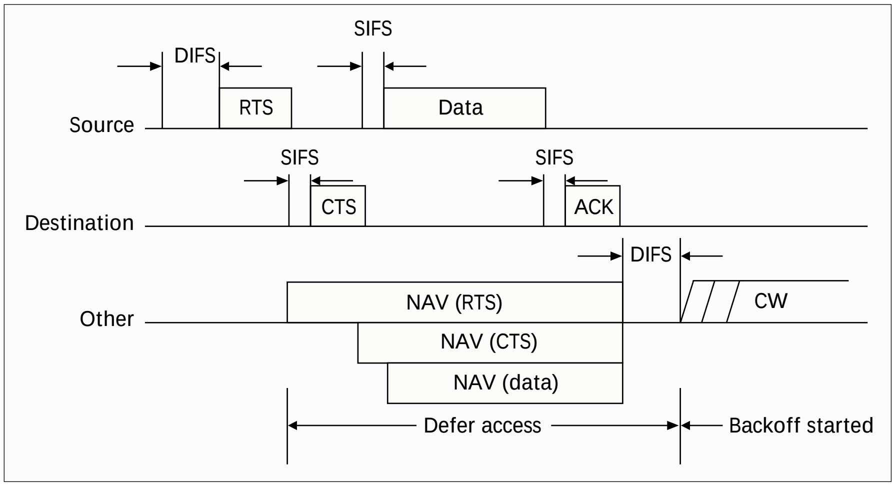


Figure 6. *Transmission of an MPDU using RTS/CTS.*

updating their NAVs based on the RTS from the source station and CTS from the destination station, which helps to combat the "hidden terminal" problem. Figure 6 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 LLC to the MAC may require fragmentation to increase transmission reliability. To determine whether to perform fragmentation, MPDUs are compared to the manageable parameter Fragmentation_Threshold. If the MPDU size exceeds the value of Fragmentation_Threshold, the MSDU is broken into multiple fragments. The resulting MPDUs are of size Fragmentation_Threshold, with exception of the last MPDU, which is of variable size not to exceed Fragmentation_Threshold. When an MSDU is fragmented, all fragments are transmitted sequentially (Fig. 7). The channel is not released until the complete MSDU has been transmitted successfully, or the source station fails to receive an acknowledgment for a transmitted fragment. The destination station positively acknowledges each successfully received fragment by sending a DCF ACK back to the source station. The source station maintains control of the channel throughout the transmission of the MSDU by waiting only an SIFS period after receiving an ACK and transmitting the next fragment. When an ACK is not received for a previously transmitted frame, the source station halts transmission and recontends for the channel. Upon gaining access to the channel, the source starts transmitting with the last unacknowledged fragment.

If RTS and CTS are used, only the first fragment is sent using the handshaking mechanism. The duration value of RTS and CTS only accounts for the transmission of the first fragment through the receipt of its ACK. Stations in the BSS thereafter maintain their NAV by extracting the duration information from all subsequent fragments.

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 that correspond 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 Slot_Time is different for each physical layer implementation. The random backoff time is an integer value that corresponds to a number of time slots. Initially, the station computes a backoff time in the range 0–7. 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, a collision will occur, and each station will have to generate a new backoff time in the range 0–15. For each retransmission attempt, the backoff time grows as $\lfloor 2^{2+i} \cdot ranf() \rfloor \cdot$ Slot_Time, where i is the number of consecutive times a station attempts to send an MPDU, *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 of this channel access method is that it promotes fairness among stations, but its weakness is that it probably could not support DTBS. Fairness is maintained because each station must recontend 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 minimum delay. With DCF, there is no mechanism to guarantee minimum delay to stations supporting time-bounded services.

Point Coordination Function (PCF)

The PCF is an optional capability, which is connection-oriented, and provides contention-free (CF) frame transfer. The PCF relies on the point coordinator (PC) to perform polling, enabling polled stations to transmit without contending for

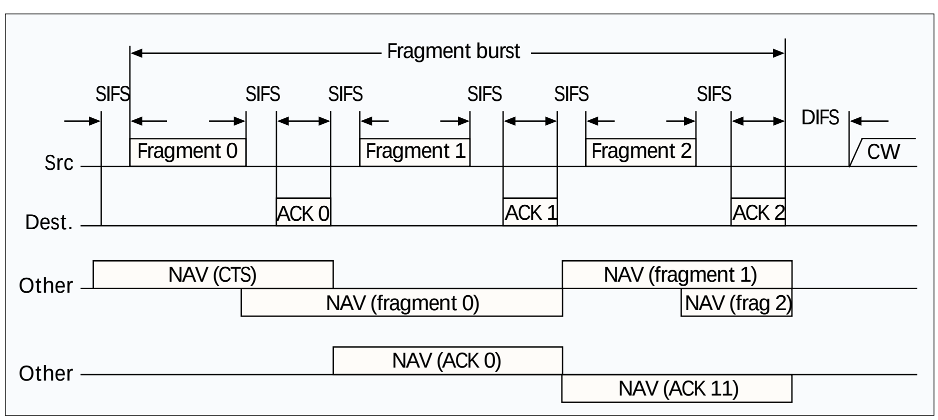
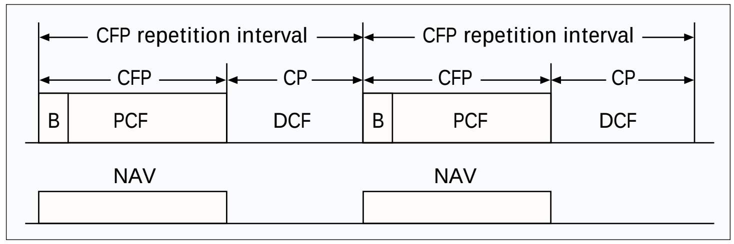


Figure 7. *Transmission of a fragmented MPDU.*

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, is left to the implementor.

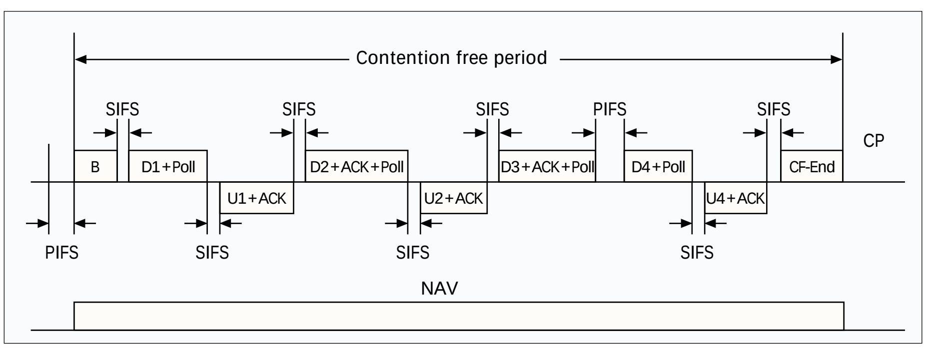
The PCF is required to coexist with the DCF and logically sits on top of the DCF (Fig. 4). The CFP repetition interval (CFP_Rate) is used to determine the frequency with which the PCF occurs. Within a repetition interval, a portion of the time is allotted to contention-free traffic, and the remainder 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. Once the CFP_Rate is established, the duration of the CFP is determined. The maximum size of the CFP is determined by the manageable parameter CFP_Max_Duration. The minimum value of CFP_Max_Duration is the time required to transmit two maximum-size MPDUs, including overhead, the initial beacon frame, and a CF-End frame. The maximum value of CFP_Max_Duration is the CFP repetition interval minus the time required to successfully transmit a maximum-size MPDU during the CP (which includes the time for RTS/CTS handshaking and the ACK). Therefore, 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 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 ACK. Figure 8 is a sketch of the CFP repetition interval, illustrating the coexistence of the PCF and DCF.



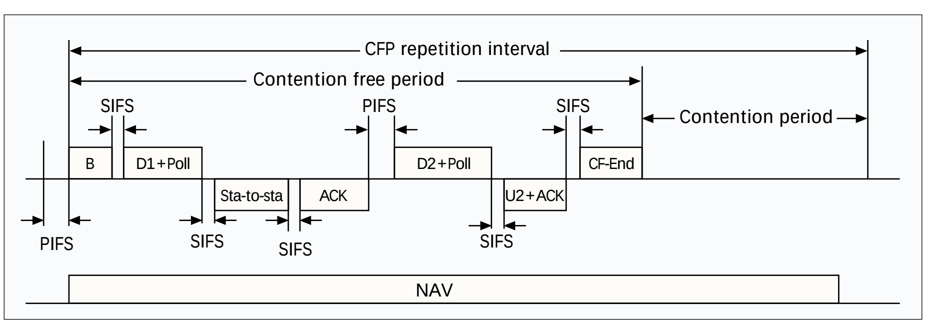
■ **Figure 8.** *Coexistence of the PCF and DCF.*

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 ACK 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. The PC starts CF transmission a SIFS interval after the beacon frame is transmitted by sending a CF-Poll (no data), Data, or Data + CF-Poll frame. The PC can immediately terminate the CFP by transmitting a CF-End frame, which is common if the network is lightly loaded and the PC has no traffic buffered. If a CF-aware station receives a CF-Poll (no data) frame from the PC, the STA can respond to the PC after a SIFS idle period, with a CF-ACK (no data) or a Data + CF-ACK frame. If the PC receives a Data + CF-Ack frame from a station, the PC can send a Data + CF-ACK + CF-Poll frame to a different station, where the CF-ACK portion of the frame is used to acknowledge receipt of the previous data frame. The ability to combine polling and acknowledgment frames with data frames, transmitted between stations and the PC, was designed to improve efficiency. If the PC transmits a CF-Poll (no data) frame and the destination station does not have a data frame to transmit, the station sends a Null Function (no data) frame back to the PC. Figure 9 illustrates the transmission of frames between the PC and a station, and vice versa. If the PC fails to receive an ACK for a transmitted data frame, the PC waits a PIFS interval and continues transmitting to the next station in the polling list.

After receiving the poll from the PC, as described above, the station may choose to transmit a frame to another station in the BSS. When the destination station receives the frame, a DCF ACK is returned to the source station, and the PC waits a PIFS interval following the ACK frame before transmitting any additional frames. Figure 10 illustrates station-to-station frame transmission during the CFP. The PC may also choose to transmit a frame to a non-CF-aware station. Upon successful receipt of the frame, the station would wait a SIFS interval and reply to the PC with a standard ACK frame. Fragmentation and reassembly are also accommodated with the Fragmentation_Threshold value used to determine whether MSDUs are fragmented prior to transmission. It is the responsibility of the destination station to reassemble the fragments to form the original MSDU.



■ **Figure 9.** *PC-to-station transmission.*



■ **Figure 10.** *Station-to-station transmissions.*

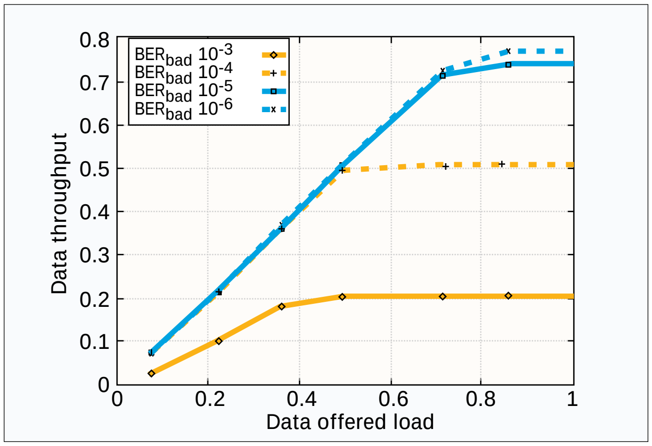


Figure 11. *Burst error effects on data throughput.*

Simulation Model

Two different simulation models are presented in this article. 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. More detailed explanation of the simulation model is found in [9].

Several assumptions have been made to reduce the complexity of the model. A short description of each of the assumptions is provided below:

- The effects of propagation delay on the model are neglected. This is a fairly realistic assumption if transmission distances are on the order of 100 ft between stations.
- The "hidden terminal" problem is not addressed in the simulation models.
- The basic rate of 1 Mb/s was simulated for the DSSS. This decision was made because the enhanced rate, 2 Mb/s, would add additional complexity since control, multicast, and broadcast frames are required to be transmitted at the basic rate (to ensure that all stations in the BSS can be properly received), while management and data frames are transmitted at any available rate (1 Mb/s or 2 Mb/s).
- No stations operate in the "power-saving" mode (PS-Mode). By requiring all stations to be "awake" at all times, transmitted MPDUs can be received immediately by the destination station without buffering at the AP.
- No interference is considered from nearby BSSs reusing the same DSSS spreading sequence.

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.

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 characterize fading in the communications channel [10]. A two-state continuous-time Markov chain is used to represent the burst error model. 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 $\mathrm{BER}_{\mathrm{good}}$). State *B* indicates the channel is operating in a fading condition with a higher error rate, denoted by $\mathrm{BER}_{\mathrm{bad}}$. The transition rate from state *G* to state *B* is denoted by α, while the transition rate from state *B* to state *G* is denoted by β. A frame is considered to be corrupt if it contains one or more bit errors.

The simulation model uses the error model above to determine whether each transmitted frame or MPDU was transmitted successfully. When the frame is transmitted, a portion of the frame can be sent over the communications medium when the channel is in state *G*, and a portion can be transmitted when the channel is in state *B*. The number of bits transmitted in the frame during state *B* is denoted by n_1, and the number transmitted during state *G* is n_2. The probability that the frame is transmitted successfully is then calculated as

$$P_r\{success\} = (1 - \mathrm{BER}_{\mathrm{bad}})^{n_1} \cdot (1 - \mathrm{BER}_{\mathrm{good}})^{n_2}.$$

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 interarrival 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. Typical default values used in the simulation of the ad hoc network are illustrated in Table 1.

Infrastructure Network Model

The effect of a single BSS with an AP is simulated, 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 Fig. 8, 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 s; therefore, approximately 94 percent of the repetition interval can be allocated by the AP for contention-free services.

During the CFP, if a station is polled by the AP to transmit, the station can transmit directly to another station in the BSS (Fig. 10) 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 size of the BSS is relatively small, all packetized voice activity is assumed to occur between stations in different BSSs. Therefore, the simulation model directs all voice traffic from a station through the AP. All voice traffic destined for a mobile station is also delivered via the AP.

The polling scheme during the CFP uses a cyclical scheduling 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; therefore, priority polling mechanisms are not required. A simple polling scheme is proposed 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 (i.e., *Null Function*), 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 transmitting and 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 s, and the mean value of the talk spurt (ON) period is 1 s. The voice transmission rate in the ON state is 64 kb/s. The transition rates are representative of real telephonic speech patterns that were obtained from measurement [11].

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. QoS parameters for voice typically limit maximum delay to 25 ms without echo canceling, and 500 ms using echo canceling [12]. Asynchronous data frames are transmitted in the CP portion of the repetition interval using the DCF protocol described above. Table 2 lists the additional default values used for simulation of the infrastructure network.

Attribute	Typical value
Data stations	10
Average MSDU length	1000 octets
Channel rate	1 Mb/s
BER_{good}	10^{-10}
α	$30\ s^{-1}$
β	$10\ s^{-1}$
RTS_Threshold	250 octets
Fragmentation_Threshold	800 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 µs
SIFS_Time	10 µs
DIFS_Time	50 µs

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

Attribute	Typical value
BER_{bad}	10^{-5}
Number of voice stations	10
Voice transmission rate	64 kb/s
Voice station buffer size	100 frames
CFP_Max_Duration	0.39 s
CFP_Repetition_Interval	0.41 s
PIFS_Time	30 µs

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

Simulation Results

Simulation results are shown for an ad hoc network and an infrastructure network. The results below are presented in the form of plots and, where applicable, with 95 percent confidence intervals. The throughput plots shown below represent aggregate throughput. Approximate throughput per station can be calculated by dividing the aggregate throughput by the total number of data stations in the BSS.

Ad Hoc Network

For the ad hoc network, we assume all mobile stations generate asynchronous data traffic with the same intensity. Figure 11 shows the aggregate data throughput in megabits per second versus the offered load in megabits per second for several BERs (i.e., the BER_{bad}). 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.

Note that the burst error transition rates for this model indicate that more time will be spent in the "bad" state than in the "good" state. When the medium is relatively clean, BER_{bad} is less than 10^{-6}, and a maximum throughput of approximately 77 percent is possible. However, the maximum throughput can drop to approximately 20 percent under harsh fading. Thus, it is clear that the channel condition can adversely affect the throughput performance of the IEEE 802.11 system. It is also noted that the throughput saturates around 90 percent under ideal channel conditions due to overhead, collisions, IFS, and backoff intervals.

The Effect of RTS on Maximum Data Throughput

As stated previously, the RTS/CTS handshaking mechanism is used to combat the effects of collisions. The RTS/CTS reserves the channel for transmission of a larger data packet, with the desired effect that if a collision occurs with the RTS/CTS handshake, less bandwidth will have been wasted than if the larger data packet had been transmitted and corrupted. RTS_Threshold is a manageable parameter used to determine when to precede a data packet with an RTS/CTS handshake. In the plots below, the maximum data throughput is plotted against RTS_Threshold for various values of data MSDU length. The maximum data throughput is defined as the maximum value of throughput obtained over all offered loads when RTS_Threshold is held constant at a specified value. A bursty channel error model is used with the transition rates given in Table 1.

Figure 12 shows the impact of RTS_Threshold on maximum data throughput when there is no fragmentation of data packets. As shown in the figure, under all of the MSDU values the peak throughput occurs when the RTS_Threshold is set at approximately 250 octets. The maximum throughput values begin to taper off considerably when the RTS_Threshold begins to exceed 400 octets, indicating that collisions are having an adverse impact on system throughput. We have also varied the number of data stations and observed the same conclusion.

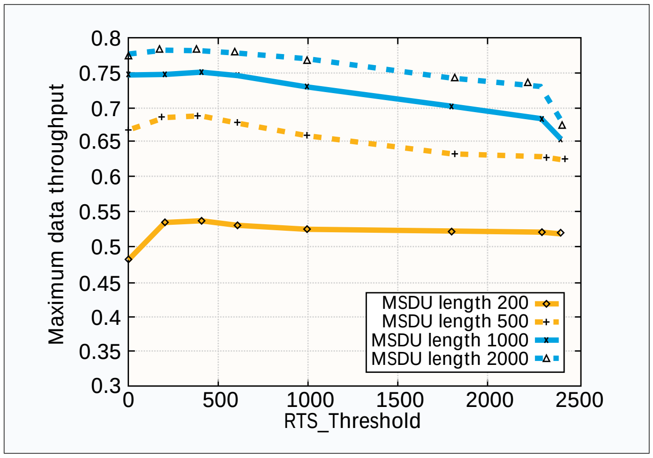


Figure 12. *RTS_Threshold effects on data throughput.*

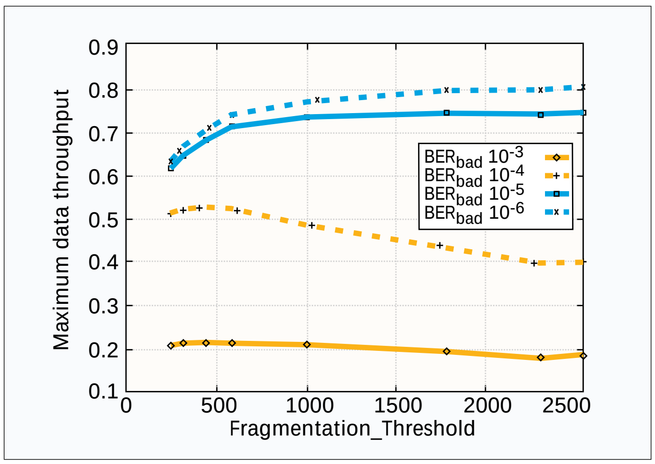


Figure 13. *Fragmentation_Threshold effects on data throughput.*

Effect of Fragmentation_Threshold on Maximum Data Throughput

The Fragmentation_Threshold is used to combat the effects of poor channel quality. By reducing the size of the packets transmitted, there is a better probability of successful transmission, especially under poor channel conditions. However, under good channel conditions, fragmentation is a hindrance because the associated overhead tends to reduce the aggregate throughput. Figure 13 shows the maximum data throughput plotted against Fragmentation_Threshold for various values of BER_{bad}, when the average MSDU length is 1000 octets. When the channel is in a good condition (i.e., BER_{bad} less than 10^{-5}), fragmentation only hinders the maximum throughput because of the additional overhead. However, when BER_{bad} is high, the benefits of using fragmentation become apparent. In the figure, the difference between the peak and smallest values of maximum throughput for the BER_{bad} 10^{-4} curve is almost 140 kb/s.

Since a typical WLAN terminal will experience the whole spectrum of channel qualities, the optimum Fragmentation_Threshold should be set between 500 and 800 octets. This range of values is ideal for neither a clean channel nor a degraded channel, but offers acceptable performance across the entire spectrum of channel qualities.

Effect of MSDU Length on Data Throughput

Figure 14 shows the effects of average MSDU length, l_{data}, on throughput performance. The curves are obtained for a BER_{bad} of 10^{-5}. The IEEE 802.11 MAC and PHY layers add a total of 58 octets for overhead. Given a clean channel like that shown in Fig. 14, the longer the MSDU is, the more efficient the system becomes. When the channel is operating in a degraded mode, we have observed that the benefits of a large MSDU length become less pronounced.

Infrastructure Network

The infrastructure network supporting voice and data traffic is now considered. Data traffic is transported through the CFP and voice traffic through the CP. All results below are shown using a CFP repetition interval of four beacon periods. With five simultaneous voice conversations in progress (10 stations total), the aggregate voice throughput is approximately 272 kb/s. This is calculated by considering that each voice station is transmitting at 64 kb/s and the channel is in the ON state for 42.5 percent of the time (based on the ON/OFF model described above).

The Effect of Voice Payload Length on Performance

The effect of voice payload length, l_{voice}, on both data and voice performances is investigated first. The number of voice stations, N_{voice}, is set to five pairs. The first five voice stations are located in the BSS; the others are located elsewhere and are generated through the AP. This voice scheme is used because it is assumed that voice traffic would not occur between stations within the same BSS, due to their close proximity to each other. For voice traffic, only the delay between an AP and another mobile station in the same BSS is considered. All measurements are done at the MSAP (MAC service access point). Figure 15 displays the influence of the voice payload length on data traffic performance.

In Fig. 15, 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. Since voice packets, unlike data packets, are bounded by a specified delay (e.g., 0.5 s), any packets exceeding the delay requirements must be discarded. A complementary cumulative distribution plot is used to determine the percentage of voice packets which are discarded because they are not transmitted within the delay bounds. The figures illustrate the complementary cumulative distribution, $P_r\{X > x\}$, for voice delay in seconds. As discussed previously, voice delay can tolerate as much as 0.5 s if an echo canceler is used. Without an echo canceler, a much more stringent voice delay (under 25 ms)

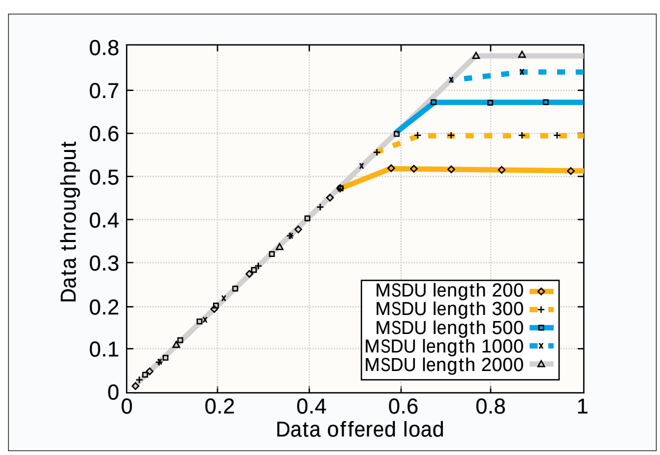


Figure 14. *Average MSDU length effects on data throughput.*

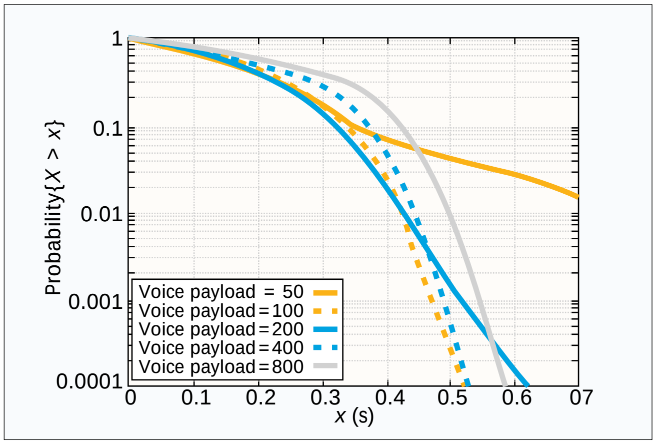


Figure 15. *Complementary cumulative distribution for voice delay.*

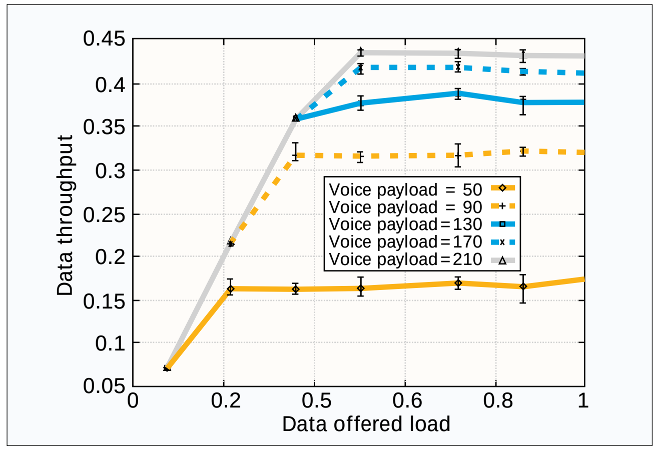


Figure 16. *Effect of* 1_{voice} *on data throughput.*

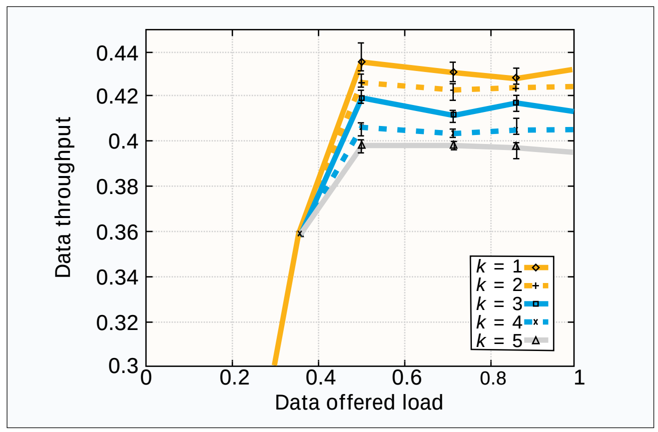


Figure 17. *Data throughput vs. offered load for several values of* k.

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. Thus, it is assumed that an echo canceler is employed, and that voice packets delayed by more than 0.5 s at the receiver 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. For clean voice quality communications, a packet loss rate of 1 percent should be maintained [13]. Shorter voice payloads incur larger overheads, translating into longer delays. At the other extreme, longer payloads imply longer packetization delays. Thus, these two parameters must be traded off. As is seen from Fig. 15, the best operating points appear to be around 100–400 octets long for voice payload. When the CFP repetition interval is 5, the recommended voice payload lengths have been shown to be 100–200 octets long [9]. Note that the average delay calculated when the voice payload is 50, 100, 200, 400, and 800 octets is 200, 186, 205, 233, and 284 ms, respectively.

Figure 16 shows the impact of voice payload length on data throughput over a range of offered loads. It is shown that data traffic will suffer more as the voice payload length is decreased. Given a fixed amount of voice information to be transmitted during the CFP, shortening the voice payload length will result in more frames (i.e., overhead) transmitted. Shortening the payload length will therefore lengthen the duration of CFP operation, leaving less available bandwidth for the transmission of data if the CFP is foreshortened. Thus, from the point of view of data traffic, the voice payload should be made relatively long. However, beyond 200 octets, the data throughput improvement is marginal.

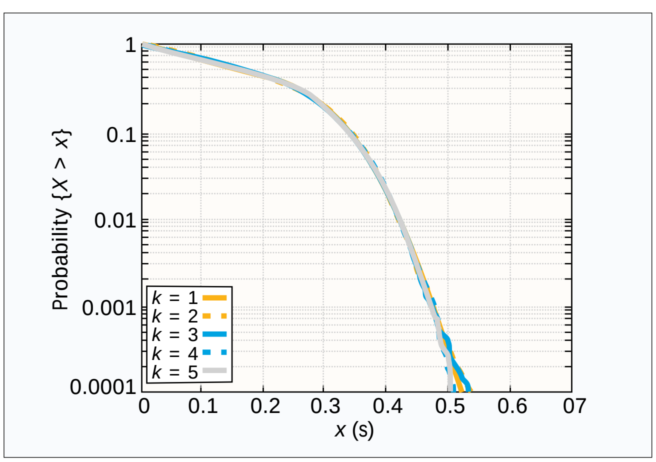


Figure 18. *Effect of* k *on voice delay.*

The Effect of Polling Scheme on Performance

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, the effect of k on data throughput and voice delay is plotted, as illustrated in Figs. 17 and 18. Figure 17 shows throughput plotted against offered load. For the PCF, five voice station pairs are used with voice payload fixed at 200 octets.

The curves indicate that a higher value of k tends to reduce the aggregate data throughput. When k increases, there is a higher probability that a voice station will receive or have traffic to transmit, which tends to prolong the duration of the CFP. Prolonging the CFP corresponds to a reduction in the amount of time that data stations have access to the channel.

In Fig. 18, the value of k has very little impact on the voice packet loss rate, mainly due to the fact that voice stations operate on an ON/OFF basis. That is, when a voice station does not have any data to send during an OFF period, 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 ($k = 1$). If the CFP is foreshortened due to light traffic at that particular instant in time, the wait until the next polling cycle is still well under the acceptable delay specifications levied by the echo canceler. Therefore, from a data throughput perspective, it is best to select $k = 1$ and have a foreshortened CFP period.

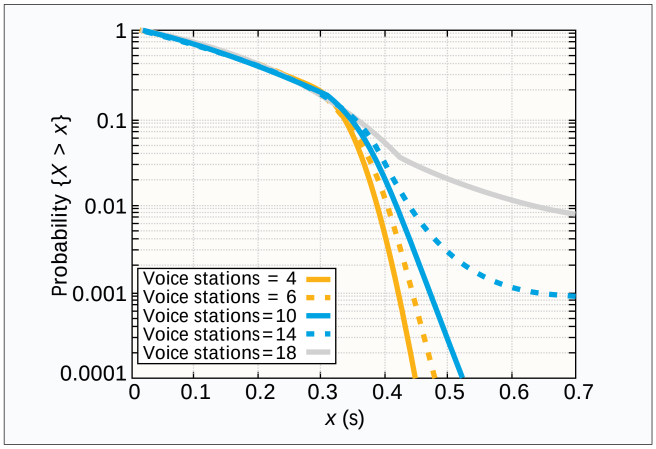


Figure 19. *Effect of voice stations on voice delay.*

Impact of the Number of Voice Stations on Performance

The results of this last section concentrate on the effect the number of voice stations has on voice delay in the infrastructure network. The PCF is simulated using a fixed voice payload of 200 octets, $k = 1$, and the number of voice stations varies between 4 and 18.

In Fig. 19, the complementary cumulative distribution is plotted for voice delay. As the number of voice stations increases, so does the amount of packet loss. This is due to the fact that more voice packets will be competing as the number of voice stations increases. When the CFP interval is set to 4, approximately up to 16 voice stations can be supported.

Conclusion

The primary contributions of this work include a detailed investigation of both the DCF and the PCF operating over a common CFP repetition interval. The simulation model includes asynchronous data being transmitted over the DCF, which is not delay-sensitive, and packetized voice traffic transmitted over the PCF, which requires bounded delay. The model includes the effect of a bursty error channel, which is typical of a wireless radio environment where multipath fading is commonly experienced. The final contribution includes a scheme to drop voice stations from the CFP if they are idle for a specific period of time. Dropping idle voice stations frees available bandwidth for stations with packets queued for transmission.

The general conclusions derived from the 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 800 octets, the RTS_Threshold is set to 250 octets, and the medium is relatively clean (BER better than 10^{-6}).
- Based on our assumptions and simulation model, 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 approximately 200 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 (i.e., $k = 1$) so that the remaining bandwidth can be allocated to other stations.

References

[1] R.O. LaMaire *et al.,* "Wireless LANs and Mobile Networking: Standards and Future Directions," *IEEE Commun. Mag.*, vol. 34, no. 8, Aug. 1996, pp. 86–94.

[2] ETSI TC-RES, "Radio Equipment and Systems (RES); High Performance Radio Local Area Network (HIPERLAN); Functional Specification," ETSI, 06921 Sophia Antipolis Cedex, France, draft prETS 300 652, July 1995.

[3] Wireless Medium Access Control and Physical Layer WG, IEEE Draft Standard P802.11, "Wireless LAN," IEEE Stds. Dept, D3, Jan. 1996.

[4] K. C. Chen, "Medium Access Control of Wireless LANs for Mobile Computing," *IEEE Network*, vol. 8, no. 5, Sept. 1994, pp. 50–63.

[5] H. S. Chhaya and S. Gupta, "Throughput and Fairness Properties of Asynchronous Data Transfer Methods in the IEEE 802.11 MAC Protocol," *PIMRC '95*, 1995, pp. 613–17.

[6] W. Diepstraten, "A Wireless MAC Protocol Comparison," IEEE P802.11-92/51.

[7] J. Weinmiller, H. Woesner, and A. Wolisz, "Analyzing and Improving the IEEE 802.11-MAC Protocol for Wireless LANs," *Proc. MASCOTS '96*, San Jose, CA, Feb. 1996, pp. 200–6.

[8] D. Bantz and F. Bauchot, "Wireless LAN Design Alternatives," *IEEE Network*, vol. 8, no. 2, Apr. 1994, pp. 43–53.

[9] B. Crow *et al.*, "Investigation of the IEEE 802.11 Medium Access Control (MAC) Sublayer Functions," *Proc. INFOCOM 97*, Kobe, Japan, Apr. 1997, pp. 126–33.

[10] E. Gilbert, "Capacity of a Burst Noise Channel," *Bell Sys. Tech. J.*, vol. 39, Sept. 1960, pp. 1253–66.

[11] P. Brady, "A Model for Generating On-Off Speech Patterns in Two-Way Conversation," *Bell Sys. Tech. J.*, vol. 48, no. 7, Sept. 1969, pp. 2445–72.

[12] M. de Prycker, *Asynchronous Transfer Mode: Solution for Broadband ISDN*, 3rd ed., Englewood Cliffs, NJ: Prentice Hall, 1995.

[13] L. Hanzo *et al.*, "A Packet Reservation Multiple Access Assisted Cordless Telecommunications Scheme," *IEEE Trans. Vehic. Tech.*, vol. 43, no. 2, May 1994, pp. 234–44.

Biographies

Brian P. Crow (bcrow@mitre.org) received a B.S. from Arizona State University in 1987 and an M.S. from the University of Arizona in 1996. From 1988 to 1992, he was a signal officer in the U.S. Army. He is currently a lead engineer with the MITRE Corporation. His research interests include wireless and broadband networks, and network and systems management.

Indra Widjaja received a B.A.Sc. degree from the University of British Columbia, an M.S. from Columbia University, and a Ph.D. from the University of Toronto, all in electrical engineering. From 1994 to 1997, he was assistant professor of ECE at the University of Arizona. Since July 1997, he has been with Fujitsu Network Communications. His research interests include mobile and wireless networks, switching architectures, and traffic management.

Jeong Geun Kim received B.S. and M.S. degrees, both in electrical engineering, from Yonsei University, Seoul, Korea, in 1990 and 1992, respectively. He is pursing a Ph.D. at the University of Arizona, studying issues of wireless ATM and performance evaluation of broadband networks.

Prescott T. Sakai received an M.S.E.E. from the University of Arizona in Tucson. He is currently a new product planning and applications engineer at Cypress Semiconductor, where he is involved with defining next-generation data communication products.